

Application Guide

IA ETHICS SELF-ASSESSMENT FOR ACTORS OF THE ENTREPRENEURIAL ECOSYSTEM

May 2021

Copyright © 2021 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the IDB, its Board of Directors or the countries they represent, nor of the MIF (IDB Lab) Donors Committee or the countries it represents.



fAIR LAC, the IDB Group initiative that promotes the ethical and responsible use of artificial intelligence, through IDB Lab has developed a practical AI ethical self-assessment tool for entrepreneurs, which allows an analysis of the technological solutions based on AI and data management. This diagnosis helps entrepreneurs to improve their product development, while identifying the main areas of attention to prevent errors, biases, discrimination and exclusions resulting from technological deployment. The ethical self-assessment of AI for entrepreneurs that you have in your hands is the first product of fAIR LAC for entrepreneurs, which is a Guideline with a multidisciplinary approach that includes six main dimensions: 1. Conceptualization and design, 2. Governance and security, 3. Human involvement in AI systems, 4. AI life cycle (data and algorithms), 5. Relevant actors and 6. Communications. The purpose is for entrepreneurs to have a quick reference of which are the most important aspects to consider in each of these dimensions, in order to have a complete vision of the ethical implications of their products and thus establish the pertinent improvement and mitigation measures.

The main innovation of this document lies in two main aspects: the first, by not only placing the onus of self-regulation solely on entrepreneurs, but also involving two key actors for the ecosystem: project funders and accelerators. The second aspect is that the guiding questions correspond to three levels of business development from early stages from the ideation to more mature ventures or small and medium-sized enterprises (SMEs) that develop innovative products. Likewise, although the document is primarily intended to guide the development and implementation of AI-based solutions, it is also useful for solutions based on data management.

We invite those interested to download the publication and to be part of the entrepreneurial journey to develop technological solutions with social impact that contribute to sustained regional development that leaves no one behind.

AUTHORS

César Said Rosales Torres
César Buenadicha
Tetsuro Narita

ACKNOWLEDGEMENTS

First and foremost, we would like to thank **Irene Arias**, CEO of IDB Lab, and **Marcelo Cabrol**, Manager of the Social Sector at IDB for their continuous support and leadership on the initiative.

We express our deep gratitude to the IDB Lab team for their time and **William Ernest**, **Paula Auerbach**, **Jessica Leite**, **Patricio Aznar**, **Cecilia Franco**, as well as to Mara Balestrini. In the same way, the IDB LAB team appreciates the collaboration and feedback of the following experts in the field of Artificial Intelligence, as well as to the members of the project's Advisory Board. The list is presented below in alphabetical order:

Adrián Soto, professor at the Faculty of Engineering and Sciences at the Adolfo Ibáñez University. Chile.

Cecilia Tham, Co-Founder and CEO of All Women. Spain.

Constanza Gómez Mont, founder and CEO of C Minds. Mexico.

Enrique Cortés Rello, director of the Artificial Intelligence Hub, Instituto Tecnológico y de Estudios Superiores de Monterrey. Mexico.

Francesca Gabetti, Founder and CEO of TeamEQ. Spain.

Irene Velasco, founder and CEO of helKi. Mexico.

Iván Caballero, founder and CEO of Citibeats. Spain.

Jesús Cepeda, co-founder and CEO of OS City. Mexico.

Juan Carlos Holguin, co-founder and CEO of LinkIn. Ecuador.

Juan Eduardo Orlandi, General Manager of Magical. Chile.

Juan Roberto Hernández Villalobos, coordinator of fAIr Jalisco of the Instituto Tecnológico y de Estudios Superiores de Monterrey. Mexico.

Leo Prieto, founder and CEO of Odd Industries. Chile.

Leopoldo Bertossi, professor at the Faculty of Engineering and Sciences of the Adolfo Ibáñez University. Chile.

Lorena Barrenechea, legal advisor to the Inter-American Development Bank.

María Paz Hermosilla, director of the GobLab of the Adolfo Ibáñez University. Chile.

Nacho Lafuente, founder and CEO of Datumize. Spain.

Natalia González, consultant for the Inter-American Development Bank.

Ricardo Baeza-Yates, Research Director at the Institute for Experimental AI in Northeastern University, USA.

Roberto Sánchez, consultant for the Inter-American Development Bank.

Romina Garrido, associate researcher at the GobLab of the Adolfo Ibáñez University. Chile.

Sylvia Chebi, CEO of ThalesLab. Uruguay.

INDEX

Introduction	6
How to use this tool	12
Conceptualization and design	14
1.1 Determining the main purpose of AI in business operations	15
1.2 Assessment of the digital ecosystem	17
1.2.1 Technical assessment	17
1.2.1.1 Determination of the Technology Readiness Level (TRL)	
1.2.2 Legal assessment	19
1.2.3 Sectoral assessment	22
1.3 Social and environmental impact	25
Governance and Security: Internal governance structures and control of AI	28
2.1 Corporate structure for the governance of AI	30
2.2 Risk management and internal controls	32
2.2.1 Data security	34
Human involvement in AI systems	36
3.1 Human-AI interaction determining the level of human supervision	38
AI Lifecycle: AI System Operations Management	40
4.1 Data source and management	41
4.1.1 Data type	42
4.1.2 Data processing	44
4.1.3 Data integrity and confidentiality	45
4.1.4 Privacy by design	46
4.1.5 Interoperability	47
4.1.6 Testing and validation	48

INDEX

4.2 Model development: Algorithms	49
4.2.1 Traceability	49
4.2.2 Explicability of AI processes and results	50
4.2.3 Replicability	52
4.2.4 Reproducibility	53
4.2.5 Auditability	54
4.2.6 Maintenance	55
Relationship with key stakeholders	56
5.1 Transparency	58
5.2 Consumers and users of the system	59
5.3 Access to financing	60
Communications	63
Annex 1 - Description of the Stages of Business Development	66
Sources	69

INTRODUCTION

INTRODUCTION

During the last decade, Artificial Intelligence (AI) has seen significant growth, despite false starts in previous decades mainly caused by overly optimistic expectations about the level of technological development of AI, the lack of historical data infrastructure and high-quality data in various sectors of the economy, as well as multiple market factors and large-scale commercial use of these technologies (MGI 2017).

Currently, the picture is very different, thanks to scientific developments in recent years and the commercial application of AI at scale in sectors such as automated vehicles, natural language processing, facial recognition, computer vision, among others. (MGI 2017; OECD 2018). This new dynamic, in turn, has aroused greater interest among investors and governments to promote and facilitate the growth of AI ecosystems, while at the same time regulating essential aspects such as data protection, human-AI interaction, and the impact it could have

on people's social and political life (OECD 2018).

However, despite this progress, the degree of AI adoption remains relatively low in the world compared to other technologies and highly concentrated in geographical areas such as the United States, Canada, China, and Europe, each with different governance and investment models (Gigler 2020). Likewise, there is a gap between large corporations and startups in getting access to financing and other



resources. In Latin America, the level of AI development remains low when compared to the rest of the leading regions and concentrated in countries such as Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Uruguay, where AI is mainly used for technological developments based in third-party models, chatbots, natural language processing, and text conversion, and to a lesser extent for classification and prediction purposes (Everis & Endeavor 2018). In addition to the disparity in the levels of development and investment with the United States, Canada, China, and Europe, Latin America faces low levels of both public and private investment, as well as limitations for the professional development of experts in data management and the creation of data value chains on which machine learning-based AI systems can operate more efficiently (OECD 2018).

In this context, at the end of 2019, the Inter-American Development Bank and IDB LAB launched an initiative called fAIr LAC to promote the ethical and responsible use of AI in the region. The main objective is to harness the technology's immense potential to create more efficient, fair, and personalized social services for the citizens of Latin America and the Caribbean (IDB 2020). Today, more countries are making use

of this type of technology that touches practically all areas of development in the region, so it is essential to respect the privacy of citizens' data and anticipate possible biases in the construction of algorithms (Google 2018).

Within the framework of this initiative, IDB LAB started in 2020 the project "Entrepreneurial Journey: articulation of an ethical and responsible ecosystem of artificial intelligence startups," which aims to contribute to the strengthening of entrepreneurial ecosystems for the adoption of AI, with the support of entrepreneurs, investors, business accelerators, data experts, academia, civil society, and government. This is achieved through tools and services that allow entrepreneurs to use this technology dynamically and innovatively, in line with ethical principles that maximize benefits and minimize risks.



The document in your hands is the first product of the “Entrepreneurial Journey” project within the fAIr LAC initiative, and it is a self-assessment guide for companies, investors, and accelerators on crucial aspects to consider ethical aspects of business operations, according to the various level of maturity of the startups. The innovation of this tool lies in two main aspects: first, it does not place the onus of self-assessment and risk management solely on entrepreneurs, but involves two key actors for the ecosystem: project funders and accelerators. The second aspect is that the guiding questions correspond to three levels of companies’ development from the early stage, going through a

later stage to mature companies that can be presented to larger rounds of financing. Likewise, the document takes a broad approach to artificial intelligence, recognizing that many Latin American companies have robust data management capabilities that could be conducive to the development and adoption of AI technologies.

More so, the development of this product rests on the following principles for the ethical and responsible adoption of AI for the development of Latin America and the Caribbean, while considering the ethical principles for AI of the OECD (OECD 2019), namely:

AI for development

This project aims at promoting initiatives that use AI for positive and scalable social impact, in order to continue improving the economic and social well-being of the citizens of the region.

Integrity

The development and use of AI systems should be consistent with their purpose of social impact, in addition to respecting the use of data and metadata from clients and third parties with systems that guarantee their reliability.

Trust and confidence

Confidence in the use of AI technology should be based on a harm prevention and risk mitigation approach and the transparency of its operations. All this to guarantee principles of equity and non-discrimination in the use of this technology.

Finally, the ethical dimension of this project is not only observed through the legal fulfillment of the obligations of the companies but also the solidity of the proposals and the objectives

of technological use for the benefit, empowerment, and protection of common goods, highlighting above all human dignity and commonly accepted social principles.



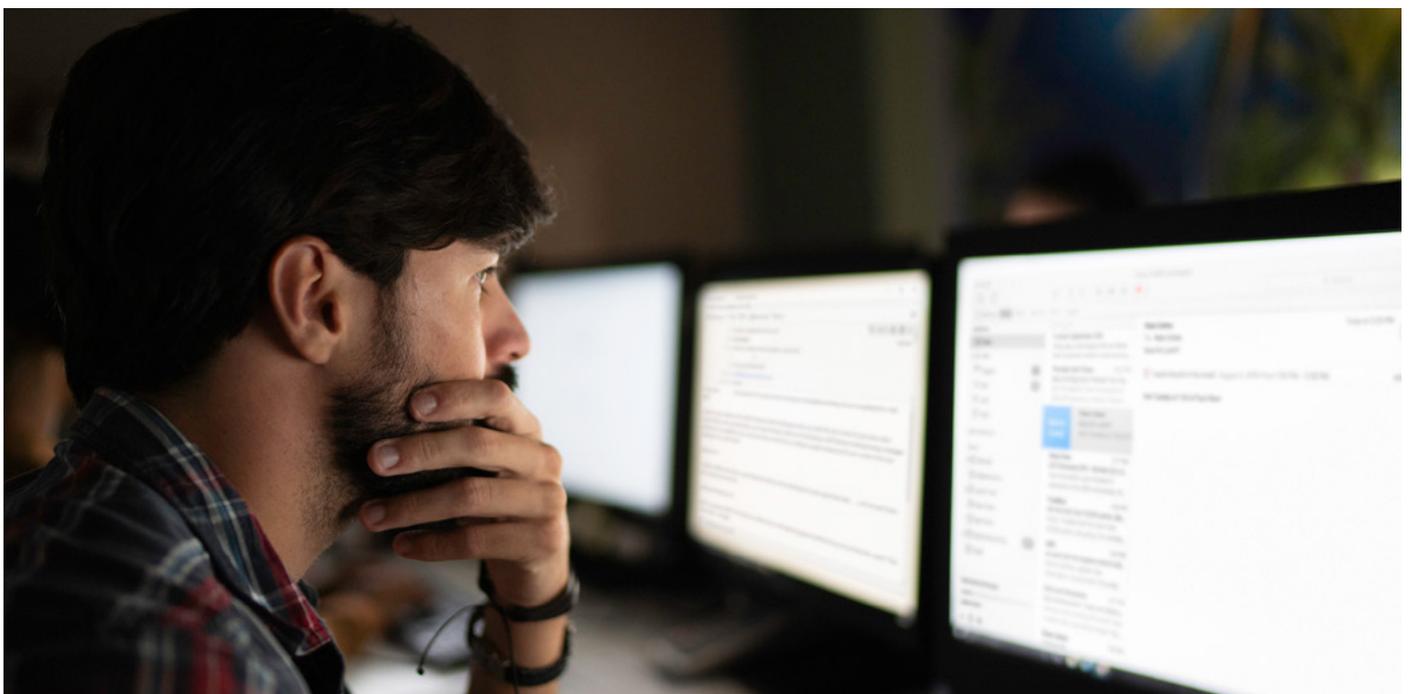
HOW TO USE THIS TOOL

This self-assessment tool of voluntary adoption is aimed at those who design, develop, deploy or use Artificial Intelligence technologies, as well as big data handling, including companies in different development stages, investors, and accelerators. But it is also a useful tool for all those interested in the processes of adopting artificial intelligence ethically and responsibly.

The tool is a non-exhaustive general guide of questions that allow identifying the purpose of using AI, strengthening governance structures and internal controls around AI systems, determining the best balance of human involvement in AI, follow up on the data control processes and algorithms, as well as suggestions on the relationship with key stakeholders such as users and clients, investors and supervisory bodies.

It is a progressive ethical self-assessment matrix since it is divided into three

stages (the details can be found in Annex 1 of this document) according to the enterprise's level of development, the level of technological development, and the actors who participate in each one of these stages. These stages are non-rigid, pre-established concepts, but rather a minimal guidance to better use and understand the guiding questions in each dimension. In this sense, although the questions are intended to be ordered according to the companies' development level, some questions may apply to different stages according to



the particularities of the business or sector in which it operates. Likewise, it is relevant to remember that this guide is not a compliance list, but a versatile list adaptable to the needs of each entrepreneur, investor, or anyone interested in the ethical and responsible use of AI. The main objective of this tool is to help entrepreneurs improve their solutions from an ethical perspective, as well as to identify and mitigate risks.

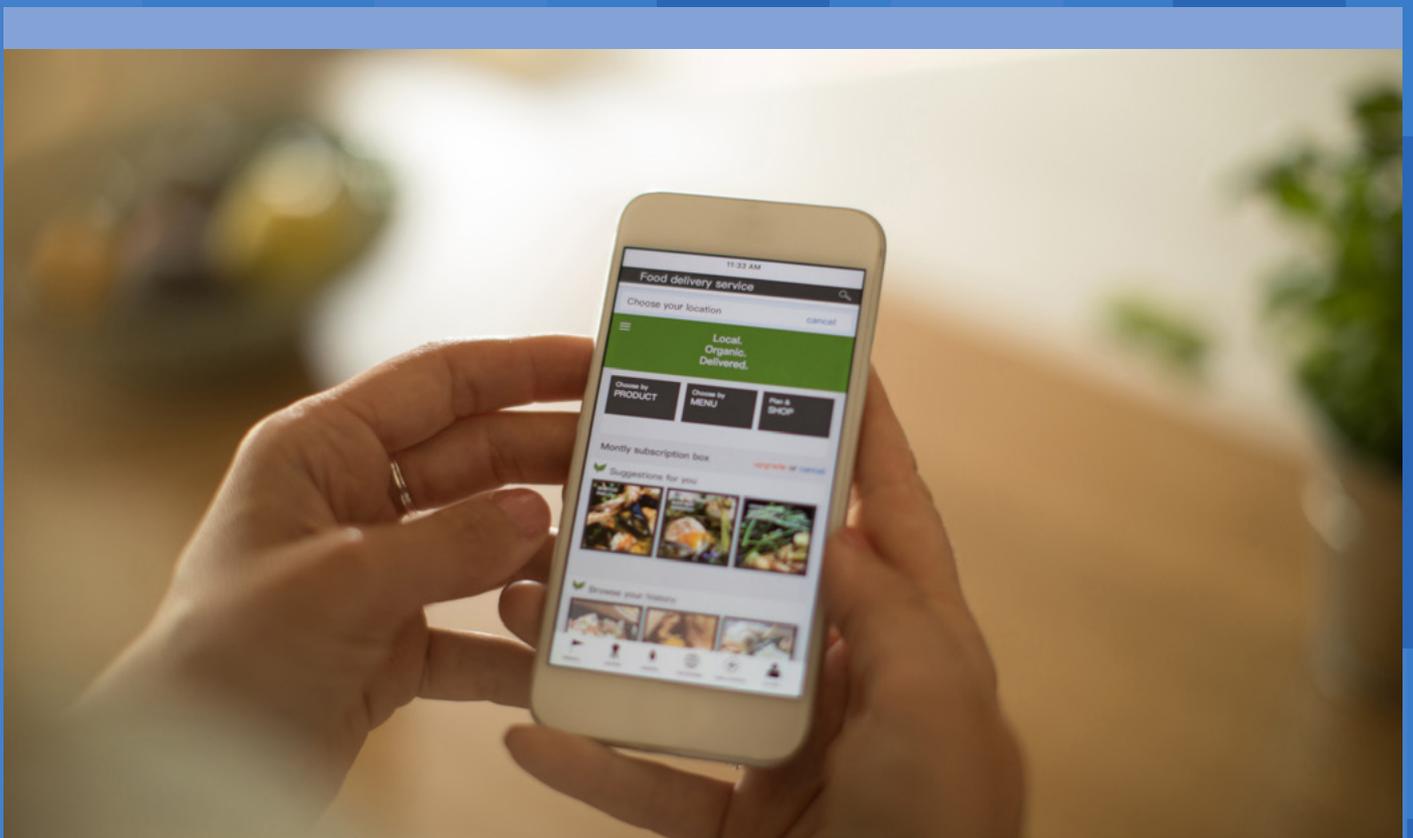
This is the first version of the self-assessment, which has received the review, comments, and guidance of entrepreneurs, investors, legal experts, and data scientists, to guarantee its ease of use and relevance of the document. It is an iterative process in which a broad consultation round is planned with more companies, academics, civil society,

governments, investors, lawyers, and data scientists to improve the ethical self-assessment and carry out a pilot testing phase that allows Identify strategic actions to strengthen the entrepreneurial ecosystem in AI.

We hope that this tool will be useful for all those interested in the ethical and responsible use of AI and contributes to the strengthening of the ecosystem to promote the competitiveness of companies in the region while achieving a positive social impact.



1



CONCEPTUALIZATION AND DESIGN

1

CONCEPTUALIZATION AND DESIGN

This first section includes various initial aspects that should be considered before the adoption of AI as a base technology for startups. These aspects include the determining the operational purpose of the AI and the legal, technical, and sectorial assessments necessary to define the degree of suitability of the technology in business operations. In addition, there is a section of guiding questions related to social impact.

1.1 DETERMINING THE MAIN PURPOSE OF AI IN BUSINESS OPERATIONS

The first step for the adoption and responsible and ethical use of Artificial Intelligence (AI) is to define the utility that such system will generate for the business model and its suitability within the corporate strategy. This stage requires a realistic and tangible assessment of the capabilities and limitations that the AI system would have (WEF & IMDA 2020). It is also advisable carrying out a detailed analysis of use cases in similar sectors to maximize the chances of success and reduce the margin of uncertainty of the initiative based on previous experiences in applying AI. Finally, in

order to understand the usefulness of the AI system, it is possible to carry out comparative assessments of the solutions with or without the use of AI models.

At this first moment, it should be noted that for the ethical and responsible adoption of AI, it is advisable to think about solutions to real problems for which AI is the most suitable technology, beyond simply incorporating AI to offer interesting products to the market based on this technology. (MGI 2017; Szramke 2017; EC 2018; IMDA & PDPC 2020).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has the organization or company set a goal with impact metrics for its business operations in which AI is the most appropriate technology, beyond just offering the incorporation of innovative technology in its products or services?</p> <p>What kinds of tasks are intended to be carried out with the AI system, such as: automatic decision-making, predictions, recommendations, interaction with users, etc.?</p> <p>What type of AI technology is going to be used, for example: <i>machine learning (ML), deep learning or deep learning, artificial neural networks, computer vision or natural language processing (NLP), virtual agents, smart robotics, or autonomous vehicles?</i></p>	<p>Have examples of use cases been explored in which AI technologies or similar have been used for the purposes established by the company for your particular case?</p> <p>Have the lessons learned from similar use cases been incorporated in terms of the type of AI technology and sector?</p>	<p>Have comparative studies been carried out between different technologies or methods to determine if AI is the simplest and most appropriate fit? For example, through the use of programming or simulation exercises that do not make use of AI.</p> <p>In case AI is the most suitable set of technologies for the company's purposes, has a comparison been made between the different AI models to determine which would be the most efficient? For example, rule-based models or models based on machine learning through big data.</p> <p>Once the type of AI model has been determined, has it been decided which kind of algorithm would be the most appropriate to use? For example: <i>regression algorithms, decision trees, clusters, or neural networks.</i></p>
INVESTORS	<p>Is the approach to AI deployment in line with the fund's investment thesis? in addition to having a clear business purpose.</p>	<p>Is the company moving towards AI from another type of data usage or processing?</p>	<p>Is the selection of the type of AI consistent with the purposes for which the AI will be used?</p>
ACCELERATORS	<p>Do the company's founding partners have a minimal understanding of the benefits and risks of adopting AI?</p>	<p>Do business partners and employees have minimal understanding of the benefits and risks of adopting AI?</p>	<p>Same.</p>

1.2 ASSESSMENT OF THE DIGITAL ECOSYSTEM

The evaluation of the digital ecosystem allows companies to carry out recognition of the technological infrastructure that allows assessing the viability and usefulness of the development of an AI system through certain technical, legal, and sectorial characteristics (Galdon 2020).

1.2.1 Technical assessment

This dimension makes it possible to evaluate in technical terms the availability of data and minimum critical infrastructure to guarantee the operability of the AI system with a risk prevention approach (EC 2018; IMDA & PDPC 2020).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are there various sources of historical data that allow the creation of an AI model? For example, data lakes, government initiatives such as smart cities, etc.</p> <p>Is it possible to access the data from these sources?</p> <p>How is it guaranteed to obtain complete, up-to-date, and high-quality data?</p> <p>Is there sufficient information and telecommunication infrastructure to guarantee the constant operability of the AI model?</p> <p>Has the initiative considered the recommendations issued by international standardization bodies that facilitate compliance with legal obligations?</p>	<p>Has the need to require additional data been considered in order to reduce biases?</p> <p>Regarding the AI system's architecture, have the set of rules and restrictions been established on which the AI system should operate and not violate?</p> <p>In the case of using AI models for robotics, have considerations been taken to determine the sufficiency of all the technical elements required based on an SMPA (Sense-Model-Plan-Act) model so that all the elements of the environment can be recognized and that the consideration of plans and actions of the system meet the expectations of behavior and results? In this sense, it is important to define the behavioral expectations of the system clearly.</p>	Same.
INVESTORS	<p>Through what type of interface will you interact with customers, websites, apps, etc.?</p>	<p>Is the information in the AI system also accessible to users using various types of technological platforms?</p>	<p>What is the level of technological maturity of the initiative?</p>
ACCELERATORS	<p>Is any business or technical data support required to consolidate the project?</p>		

1.2.1.1 Determination of the Technology Readiness Level (TRL)

Particularly for some financing rounds, it is advisable to determine the technology readiness level of the proposed development. Usually, international organizations such as the European Union use a conceptual framework with the following levels of technological development (European Commission 2014-2015):

Level 1
Observed and reported basic principles;
Level 2
Concept and technological application formulated;
Level 3
Experimental proof of concept;
Level 4
Validation of components and layout in a laboratory environment;
Level 5
Assessment of component and arrangement in a relevant environment;

1.2.2 Legal assessment

The use of AI and related technologies must usually comply with the legal provisions regarding the protection of personal data and consumer protection and some other sectoral regulations specific to the environment for which the AI system is developed (Galdon 2020).

Level 6
Prototype demonstration in a relevant environment;
Level 7
Demonstration of a system prototype in an operating environment;
Level 8
Complete system and certified through tests and demonstrations;
Level 9
System successfully tested in a real environment.



Moreover, depending on the level of impact of the AI solution on people's life, it should comply with regulations for the protection of vulnerable communities such as children, the disabled, and historically disadvantaged populations or at risk of exclusion (General Data Protection Regulation of the European Union 2016-2018). In this same sense, it should comply with international standards on human rights, especially

for those initiatives addressing social problems in sectors such as health, education, work, and economy. In particular, it is important to verify that the use of AI respects the right to non-discrimination since biased data could lead to the perpetuation of unequal treatment of population groups that traditionally suffer from unfavorable treatment.

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Does the system comply with the regulations on data protection and privacy, and consumer protection from its design?</p> <p>Are there specific sectoral regulations that apply to the system? Does the system comply with these regulations?</p> <p>Will the AI system have any kind of interaction with vulnerable populations that require special considerations?</p>	<p>Will the system have any relevant impact on the community's social, economic, or political life that requires the revision of particular regulations in this regard?</p>	<p>In case the AI system is being used to solve social problems, have the most relevant international standards regarding the protection of human rights been reviewed, such as the right to non-discrimination and other fundamental rights, mainly in the framework of the <i>Inter-American Convention on Human Rights, protocols and additional resolutions</i>?</p>
INVESTORS	<p>Has it been determined if the initiative complies with the applicable regulations and if the eventual breach of the company's legal obligations could cause the interruption of the business, sanctions, or fines that make it unviable or, in limited cases, in the region, transfer the responsibility to funding entities?</p> <p>Even if it complies with national regulations, does the company adhere to the rest of the countries' regulations where it has operations?</p>	<p>Is the information of the system available to access from different technological platforms?</p>	<p>What is the level of technological development of the solution?</p>
ACCELERATORS	<p>Has a legal feasibility analysis of the AI model been carried out?</p> <p>Has the company protected its intellectual property by registering a patent or corresponding as the case may be?</p> <p>If the operations of the proposed technological solutions are not contemplated in current legislation, have the legal mechanisms been structured to allow their operability? Such as permits and authorizations or risk mitigation measures.</p>		

1.2.3 Sectoral assessment

AI is a set of technologies that, due to its versatility, can be used in a wide variety of sectors and industries. This heterogeneity requires a very specific analysis of the dynamics that occur in each industry. However, this section

offers some minimum parameters to consider when identifying opportunities for the ethical and responsible use of AI (EC 2018; Omidyar 2019; Galdon 2020; MacCarthy 2020; WEF & IMDA 2020).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>What type of need does the AI system cover, one previously identified or one that had not been yet registered in the sector?</p> <p>In case the company has decided to build its data, has it implemented a methodology that allows knowing the systemic impact of the sector in which it will work?</p> <p>Is the available data representative of the sector's user base?</p> <p>Does the available data come from heterogeneous sources within the same industry, and are there enough data sources for AI systems based on machine learning?</p> <p>Do engineers and developers know the specific industry environment in which they will work in value chains and data systems?</p> <p>Has the importance of data quality been discussed with decision-makers for the impact on decision making?</p>	<p>Does the company have a dedicated AI governance group that conducts regular assessments of emerging public or policy concerns about using the AI system in the industry in which it operates?</p> <p>Have industry standards been adopted for the best use of data and technologies such as AI?</p> <p>Are there ongoing internal and external training and education methods on ethics, AI systems thinking, and Data Science?</p>	<p>Is there the collaboration of other key players in the sector for periodic reviews of the AI system's operation?</p> <p>Have data protection agreements or protocols been established with the rest of the partners and actors within the sector value chain?</p> <p>Have internal audits been carried out to assess the implementation of the best practices established in the early stages?</p> <p>Are measures of the impact of interventions carried out in the sector?</p>
INVESTORS	<p>How well do companies know the sector for which they plan to develop an AI solution?</p> <p>Does the product or solution solve a relevant problem for the sector?</p> <p>Have investors been communicated that the acquisition and structuring of quality data require consideration within the investment items?</p>	<p>Could the system's benefits be scalable to other areas of the value chain or even to other related sectors?</p>	

INCUBATORS

Do they have advisors, mentors, or tools to **communicate the importance of ethical considerations** in using AI and data science?

Is there a mechanism for **rapprochement between entrepreneurs and actors in the sector** that allows knowing the technological solution's impact?



1.3 SOCIAL AND ENVIRONMENTAL IMPACT

Depending on the type of functionality that the AI system has, there are times when it could have significant repercussions for the communities' social and environmental well-being in which it operates. Therefore, the organization that adopts or uses AI must consider measures that ensure diversity, non-discrimination, and equity in sectors such as education, health, environment, economy, and labor markets (EC 2018; Omidyar 2019).

Likewise, the AI system must take into special consideration its impact or influence on vulnerable populations, such as minors, people with disabilities, historically disadvantaged groups, and at risk of exclusion (General Data Protection Regulation of the European Union 2016 -2018).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Does the system respect fundamental principles of equity and non-discrimination in general and towards vulnerable populations?</p> <p>Is the AI system the best tool to generate a positive social impact, or are there other alternatives?</p> <p>Has the potential negative social impact of AI been analyzed to assess the severity of the errors? For example, through a <i>confusion matrix</i> that allows visualizing the performance of the algorithms.</p> <p>Has the scope of the social impact of the use of the AI system been assessed to establish proportional risk mitigation measures? E.g. the development of a user recommendation system would not have the same level of impact as that of a decision-making system for the selection of beneficiaries of a health or social assistance program.</p> <p>If the AI system is designed to be used by children, people with disabilities, or at risk of exclusion, what protection measures exist for these groups?</p>	<p>Does the AI system feed any decision-making process that has a social impact? If so, what measures are taken to ensure the principles of equity and non-discrimination?</p> <p>Is there a strategy or procedure to counteract biases from reality, such as discrimination and prejudice?</p> <p>Are there mechanisms to indicate problems related to bias and discrimination due to the use of AI?</p> <p>In addition to the considerations of the previous question, does the system have data correction mechanisms that allow knowing and correcting automated decision criteria?</p> <p>Derived from the previous point, are there mechanisms that allow users to present, process, and follow up on claims?</p> <p>Does the consent policy on the use of the AI system include provisions regarding the authorization of use to minors by their parents or guardians and disabled persons?</p>	<p>Does the solution generate some kind of power or information asymmetry? If so, are there mechanisms to promptly identify and correct these imbalances?</p> <p>Is there an evaluation and correction procedure established in the event of a negative social impact?</p> <p>Where appropriate, have mechanisms been established to measure the environmental impact of the AI system's deployment?</p> <p>Have measures been designed to reduce potential negative impacts on the AI system's use during the project cycle?</p>
INVESTORS	<p>Is there an approach and delimitation on a specific social problem to be solved for which AI is the most straightforward and appropriate technology to use?</p> <p>Has the community in which the solution will be deployed been involved or consulted on the development of the AI system?</p>	<p>Are there metrics to assess the social and economic impact of the deployment of the AI system?</p>	<p>Does the company carry out a periodic impact assessment on using the AI system around the specific problem it addresses, using methodologies such as Maslow's Pyramid of Needs?</p>

ACCELERATORS

Is the product commercially viable and **offering benefits to consumers, individuals, and communities** where it will be used?

Does the company **document all the possible risks of using the AI system** and accompany them with mitigation strategies?



2



GOVERNANCE AND SECURITY: INTERNAL GOVERNANCE STRUCTURES AND CONTROL OF AI

2

GOVERNANCE AND SECURITY: INTERNAL GOVERNANCE STRUCTURES AND CONTROL OF AI

Companies can create or update their governance frameworks and internal controls to ensure transparency and accountability of AI related to its development, deployment, and use. The selection of these models depends on the structure of the organization, size, available resources, and the particular sector in which it operates (Omidyar 2019; WEF & IMDA 2020).

This section offers some guiding questions to help analyze and identify the best corporate governance structures to ensure ethical and responsible AI adoption.



2.1 CORPORATE STRUCTURE FOR THE GOVERNANCE OF AI

The corporate structure of companies willing to adopt, deploy or use AI should have organizational arrangements that allow the identification, evaluation, documentation and resolution of fundamental tensions between its principles and the technical problems that result from the use of this technology (WEF & IMDA 2020). In this sense, it is necessary to assess what type of management model on AI is more appropriate, for example, between centralized or decentralized systems with a certain degree of human intervention or without it (Google 2018).

From a crosscutting perspective, it is recommended to create an ethics

council with executives and managers directly involved with the various areas of AI operation, such as management and the board of directors, the legal department, product development, quality assurance, human resources, purchases, and acquisitions, developers and daily operations, in order to resolve controversies resulting from the use of the AI systems (WEF & IMDA 2020). As a result, it is recommended that companies have direct communication structures with external public or sector supervision groups that allow the exchange of acceptable practices, the debate on problems or controversies, and the timely notification of emerging issues resulting in the use of such technology.



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has an officer within the company been appointed to conduct the ethical assessment of the AI solution?</p> <p>Has an assessment been made as to whether the company's corporate structure should be modified in any way to facilitate AI adoption, especially in case of accelerated growth of the company?</p> <p>Is the personnel in charge of the supervision, management, and control of the AI system adequately trained and aware of their role and responsibilities?</p> <p>Do the teams linked to the AI system's management have the necessary tools and capacities at their disposal to adequately fulfill their responsibilities?</p>	<p>Has the creation of an ethical review board been considered on the adoption of AI?</p> <p>Is the use of external guidance from a trusted source envisaged to monitor ethics and accountability issues and internal initiatives?</p> <p>Has the adaptation of risk management structures been considered for the incorporation of AI control processes?</p> <p>Is there a Code of Conduct on the use and operations of the AI system?</p>	<p>Are there institutional mechanisms for the coordination of the AI system with third parties, such as suppliers, consumers, distributors, and other workers?</p> <p>Has the creation of a collegiate body for the control of AI been considered, made up of representatives of departments that work directly with the AI system?</p> <p>Has a Code of Conduct on AI been developed to distribute to all company employees to be aware of the AI governance and control processes?</p>
INVESTORS	<p>Have the organization's corporate structure and the specific roles that executives have in terms of AI governance and control been shared?</p> <p>Does the company have a Chief Technology Officer, Chief Product Officer, and/or Chief Data Officer or Data Protection Officer? Have you been assigned specific responsibility for leading the ethical assessment of AI adoption?</p> <p>Does the company have a strong team of data scientists and engineers?</p>	<p>Is there a distinction between technical staff and business staff's responsibilities and involvement regarding their attributions concerning the AI system?</p>	
ACCELERATORS	<p>Are learning and education mechanisms provided for employee capacity development leading to the adoption and maintenance of AI?</p>		

2.2 RISK MANAGEMENT AND INTERNAL CONTROLS

AI governance systems must be based on a risk management approach allowing early identification of potential difficulties for the system and the mitigation of unintended damage that may have resulted from the misuse of AI. Therefore, it is suggested that risk management systems be based on privacy protection models by default (EC 2018; Galdon 2020; Omidyar 2019, Hoepman 2020), so that the personal data used in the security system AI is protected at all

times and that the user has control over their data.

Internal controls should also describe how data processing monitoring is carried out, including details on who processes, who controls, and with whom the data is shared.



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are there risk management mechanisms that can be expanded to include AI systems?</p> <p>In the case of making use of massive data and invasive technologies, among others, has the possibility of carrying out data protection impact assessments (<i>EIPD</i> or <i>PIA</i>) been considered?</p> <p>Does the AI system have a data management system in which the user can have a minimum management level over their personal information?</p> <p>Are security mechanisms in place for the AI system's critical elements throughout the entire operational period of the system?</p> <p>Does the company's board of directors monitor and approve risk management methodologies?</p>	<p>Have evaluations been carried out within the organization to identify risks to personnel or business strategies?</p> <p>Is there a periodic evaluation of possible risks and spaces for discussion to address the algorithms' concerns?</p>	<p>Has the board of directors of the company adopted any tool that allows identifying key risks according to the level of technological maturity of the company, such as those issued by the <i>National Institute of Standards and Technology (NIST)</i> or other international standardization bodies around cybersecurity?</p> <p>Are regular assessments and updates of risk management systems carried out?</p> <p>Is there a knowledge management strategy that allows monitoring and control over the processes of the AI system?</p>
INVESTORS	<p>Does the company have risk management mechanisms based on a default privacy policy?</p> <p>Does the company's board of directors regularly participate in assessments of the risk management system and internal controls to ensure the AI system's integrity?</p>		

2.2.1 Data security

Data security is one of the essential elements within the evaluations of an AI system since data is the input on which the system operates and the potential private content by users. In this way, it is important to assess the actions leading to the technical and organizational protection of the AI system against vulnerabilities and malicious agents that can alter data, cause failures in AI models, and its critical infrastructure in terms of

software and hardware. For an AI model to be considered safe, in most cases the developers of AI systems should explicitly consider those circumstances in which potential abuses by malicious agents could occur (EC 2018; Google 2018; Omidyar 2019; Méndez 2020; WEF & IMDA 2020).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Have vulnerabilities assessments of the AI system been conducted?</p> <p>Has the risk level of the AI system been considered for each specific use?</p> <p>Has it been considered to what extent the system can be used for purposes other than those established by the organization? E.g. there are numerous cases of AI systems used in electoral campaigns that have significantly impacted political and democratic systems.</p>	<p>Have acceptability thresholds and AI governance mechanisms been established that are actionable in the event of a possible eventuality?</p> <p>In tracking and user proximity models, decentralized protocols for data protection such as DP-3T or equivalent been considered?</p>	<p>Have the possible risks of using AI to third parties been assessed, including potential harm, characterization of the affected audience, and severity of harm?</p> <p>Does the organization have emergency protocols in place for potential cyber-attacks and data loss, among others, such as immediate technical changes or human intervention to reduce risks?</p> <p>Have cyberattacks simulations been carried out in which <i>human intervention mechanisms or total or partial disconnection of the system</i> have been tested?</p>
INVESTORS	<p>What kind of data security models are used, such as centralized or decentralized protocols?</p>	<p>Have the risks of possible fines or penalties for legal breach in data security that could seriously compromise the business's operations and profitability been assessed?</p>	<p>Is a coverage clause considered for any possible damage caused by the AI system?</p> <p>If necessary, has the company considered contracting coverage insurance against possible damages caused by the AI system?</p> <p>Does the project consider the responsibilities attributable to consumer protection regulation?</p>

3



HUMAN INVOLVEMENT IN AI SYSTEMS

3

HUMAN INVOLVEMENT IN AI SYSTEMS

One of the most relevant elements for the ethical and responsible use of AI is identifying the tasks that require supervision, intervention, or human interaction, either to correct inaccuracies in the system or to generate an adequate interaction between users and the IA model. (Google 2018; IMDA & PDPC 2020). Similarly, and from the position of a risk prevention approach, the appropriate degree of human involvement in AI systems should enhance human autonomy while minimizing possible conditioning, deception, or unjustified subordination to those who use this technology.

Below, a classification table of AI by type of use and interaction with human beings is shared, prepared by the Social Sector Division of the IDB, which allows

assessing the correspondence of human involvement according to the type of decision-making process:

	HUMAN INVOLVED	NO HUMANS INVOLVED
SPECIFIC OBJECTIVE	Assisted intelligence	Automated intelligence
ADAPTIVE SYSTEMS	Augmented intelligence	Autonomous intelligence

3.1 HUMAN-AI INTERACTION DETERMINING THE LEVEL OF HUMAN SUPERVISION

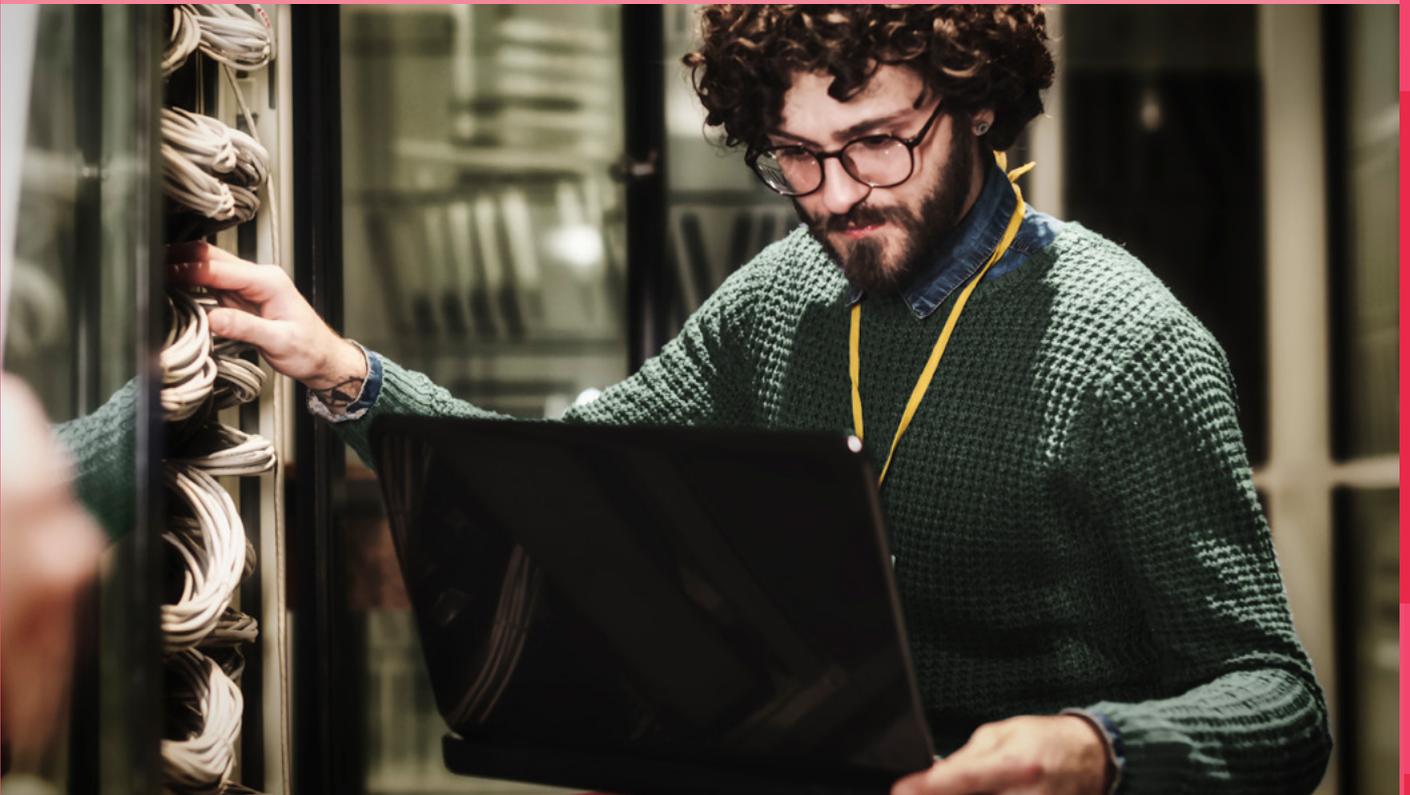
The determination of the level of human involvement in the processes, results, and decision-making of the artificial intelligence system will largely depend on the impact that such processes have on users. The most common example used to get an idea of the gradual impact of AI systems in its environment is the comparison between an AI system that is limited to issuing purchase recommendations to users based on their behavior and preferences versus an

AI system that has implications for road safety of users on autonomous vehicles (MGI 2017; EC 2018). It is evident that both systems require different degrees of supervision and human intervention to prevent or correct possible adverse expected or unexpected results of the system (Google 2018).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has the use of tools or models been considered to determine the level of human involvement based on the system's impact on the lives of users? For example, through methodologies such as <i>human in/out/over the loop</i>.</p> <p>Has the operating cost of the different levels of human involvement in the AI system been considered when assigning specialized personnel dedicated to these tasks?</p> <p>Depending on the degree of impact of the system on users' lives, does the system have an immediate stop option for human intervention?</p>	<p>Has the appropriate level of human control been selected following the relevant AI system and use cases?</p> <p>Do the activities or processes in which AI will operate require recognizing emotional nuances, or does it merely involve repetitive activities?</p> <p>Has the development of Key Performance Indicators (KPIs) of the AI system been considered to help determine human involvement's relevance in case the system deteriorates?</p>	<p>In the case of autonomous AI and machine learning mechanisms, are there more specific control mechanism on already identified problems?</p> <p>Has the development of emergency responses based on the analysis of scenarios been considered if risk mitigation measures fail?</p> <p>In case the AI system can understand or infer the emotional status of its interlocutors, what measures have been taken to avoid problems of empathy or attachment of the users?</p>

4



AI LIFECYCLE: AI SYSTEM OPERATIONS MANAGEMENT

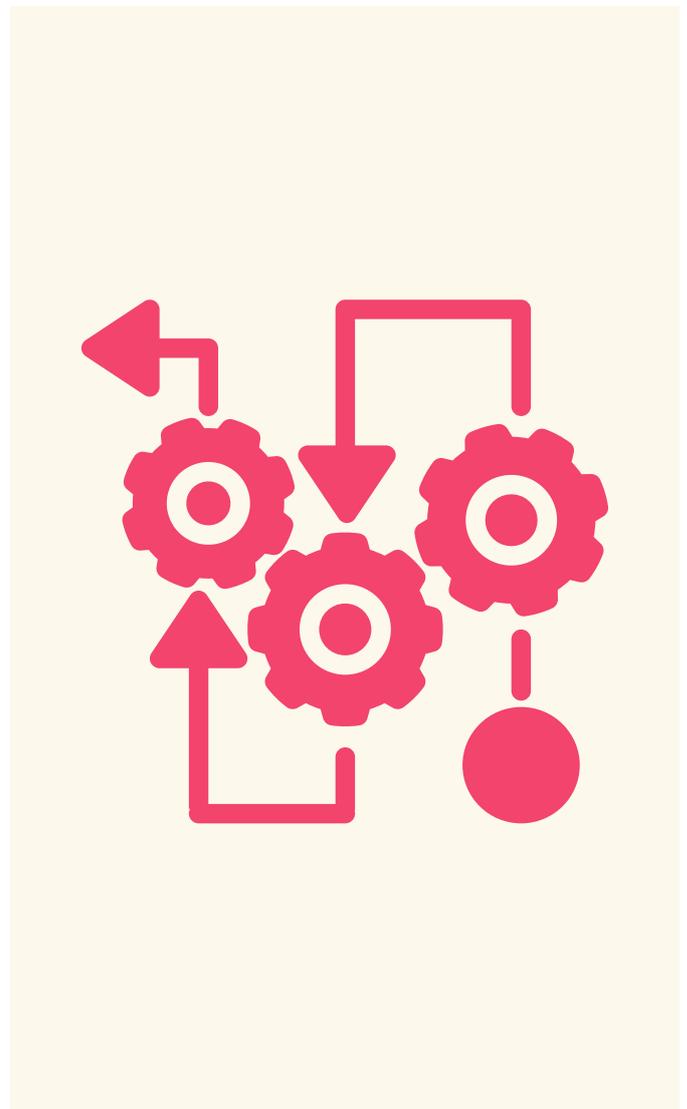
4

AI LIFECYCLE: AI SYSTEM OPERATIONS MANAGEMENT

This section is focused on the analysis of the operations of the AI system from the collection and structuring of data, based on a privacy perspective by design or by default, to the processing of this data through AI algorithms.

4.1 DATA SOURCE AND MANAGEMENT

The basic input of AI systems is data. Most machine learning-based systems (as opposed to rule-based systems) require a large amount of high-quality structured and historical data to refine their behavior and results in predicting and making automatic decisions. In this sense, the quality of the data is of great importance for the responsible use of AI, as it is a way to guarantee the correct operation of the system so that it does not have biases or problems that affect users (Posey 2019; WEF & IMDA 220). On the other hand, much of the data used by AI systems usually contain personal information or metadata related to its users' behavior, so the protection of this data is another critical element for the ethical and responsible use of AI.



4.1.1 Data type

Before the deployment or adoption of an AI system, an assessment should be made of the type of data required for the correct operation of the system and the type of data available to feed it. Data is currently an element that has acquired an increasing commercial value so that its exchange requires careful treatment not to violate various aspects of user privacy (MGI 2017). On the other hand, in recent years, a dynamic of competitiveness has been generated by data that users

tend to protect more, such as their geolocation and opinions or sentiment around different products, services, or topics, so that there is an awareness growing on the care of these elements that should be considered by those companies interested in the use of this type of data.



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has it been precisely assessed what type of data the AI system will require, including its availability, access to sources and necessary formats?</p> <p>Are there measures to trace the origin and destination of the data used by the AI system?</p> <p>Is the data complete, up-to-date and from reliable sources?</p> <p>Has the use of data extraction methods such as <i>ETL (extract, transform, load)</i> been considered to ensure data and format compatibility?</p>	<p>Have quality control evaluations been conducted on the use of data from external sources?</p> <p>Are there inventories or data repositories, taxonomies, and documentation of control mechanisms over the data based on the principles of <i>data governance of standardization agencies such as ISO or IEEE</i>?</p> <p>Are there monitoring mechanisms that allow assessing whether changes in the source of origin generate substantial changes in the AI system's results?</p>	<p>Does the AI system use specialized software tools for data storage, retrieval, or transformation to ensure consistency? Some examples are: <i>NoSQL or the MapReduce model used by Apache Hadoop</i>.</p> <p>Is there a team dedicated to managing data policy?</p> <p>Does the company have the ability to verify the validity and representativeness of the data obtained through third parties (<i>including verification strategies through metadata</i>)?</p>
INVESTORS	<p>What is the source of the AI system's data, and what is the data strategy?</p> <p>Does the company have its data-generating source (single database), or does it depend on third parties to obtain it?</p>	<p>Are there association agreements with other companies to guarantee the maintenance of unique databases not accessible to competitors?</p>	<p>Does the company use historical data, static data, or real-time data?</p> <p>Does the company have the capacity to process training data to make its algorithms more efficient?</p>
ACCELERATORS	<p>Are technical experts available to guide data architecture, data collection, storage, and transformation?</p> <p>Has it been assessed which are the legal instruments that allow processing the data for the AI system?</p>		

4.1.2 Data processing

Data processing refers to collecting, storing, organizing, structuring, altering, consulting, using, transmitting, disseminating, and any other operation executed over this data (Omidyar 2019; WEF & IMDA 2020). It should be remembered that various international instruments prohibit, in principle, the processing of data that reveals the ethnic or racial origin of people, political opinions, religious or philosophical beliefs, affiliation or membership of labor unions, genetic and biometric data, sexual preference, or orientation, as well as a state of health, except if conditions are met in some instances, such as, among others, the explicit consent of the

user (EC 2018; General Data Protection Regulation of the European Union 2016-2018). Therefore, it is recommended that the appropriate instruments and regulations be consulted according to international and national regulations. The guiding questions presented below are mostly based on the provisions contained in the General Data Protection Regulation of the European Union, as one of the most advanced instruments in this matter. This instrument could offer inputs for the elaboration of further regulations in various countries.

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Is the data collection process part of data minimization protocols that collect only personal information from users only when necessary and for a limited time under secure storage conditions?</p> <p>Have protocols been established to process data concerning storage, disclosure, and transformation of processing and its treatment by third parties?</p> <p>Has an informed and disclosed consent policy been established on the use of users' data?</p>	<p>In the event of manual data editing, are there sufficient protocols to ensure the quality, auditability, and traceability of the modified data?</p> <p>Are there provisions to safeguard the privacy of personal information used by the AI system, such as aggregation and anonymity of data?</p> <p>Have the regulations around security policies been considered for limiting personal data storage, including protection against unauthorized use, accidental loss, destruction, or data damage?</p>	<p>What steps does the organization take to mitigate bias from pattern reinforcement?</p> <p>Has the development of an AI model contrary to the original system been considered for testing it and assessing the reliability of the results?</p> <p>Does the organization fragment large databases to mitigate risks and validate the AI model from multiple perspectives?</p>

4.1.3 Data integrity and confidentiality

Data integrity assessment is closely linked to data quality throughout the chain of use within the AI system, as the introduction of incomplete or poorly structured data could have significant repercussions on system performance. The various data sets from which the system is fed should be regularly tested and well documented whether the data was developed in-house or acquired through a third party (General Data

Protection Regulation of the European Union 2016-2018; EC 2018). Besides using the questions for guidance on this topic, it is recommended that you review the OECD privacy principles and consult the respective ISO and IEEE standards.

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are procedures in place to ensure data quality and integrity?</p> <p>Are the possible limitations caused by the composition of the data used recognized?</p> <p>Are wide margins of diversity and representativeness of data considered for users?</p> <p>Is there a possibility that the user chooses to delete their history and registry from the system?</p>	<p>Has the implementation of tests that assess the integrity of the data been considered? For example: <i>Database Integrity Testing by service, area, technology, or discipline.</i></p> <p>How does the company verify that the databases maintain a consistent state or are not accessed by third parties improperly?</p> <p>Has the diversity and representativeness of the data been tested concerning particular populations that have reported problematic use cases?</p>	<p>Are research methods and technical tools used to improve understanding of the data, its dynamism, the AI model, and performance?</p> <p>Has independent agency certification been considered for data privacy and security?</p>

4.1.4 Privacy by design

Access protocols should regulate the handling of users' data to use AI systems to this data that determine precisely who and under what circumstances they can access said information. The company

personnel that has access to it must be duly accredited and qualified for its treatment (KPMG, 2018 Galdon 2020, Hoepman 2020).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>In case of having databases that include the users' personal information, what mechanisms to protect this information exist?</p> <p>Has the use of AI models been considered to minimize dependency on the use of personal data?</p> <p>Have consent mechanisms been established to use personal data and the possibility of withdrawing the data if requested?</p> <p>Has it been considered in the data processing and archiving design that personal data is not retained for longer than the time necessary for its use?</p> <p>Are personal data protection impact assessments (PIA: Privacy Impact Assessment) carried out on products and services linked to the AI system, including personnel?</p>	<p>Have data protection measures been taken, such as encryption, anonymization, and aggregation?</p> <p>Are data protection protocols from international standards such as ISO or IEEE standards followed?</p> <p>Is there a record of access to personal data? For example: who accesses, when, from what location.</p>	<p>In the event of the existence of personnel assigned to the supervision of personal data, has this officer been involved in the design of the system from the beginning?</p> <p>Are there mechanisms for disclosing the data protection policy that ensure its knowledge by company employees and third parties?</p>
INVESTORS	<p>Entities receiving financing or investment should know and follow the data privacy policies of their counterparts.</p>		

4.1.5 Interoperability

Interoperability refers to a system’s ability to make processes and data sources compatible with other providers or customers (Galdon 2020), through the use of similar formats that allow the correct functioning of AI systems and similar syntax between various databases. However, on some occasions, this

dimension is usually omitted. It is assumed that the data from other institutions are in compatible formats and with content structured in the same way. However, this is often a great challenge for many systems’ operations, mostly syntactic compatibility (Potgieter 2018).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has a mapping been carried out of the sources with which it is required to integrate the AI system, including the type of formats used and the readable structure of the data?</p> <p>Who are the counterparts (public or private sector) with whom the operational integration will be carried out?</p> <p>Is the conclusion of contracts required for operational integration?</p> <p>Has the collection of data been considered through providers that comply with data protection standards?</p>	<p>Are details such as the physical location of servers and databases known?</p> <p>What kind of process redesign is required to integrate external databases into the workflow?</p> <p>Is there a possibility for the user to obtain his data record to port it and use it with another AI system?</p>	<p>Have specific measures been adopted to make the syntax and semantics of the data sources compatible with the particular requirements of each AI system? For example, in the health sector through resources similar to the <i>Fast Healthcare Interoperability Resources (FHIR)</i> in the United States for the health sector.</p>
INVESTORS	<p>Is there a prior history of the company’s operational integration with the proposed counterparties, including contacts, contracts, or some other link?</p>	<p>Is the interoperability relationship between the company and its counterparts sustainable?</p>	
ACCELERATORS	<p>What analogous interoperability initiatives can be developed and used? For example, initiatives such as smart cities or data lakes.</p>		

4.1.6 Testing and validation

The testing and validation processes allow identifying and resolving possible errors in the AI system within a stable environment. These trials require accurate

databases that are as close to reality as possible to guarantee the correct functioning of the AI system (Galdon 2020; Omidyar 2019).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>During the trials' design, are the basic parameters for the validations that allow us to know the degree of reliability of the AI system determined?</p> <p>Are trials and validations performed for each of the subgroups of the target population of the AI system?</p> <p>Do the tests carried out correspond to the operational expectations in the environment in which the AI system will be used?</p>	<p>Are there different databases to carry out differentiated algorithm training, testing, and validation activities?</p> <p>Are reproducibility tests carried out in scale production environments?</p> <p>Has the creation of a test template been considered that can have user feedback?</p>	<p>Are simulations carried out on collecting data from different sources to ensure data quality and verify its real world validity?</p> <p>Are tests carried out with methods such as A/B on different versions of the AI model to verify its reliability?</p>
INVESTORS	<p>Are trial results shared with funding entities, investors or donors?</p>		

4.2 MODEL DEVELOPMENT: ALGORITHMS

Algorithms are the structure of AI systems, as they are in charge of processing data to perform various functions, including predictions, clustering, categorization, and automated decision-making (MGI 2017; Google 2018; IMDA & IDPC 2020). There are several types of algorithms based on rules or learning through big data. Their classification may also include the level of supervision or human instruction they require. Due to their conceptual complexity and high level of mathematical abstraction, algorithms are usually explained by the type of result they generate, rather than by how their internal processes act (Google 2018; IMDA & IDPC 2020).

Based on the previous, building trust around AI necessarily requires scrutiny of the different algorithm models used and their results. This section is a thematic

guide on the most relevant topics around the algorithms' ethical and responsible use.

4.2.1 Traceability

The traceability of the algorithms used in artificial intelligence refers to the documentation of the AI system's inputs and processes leading to certain outcomes. This record is important, as it is one of the elements that contribute to transparency and confidence-building around AI adoption. Besides, it allows to accurately identify any problem that may arise throughout complex operations from data collection, structuring, processing, interpretation, and the result (EC 2018; Google 2018; WEF & IMDA 2020).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are there ways to document the following element: <i>System objectives, data, research approach, description of the algorithm, and performance evaluation parameters, errors, and notes from the technical team?</i></p> <p>Are there records on the process of obtaining data from open sources?</p> <p>Have documentation and monitoring methods been implemented for AI processes based on pre-established rules or self-learning?</p>	<p>For AI systems based on preset rules, has the model building process been documented?</p> <p>For AI systems based on self-learning, is there a data source tracking system that explains the selection process and how it happened?</p>	<p>Regarding the methods of testing and validation of algorithms: Does the company have information about the data used for testing and validation for systems based on pre-established rules?</p> <p>Does the company have data tracing from various sources and the results obtained from multiple simulations for systems based on machine learning?</p>

4.2.2 Explicability of AI processes and results

Explicability is the ability to clarify the technical processes of an AI system and the decisions that result from the use of the system. An effective explanation is one that is understandable to the different audiences to whom it is exposed. As AI systems have a greater impact on people's lives, the greater their obligation to explain the processes

and results of using AI. Likewise, the concept of explicability has become part of regulatory obligations that have arisen from legal instruments such as the General Data Protection Regulation of the EU, and that has been incorporated into various laws in Latin America (Google 2018; Dassen et al. 2019; Omidyar 2019; Molnar 2020; WEF & IMDA 2020).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>To what extent can the processes and results of the AI system be easily understood?</p> <p>Can the company explain why the system made a certain prediction or decision so that all users can understand it?</p> <p>Has the AI model been designed considering different interpretation factors including language varieties?</p>	<p>Suppose the results of the AI system's operation have an impact on the company's decision-making. Is it communicated on time to the rest of the employees and partners affected by the decisions?</p> <p>Are there different models of explanation corresponding to the various features or actions of the AI model?</p> <p>Has the construction of simple explanatory narratives been considered through measurement parameters of the type: <i>high /medium/low, percentages, or other tools</i> that help understand the system?</p>	<p>Is there a document that explains to the public how the AI system works, including performance and maintenance measurement parameters?</p> <p>Is there a simpler model of the AI system created to provide didactic explanations?</p> <p>Has the use of specialized techniques been considered to explain the results of AI models such as <i>LIME, SHAP, LOCO, PDP, ICE?</i></p> <p>Does the company have the supplementary infrastructure to automatically and proactively offer explanations to users according to the different predictions or decision-making results generated by the AI system?</p>
INVESTORS	<p>What is the company's business model, and how does it generate value through the use of AI?</p>	<p>Has it been assessed to what extent the AI model's interpretability can be examined, and the model's work processes accessed to understand them better?</p>	
ACCELERATORS	<p>Is there an obligation to explain the processes and results of the AI system? And does the company have the mechanisms that allow these explanations?</p>		

4.2.3 Replicability

Replicability consists of the repeated execution of an operation that generates the same result consistently. For this reason, some AI models substitute certain measures of explicability by demonstrating the consistency in the results of a given model through controlled repetitions. This occurs when the AI model processes are usually very complex or abstract in such a way that understanding the system is simpler and more useful to communicate the degree of reliability of the system (EC 2018;

Google 2018; Omidyar 2019; WEF & IMDA 2020).

Furthermore, sometimes it is convenient to carry out replicability tests to assess in a counterfactual way how “fair” a system is concerning different population groups, avoiding the perpetuation of biases, especially in sectors such as insurance, loans, hiring, and law enforcement predictively (Kusner et al. 2017).

	STAGE I	STAGE II	STAGE III
COMPANIES	Is there a need to carry out replicability tests and counterfactual tests to ensure that certain population groups will not be negatively affected by the AI system’s automated decisions?	Using replicability measures, have exceptional results been identified , and an assessment made of how to address them? In the case of replicability tests, are there updated data for each test?	Are comparative replicability tests carried out between different versions of the same AI model?



4.2.4 Reproducibility

The concept of reproducibility refers to the AI model’s ability to be reproduced in another environment external to that of the organization that created it, unlike the concept of replicability, which refers to the possibility of obtaining the same

results within of the same environment. By definition, this concept requires two environments in which tests are carried out to verify the model (Google 2018; Omidyar 2019; WEF & IMDA 2020).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has it been determined what conditions are necessary to guarantee the reproducibility of the system?</p> <p>Has any external organization been contacted with whom AI system reproducibility activities could be carried out?</p>	<p>Has a dedicated team been engaged to review whether similar results can be produced using the same AI models in different settings?</p> <p>Have the processes for testing and verifying the reliability of AI systems been documented and operationalized?</p>	<p>Has the development of replicability files been considered for behavioral testing phases of the AI system?</p> <p>Have communication mechanisms been established to let users know the degree of reliability of the AI system based on the reproducibility tests’ results?</p>
INVESTORS	<p>Is the company the developer of its own AI system, or does it depend on an external developer to replicate and test results?</p>		

4.2.5 Auditability

The auditability of AI systems is usually reserved for particular cases in which a third party assesses the algorithms' processes and results. In this sense, the AI models that could have a greater propensity to be audited are those that could have an impact on human rights or the safety of people. Although it is a concept whose application is limited,

it is convenient to consider some preparatory aspects that facilitate the process of collecting and structuring information in case an audit is required (General Data Protection Regulation of the European Union 2016-2018).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Have mechanisms been established to facilitate the auditability of the system? For example, through the <i>documentation and traceability of processes and results</i>.</p> <p>Have you considered how commercially sensitive information should be shared in the event of an audit?</p>	<p>In cases where the AI system has implications for user safety, could it be independently audited?</p>	<p>Does the company have certifications from international standardizing agencies, which facilitate documentation delivery to auditing bodies?</p>
INVESTORS	<p>Is the company responsible for data processing, facilitating the traceability of data flows?</p> <p>Is there a periodic report on the application of governance codes to investors?</p>	<p>Are the company's business practices such as transactions, acquisition of licenses, or other agreements aligned with the ethical principles of data processing?</p>	

4.2.6 Maintenance

AI systems require periodic maintenance to have up-to-date data

to allow efficient operation (EC 2018; Galdon 2020; WEF & IMDA 2020).

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are periodic reviews and updates of the databases carried out to ensure their accuracy, validity, and relevance?</p> <p>Have specialized personnel been dedicated to the periodic review and update of the data and algorithms?</p>	<p>Has the use of automatic notification systems been contemplated on the availability of new data relevant to the AI system?</p> <p>Have systems been verified to behave unexpectedly in certain contexts?</p>	<p>Are various alternative models of the AI system used to validate its operability concerning the different elements or types of processed data to mitigate the system's bias?</p>



5



RELATIONSHIP WITH KEY STAKEHOLDERS

5

RELATIONSHIP WITH KEY STAKEHOLDERS

Stakeholder engagement refers broadly to the involvement of relevant stakeholders throughout the AI lifecycle, which includes communications, collaboration, consultation, and accountability with information owners, users, clients, investors, partners, suppliers, regulators, and civil society in general (EC 2018; Omidyar 2019; WEF & IMDA 2020).



5.1 TRANSPARENCY

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Has the organization contemplated and designed communication pieces aimed at the different actors involved with the AI system or affected in some way by it?</p> <p>Has the organization assessed and designed communication pieces based on scenarios in which an explanation of the AI system's operation or results would be necessary?</p>	<p>Has the creation of an explanation strategy been considered based on data, models, human elements, interferences, impact, etc.?</p> <p>Has the progressivity of using the AI system by different audiences been taken into consideration to avoid communications fatigue?</p> <p>Does qualified personnel supervise the communication and feedback channels?</p>	<p>Has various tools and audiovisual media been considered to improve the communication strategy on the AI system?</p> <p>In those moments in which the explanation of the AI system's operations is not useful, has the use of counterfactual explanation techniques based on replicability tests been considered?</p>
INVESTORS	<p>Does the company have a policy of accountability and transparency or similar to its clients and partners?</p>		<p>Does the company efficiently communicate the results of impact evaluations on protecting its clients' data, if necessary? Especially if the AI system carries out automated decision-making tasks that have an impact on customers</p>

5.2 CONSUMERS AND USERS OF THE SYSTEM

	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are users notified that they are interacting with an AI system, and what implications does it have, before starting the interaction, for example, with chatbots?</p> <p>Are users informed about the reason for using this technology, and what specific functionalities are linked to AI?</p> <p>Is the expectation and extent of the behavior of the AI system communicated to the user?</p> <p>Is there a possibility that the user chooses to exit the system and delete all their data?</p>	<p>In addition to text-based communication tools, are there infographics and other multimedia tools that allow the public to understand how the AI system works?</p> <p>For some instances: Once a result or a decision derived from the AI system is generated, are factors that led to that result explained to the user? Especially in cases of granting loans, government aid, or insurance coverage.</p> <p>Are there means of communication through which the user can request a review of the result of the AI system that has affected them in any way?</p>	<p>Is there an explanation policy that includes the various cases in which additional information must be offered to users?</p> <p>Are there pedagogical tools and workshops that allow clients to have a better level of understanding and interaction with the company about the behavior of the AI system?</p> <p>In using chatbots for customer relations, is there a method of recording the conversation, for purposes of improvement and learning only?</p>

5.3 ACCESS TO FINANCING

Access to financing for projects related to AI is the result of a match between the capital fund interest and the start-up. It is a process of building trust and mutual selection based on the exchange of information to reduce the knowledge gap between both parties about AI applications and the return on investment. It is crucial to have accurate communication about the scope and limitations of the AI system, and even tools that are not properly AI, but which manage large amounts of data.

Similarly, teams looking for funding rounds must include a balance of business professionals, industry experts, and data scientists who can offer a complete and robust view of the solution being presented (EC 2018; Gonfalonieri, 2019).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Is the use of AI being promoted as the best tool to solve the problem posed?</p> <p>Is there a simple and straightforward narrative that links the identification of problems and the solutions that the AI system will offer with the corporate strategy to identify <i>niche markets, competitors, and SWOT</i>?</p> <p>Is the type of AI system or proprietary technology used being communicated accurately and transparently, even if it is just a large data processing system? <i>Automated tasks and data analysis are sometimes confused with the aspiration to use these tools to carry out autonomous and intelligent predictions or decision-making as AI systems do.</i></p> <p>Does the company's pitch focus on showing the added value and ease of implementing the solution beyond focusing on using a specific AI model?</p> <p>Does the company have a diverse and balanced team of professionals in business, data science, marketing, distribution, sales, and industry specialists?</p> <p>What type of guarantee clauses does the fund offer to protect your investment and is it feasible/acceptable to comply with them?</p>	<p>Has the AI system been piloted with incipient indicators that show its ability to be deployed in the market?</p> <p>Which key performance indicators (KPIs) are in place to communicate the robustness of the AI model?</p> <p>What is the relationship strategy with strategic partners within the value chain?</p> <p>Are elements communicated that show the possibility of scalability of the business model, including "go to market" strategies?</p>	<p>Does the company have indicators that allow transmitting the interest and validation of clients that justify the <i>return on software investment as a service (SaaS)</i> versus other digital solutions?</p>

INVESTORS

In case of not having an actual AI system, but a big data management system that in the short or medium term might evolve into an AI model, **what is the progressive level of human intervention?**

What **track record does the company** have in using AI for real-world solutions and what have been their returns on investment?

How **dependent are the solutions on obtaining intellectual property rights on AI data and models if you don't have your own technology?**

Does the company have the **ability to become a segment leader** and dominate its market?

For products aimed at consumers/user, does the AI system uses the data provided by the user to **create greater personalization and link with the service?**

What is the **level of social impact** that the company has had by using its AI solution, and **how sustainable is its return on investment?**

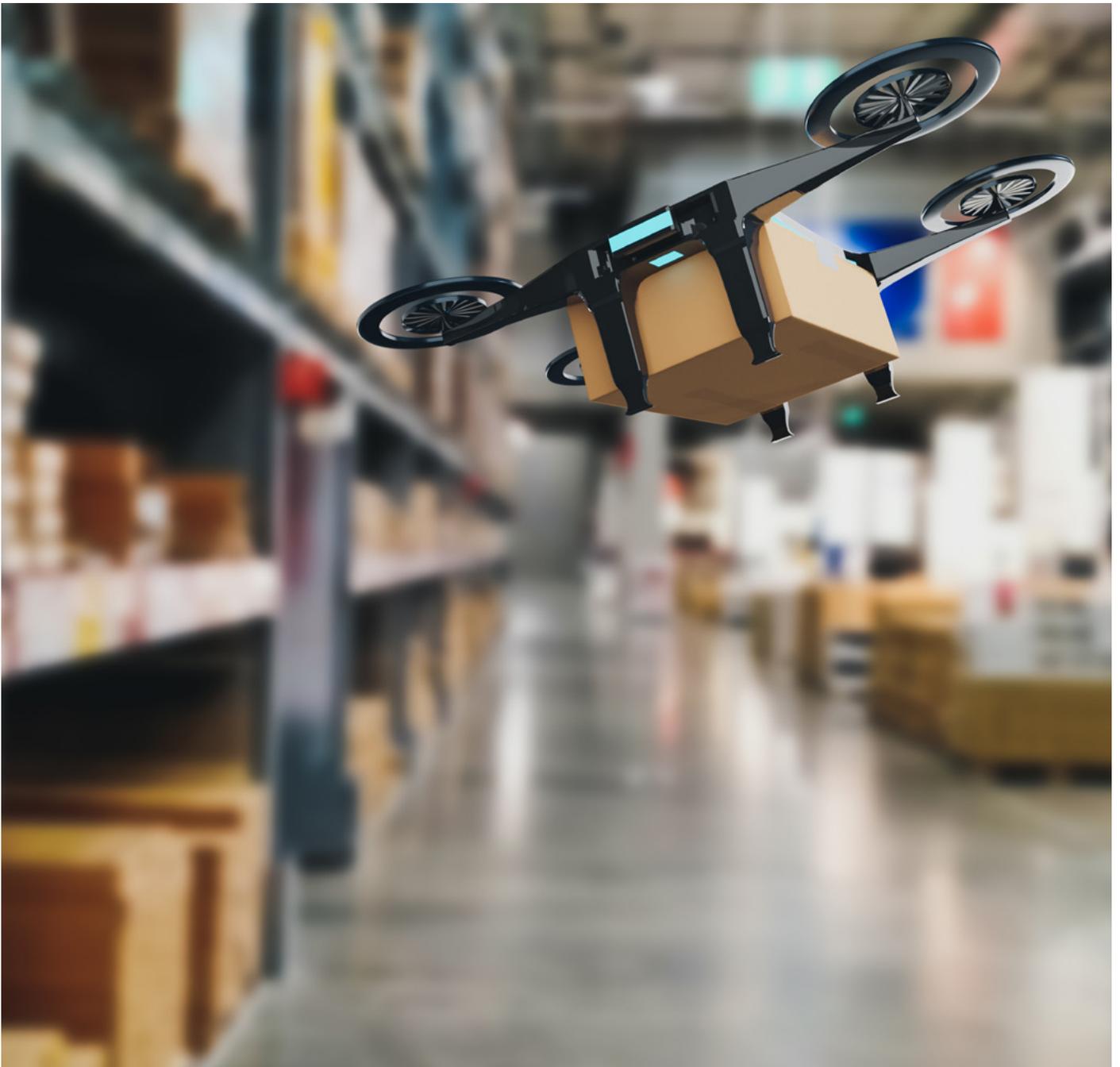
6



COMMUNICATIONS

6 COMMUNICATIONS

Companies that make use of AI systems should carry out proactive, transparent, and systematic communications to stakeholders about the capabilities and limitations of AI systems. In this sense, it is important to manage expectations about the technological scope of the AI solution (Omidyar 2019; Engler 2020; WEF & IMDA 2020).



	STAGE I	STAGE II	STAGE III
COMPANIES	<p>Are the terms of service offered by the AI system communicated?</p> <p>Are there appropriate communications channels to explain the user it is interacting with an AI system?</p> <p>Have alternatives for interaction with human beings been offered?</p> <p>Have mechanisms for receiving feedback from end-users of the system been established?</p> <p>Have communication channels, procedures, and contact points been shared to resolve problems and questions about the AI system?</p>	<p>Have interested parties, including shareholders and users, been communicated about previously identified or perceived biases or errors?</p> <p>Has any mechanism been considered to involve stakeholders in the development or feedback of the AI system?</p> <p>Has the adoption of AI been communicated within the company, including possibly affected workers and representatives?</p>	<p>Have various scenarios been considered on which certain critical operations of the AI system should be communicated?</p> <p>Has the possibility that explanations run the risk of generating confusion, confirmation bias, or cognitive fatigue of audiences has been considered?</p> <p>Has a taxonomy or glossary been developed to clarify concepts such as “fairness” regarding the AI system?</p> <p>Is there a metric or parameter for determining concepts such as “fairness” in the use of the AI system?</p>
ACCELERATORS	<p>Is it possible or necessary to determine the degree of understanding of the counterparts on the AI system’s terms of service?</p>		

ANNEX 1

ANNEX 1 - DESCRIPTION OF THE STAGES OF BUSINESS DEVELOPMENT

To cover the broadest possible spectrum of entrepreneurs and sectors that make use or intend to incorporate artificial intelligence into their business, this matrix develops three stages of development for the self-assessment of the ethical and responsible use of AI. These stages give the self-assessment its progressive condition since the evaluations change according to the level of development of the enterprise (Churchill and Lewis 1983; Dibner 2018; Areitio 2019), the level of TRL (European Commission 2014-2015), and the participation of different actors in the various stages.

The main component determining the stages is the level of development of the ventures, including the seed and early stages in stage I, start-ups in series A and B in stage II, and start-ups in series C and corporations in stage III. In turn, this usually corresponds to the level of technological development that enterprises have reached, according to the nine TRL levels of the European Commission, categorized into three levels: research, development, and

deployment. Finally, it should be noted that each stage entails a different interaction between actors; thus, incubators usually operate in stages I and II, while accelerators could take a more active role in stage III.

DETERMINANTS OF STAGES	STAGE I	STAGE II	STAGE III
<p>Development level of the start up <i>* Based on the conceptualization of Harvard Business Review, Medium, Mattermark, Crunchbase, The Venture City.</i></p>	<p>Seed/early stage</p> <p>Product ideation and development.</p> <p>Arrival to the first users and market discovery.</p> <p>First hires.</p> <p>Preparing for Series A</p> <p>Alistamiento Serie A</p>	<p>Later stage</p> <p>Startups in Series A and B + Bridge rounds.</p> <p>Series A Business model definition.</p> <p>Start of product or service distribution on a larger scale.</p> <p>Series B Commercial traction with clients and well-defined business model.</p> <p>Scale business model, user base and acquisitions.</p>	<p>Mature</p> <p>C series startups and corporations.</p> <p>Series C Acquisition of greater capital and accelerated growth.</p> <p>Identification of profitability, but with the need for more capital.</p> <p>Internationalization.</p> <p>More acquisitions and hiring.</p> <p>Corporations Greater consolidation in the market and profitability.</p>
<p>Tech Readiness Level (TRL) <i>*Based on Tech Readiness Levels developed by the European Commission 2014-2015.</i></p>	<p>Research</p> <p>Level 1 - Observed and reported basic principles.</p> <p>Level 2 - Concept and technological application formulated.</p> <p>Level 3 - Experimental proof of concept.</p>	<p>Development</p> <p>Level 4 - Validation of components and layout in a laboratory environment.</p> <p>Level 5 - Assessment of component and arrangement in a relevant environment.</p> <p>Level 6 - Prototype demonstration in a relevant environment.</p>	<p>Deployment</p> <p>Level 7 - Demonstration of a system prototype in an operating environment.</p> <p>Level 8 - Complete system and certified through tests and demonstrations.</p> <p>Level 9 - System successfully tested in a real environment.</p>
<p>Engaged Actors</p>	<p>Entrepreneur</p> <p>Investor</p> <p>Incubator</p>	<p>Entrepreneur</p> <p>Investor</p> <p>Incubator/Accelerator</p>	<p>Entrepreneur</p> <p>Investor</p>

SOURCES

Areitio, Andy (2019). *From idea to scale up: Phases of a start up.* Medium. The Venture City. <https://medium.com/theventurecity/fases-de-la-startup-de-idea-a-scale-up-73c8e81cd37>

Inter-American Development Bank (IDB) (2020). *Ethical and responsible adoption of Artificial Intelligence in Latin America and the Caribbean.*

Bartram, Robert (2018). *The New Frontier for Artificial Intelligence.* ISO Committee on Artificial Intelligence. *ISO/IEC JTC 1/SC 42.* <https://www.iso.org/news/ref2336.html>

Berkman Klein Center for Internet & Society at Harvard University (2020). *Principled Artificial Intelligence. A Map of Ethical and Rights-Based Approaches to Principles for AI.*

Cabrol, Marcelo; Baeza-Yates, Ricardo; González Alarcón, Natalia; Pombo, Cristina (2020). *Is data privacy the price we must pay to survive a pandemic?* Inter-American Development Bank. <http://dx.doi.org/10.18235/0002292>. <https://publications.iadb.org/es/es-la-privacidad-de-los-datos-el-precio-que-debemos-pagar-par-a-sobrevivir-una-pandemia>

Churchill, Neil C. & Lewis, Virginia L. (1983). *The Five Stages of Small Business Growth.* Harvard Business Review. May 1983 Issue. <https://hbr.org/1983/05/the-five-stages-of-small-business-growth>

European Commission (2018). *Ethical Guidelines for a Reliable AI.* Independent Group of High Level Experts on Artificial Intelligence.

Dassen, Thommy; Hou, Naiwen; Kronseder, Veronika (2019). Chapter 2. *Introduction to Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE).* https://compstat-lmu.github.io/iml_methods/limitations/pdp.html

Dibner, Gil (2018). *There are only three startup stages.* Medium. Angular Ventures. <https://medium.com/angularventures/there-are-only-three-stages-for-startups-b8783d6b0f1>

Engler, Alex (2020). *The Case for AI Transparency Requirements.* Brookings Institution. <https://www.brookings.edu/research/the-case-for-ai-transparency-requirements/>

European Commission (EC) (2014-2015). *Technology Readiness Levels.* Horizon 2020 Work Programme 2014-2015. Annex. https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp-1415-annex-g-trl_en.pdf

_____ (2018). *Trustworthy AI Assessment List.* Independent High Level Expert Group on Artificial Intelligence.

European Parliament Research Service (2019). *A Governance Framework for Algorithmic Accountability and Transparency.* [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU\(2019\)624262_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf)

European Union (2019). *Annual Report 2019.* European Data Protection Supervisor.

Everis & Endeavor (2018). *The impact of Artificial Intelligence on entrepreneurship.* <http://www.endeavor.cl/wp-content/uploads/El-impacto-de-la-IA-en-el-emprendimiento-en-Am%C3%A9rica-Latina-everis-y-Endeavor.pdf>

Galdon, Gemma (2020). *Ethical Evaluations Protocol.* Ethics Consulting.

Gigler, Bjorn-Soren (2020). *Closing The Investment Gap For Artificial Intelligence And Blockchain In Europe.* Panodyssey. <https://panodyssey.com/en/article/entrepreneurship/forging-new-frontiers-in-finance-for-digital-innovations-m4qcv3ktj9ft>

Gonfalonieri, Alexandre (2019). *How to get Funding for AI Startups.* KDnuggets. <https://www.kdnuggets.com/2019/06/funding-ai-startups.html>

González Alarcón, Natalia; Pombo, Cristina (2020). *How can artificial intelligence help in a pandemic?.* Inter-American Development Bank

<http://dx.doi.org/10.18235/0002300>.

<https://publications.iadb.org/es/como-puede-la-inteligencia-artificial-ayudar-en-una-pandemia>

Google (2018). *Whitepaper: Perspectives on Issues in AI Governance.*

Hoepman, Jaap-Henk (2020). *Privacy Design Strategies (The Little Blue Book).* Computing Science Department. Radboud University. <https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf>

IEEE P7000TM Standards.

Infocomm Media Development Authority and Personal Data Protection Commission, Singapore (IMDA & IDPC) (2020). *Model Artificial Intelligence Governance Framework. Second Edition.*

ISO/IEC JTC 1/SC 42 Artificial intelligence.

KPMG (2018). *The importance of the Privacy Impact Assessment (PIA) in data protection.* <https://www.tendencias.kpmg.es/2018/01/la-importancia-de-los-privacy-impact-assessment-pi-a-en-la-proteccion-de-datos/>

Kusner, Matt; Loftus, Joshua; Russell, Chris & Silva, Ricardo (2017). *Counterfactual Fairness.* Neural Information Processing Systems. <https://papers.nips.cc/paper/6995-counterfactual-fairness.pdf>

Lewis, Barnaby (2019). *Standards Cooperation is Key to Making AI and Smart Cities a Reality.* ISO Committee on Artificial Intelligence. <https://www.iso.org/news/ref2381.html>

Lockett, Kath (2019). *Embracing the Power of Technology.* ISO Committee on Artificial Intelligence. <https://www.iso.org/news/ref2451.html>

MacCarthy, Mark (2020). *AI needs more regulation not less.* Brookings Institution. <https://www.brookings.edu/research/ai-needs-more-regulation-not-less/>

McKinsey Global Institute (MGI) (2017). *Artificial Intelligence: The Next Digital Frontier?. Discussion Paper.*

Méndez, Manuel Ángel (2020). *The Government Joins a Questioned Digital Tracking Initiative: "It's a Trojan Horse."* Confidential. Apps.

https://www.elconfidencial.com/tecnologia/2020-04-17/gobierno-apps-rastreo-pepp-pt-dp3t-coronavirus_2553776/

Molnar, Cristoph (2020). *Interpretable Machine Learning: A Guide for Making Blackbox Models Explainable.* <https://christophm.github.io/interpretable-ml-book/>

Official Journal of the European Union (2016-2018). *General Data Protection Regulation 2016/679.*

Omidyar Network (2019). *Due -Diligence Checklist for Privacy-Aware Investing. Race to the Top Initiative.* _____ (2019). *Progression Matrix. Race to the Top Initiative.*

_____ (2019). *The Emerging Market in Digital Trust: New Obligations and New Opportunities for Venture Capital.*

Organisation for Economic Co-operation and Development (2018). *Private Equity Investment in Artificial Intelligence.*

_____ (2019). *OECD Principles on Artificial Intelligence.* <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

European Parliament (2016). *General Data Protection Regulation.* <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679&from=es>

Posey, Luke (2019). *Intro to AI Ethics. Towards Data Science.* <https://towardsdatascience.com/intro-to-ai-ethics-2a31fbb6e87>

Potgieter, Luke (2018). *Semantic Interoperability: Are you training your AI by mixing data sources that look the same but aren't?. KDnuggets.* <https://www.kdnuggets.com/2018/10/semantic-interoperability-training-ai-mixing-different-data-sources.html>

Szramke, Tomasz (2017). *Types of AI Algorithms.* Growly. <https://www.growly.io/types-of-ai-algorithms/>

World Economic Forum in collaboration with the Info-communications Media Development Authority of Singapore (WEF & IMDA) (2020).

*Companion to the Model AI governance
Framework - Implementation and Self-Assessment
Guide for Organizations.*

Wright, Ray (2019). *Interpreting Black
Box Machine Learning Models Using Partial
Dependence (PDP) and Individual Conditional
Expectation (ICE) Plots.* SAS Institute.

[https://www.sas.com/content/dam/SAS/support/
en/sas-global-forum-proceedings/2018/1950-
2018.pdf](https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1950-2018.pdf)