Telefónica´s Approach to the Responsible Use of AI

TELEFONICA, S.A.

Introduction

Artificial Intelligence (AI) is on the rise. It can be applied to many different domains such as content recommendations, chatbots, image recognition, machine translation, fraud detection, medical diagnosis, autonomous vehicles, legal, education, transport, and logistics to name just a few. It can not only be used for business, but also for social purposes such as better understanding and reducing the impact of climate change, natural disasters, and migration. Also, in Telefonica, AI and Big Data are used increasingly. There are four main areas of application: i) optimization of core business; ii) innovation in the customer relationship using cognitive technologies for digital assistants in apps, webs, call centres, shops, etc; iii) offering AI and Big Data services to business customers through Telefónica's B2B area; and iv) using AI and Big Data for social good such as the fight against COVID-19 and contributing to the Sustainable Development Goals. The last two areas are always based on anonymized and aggregated data.

However, recently several concerns have been expressed about the use of AI, in particular related to potential discrimination (bias, discrimination, predictive parity), lack of interpretability of algorithmic conclusions (explainability, black box problem), and lack of transparency of personal data used.

To deal with those potential problems, Telefonica published its AI Principles in October 2018. This document describes Telefónica´s Responsible Use of AI, which is a part of the broader Responsibility by Design-Approach of Telefónica. Given that there is still little industry experience, we will adapt and adjust the approach based on experience and new external developments.

Telefónica´s Approach to the Responsible Use of AI

The approach for the responsible use of AI includes:

- a strategic model: the strategic vision on how the responsible use of AI aligns with wider company objectives
- an organizational and relational model, which defines the functions needed, aligned with the corporate structure, in addition to the roles & responsibilities, and relationships
- an operating model defining the relevant processes along with roles responsible for the tasks to be carried out. This includes the Responsible AI by Design methodology.

The strategic model: Telefonica's Principles of Artificial Intelligence

Telefonica is strongly committed to respecting Human Rights, as is stated in its Business Principles and Human Rights Policy. This includes a commitment to developing products and services aimed at making the world a better place to live in and mitigating any negative impacts technology may have on society or the environment. Technology should contribute to making society more inclusive and offer better opportunities for all, and AI can contribute to these goals.

In order to guide the organization in its uptake of AI and Big Data across the business, Telefonica has published its "Principles of AI". The principles include:

- **Fair AI** seeks to ensure that the applications of AI technology lead to fair results. This means that they should not lead to discriminatory impacts on people in relation to race, ethnic origin, religion, gender, sexual orientation, disability or any other personal condition. When optimizing a machine learning algorithm, we must take into account not only the performance in terms of error optimization, but also the impact of the algorithm in the specific domain.

- **Transparent and Explainable AI** means to be explicit about the kind of personal and/or non-personal data the AI systems uses as well as about the purpose the data is used for. When people directly interact with an AI system, it should be clear to the users that this is the case. When AI systems take, or support, decisions, a certain level of understanding of how the conclusions are arrived at needs to be ensured, by generation explanations about how they reached that decision, like is illustrated in for the particular case of supervised machine learning. Those explanations should always consider the user profile to adjust them to the transparency level required. This also applies in case of using third-party AI technology.

- **Human-centric AI** means that AI should be at the service of society and generate tangible benefits for people. AI systems should always stay under human control and be driven by value-based considerations. AI used in products and services should in no way lead to a negative impact on human rights or the achievement of the UN's Sustainable Development Goals.

- **Privacy and Security by Design** means that when creating AI systems, which are fueled by data, privacy and security aspects are an inherent part of the system's lifecycle. This maximizes respecting people's right to privacy and their personal data. Notice that the data used in AI systems can be personal or anonymous/aggregated. Notice also that this principle is broader applicable than only to AI systems, and Telefonica already has processes in place to ensure proper privacy and security.

- **Working with partners and third parties** means Telefónica is committed to verifying the logic and data used by the providers to ensure that its principles are respected.

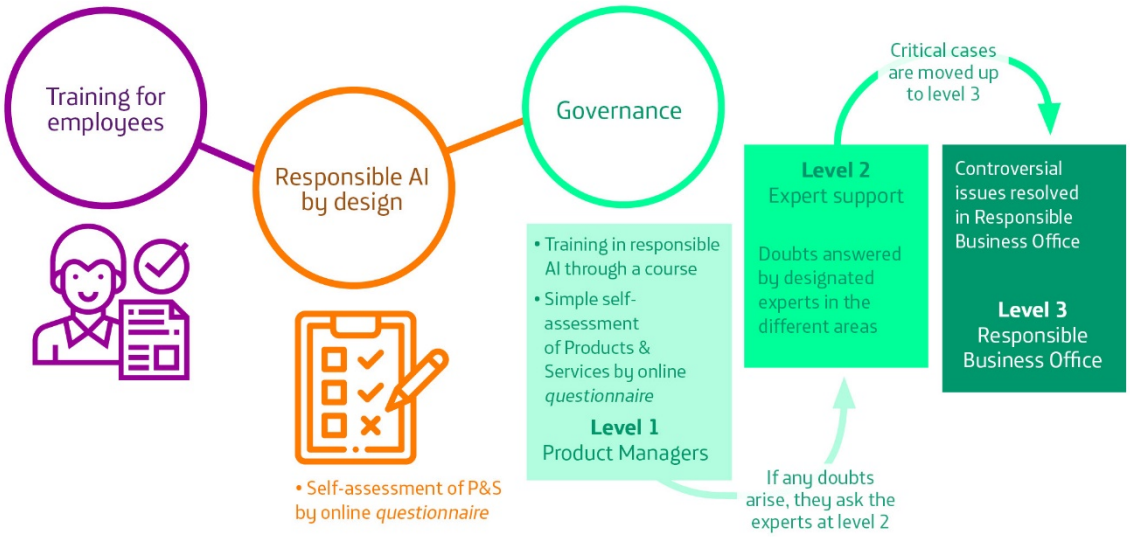
Those Principles are also based on a broad consensus in expert communities, as well as on specific aspects of the telecommunications industry.

There is, however, little published experience on how such principles can be practically implemented in large organizations such that they have the desired effect. In this sense, Telefonica is making an important step with this approach for the responsible use of Artificial Intelligence.

The organizational and relationship model

We are implementing responsible AI through an organizational and relationship model that defines what areas of the company are involved, what their roles are and how they relate to each other for the achievement of a responsible use of AI.

We promote a self-responsibility approach with on-demand escalation. There is a 3 step escalation process as illustrated in the figure below.



Product managers/developers who purchase, develop and/ or use Artificial Intelligence are to carry out a simple self-assessment of the product/service they are developing already at the design phase through an online questionnaire. This self-assessment explicitly covers potential human rights risks associated with the use of Artificial Intelligence. This self-assessment will be integrated into a 3 tiered governance model, supported by a broader Community of Experts (among them a single-point-of-contact for questions relating to AI & Ethics, the Responsible AI Champion). If a product manager/developer (level 1) has doubts about a potential adverse impact of a given product/service after completing the self-assessment, and this doubt cannot be resolved with the help of the RAI, she will be automatically directed to a group of predetermined, multidisciplinary experts within the company (level 2), that together with the product manager/developer try to solve the issue at hand. In case this issue turns out be a potential risk to the company´s reputation, the matter is elevated to the Responsible Business Office which brings together all relevant department directors at global level (level 3).

The operating model

The operating model describes the processes of how to implement the Responsible AI Approach in the organization on a day to day basis. Integrated within the broader Responsibility-by-Design-Approach, it includes a methodology called “Responsible AI by Design”, inspired by methodologies such as Privacy and Security by Design. The operating model consists of, among others:

- Training & awareness activities
- The self-assessment on line questionnaire, where each AI principle is operationalized through a set of questions to answer along with recommendations.
- A set of technical tools that helps in answering the questions

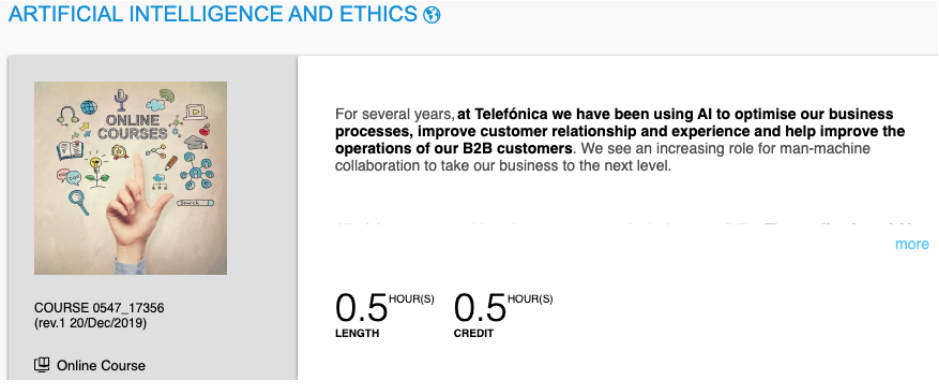
The design process of the methodology required a cross-enterprise initiative involving different departments such as Engineering, Corporate Ethics & Sustainability, Security,

Legal, Business, Human Resources, Procurement, as well as an endorsement of top management.

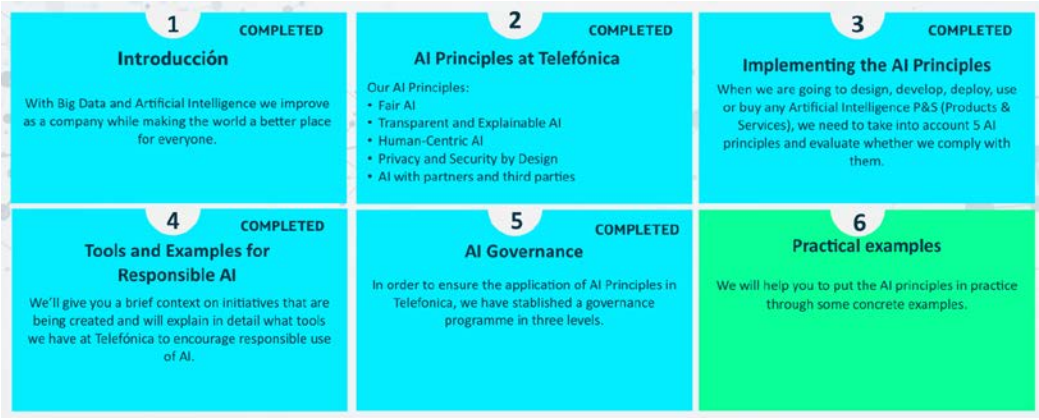
Training & awareness

While Artificial Intelligence is still a rather new technology used in large organizations, the ethical and societal impacts are even more recent. Therefore, it is of utmost importance to explain to employees what AI is, how it works and how it might lead to undesired consequences.

Telefónica has therefore developed courses related to AI & Ethics that are accessible to all employees through the standard corporate portals in three languages (Spanish, English and Portuguese). Depending on the profile of the employee he or she can access a light version of the course which is divided in three modules of each 10 min, or a more profound version which takes about 1 hour to complete.



The full course is divided into 6 modules as illustrated in the figure below:



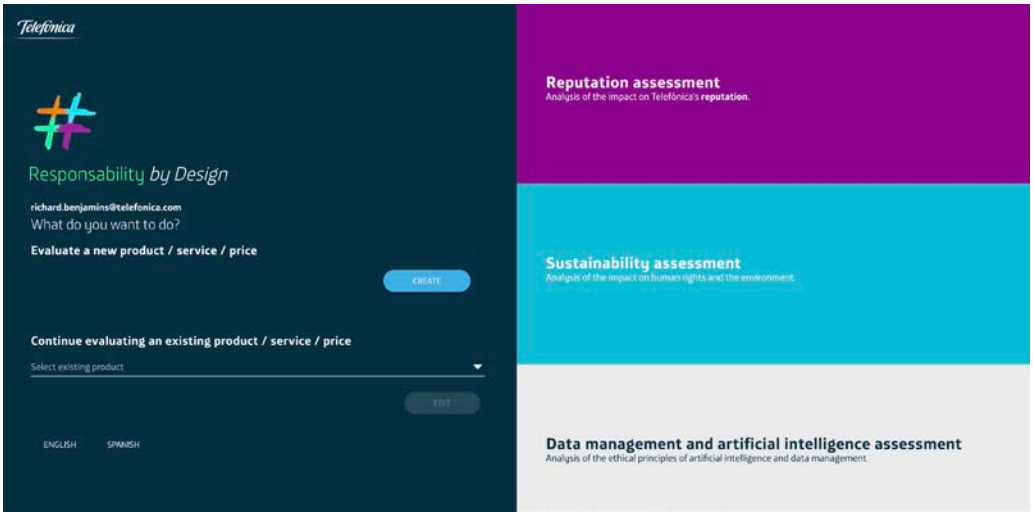
Apart from online courses, each business unit using AI is participating in dedicated workshops explaining the governance model. Moreover, there is a mini-guide available on the intranet where employees can get a quick overview of how to apply the AI Principles for ensuring a responsible use of AI.

Questionnaire with recommendations

For all products and services that use AI or Big Data, the responsible manager needs to complete the self-assessment questionnaire where for each principle, several questions must be answered. The questionnaire is available online in Spanish and English and integrated in the global "Responsibility by Design" initiative of Telefónica Group. All completed questionnaires are logged for inspection, statistics and actions if so required.

Managers who have completed the questionnaire will receive an email with their completed questionnaire, a set of recommendations where appropriate, and an indication of issues to resolve or revisit later.

The designing of the methodology has involved a cross-enterprise initiative have involved different departments such as Engineering, Corporate Ethics & Sustainability, Security, Legal, Business, Human Resources, Procurement, as well as an endorsement of top management.



The questionnaire consists of 12 questions covering all AI Principles. An example can be seen in the figure below:

ETHICAL PRINCIPLES - Data management, Big Data and Artificial Intelligence

One of the major concerns of the advances in technology is related to the widespread application of Artificial Intelligence and Big Data, and how to ensure an ethical use. One of the solutions is to apply ethical principles during the design and development of such systems. For this reason, when working with AI algorithms and Big Data technologies, we should incorporate such principles "by design".

PRINCIPLE 1: Fair AI

We seek to ensure that applications do not produce biased results and do not lead to unfair discrimination

1. If the data set contains sensitive variables and the use of this variable is not justified by the purpose of the application:

- For AI: Did you check during the algorithmic training process that the model does not discriminate with respect to the sensitive variable?

- For Big Data: Did you refrain from using the sensitives variables in the operations/calculations?

The use of sensitive variables is forbidden by law for decision making, and include racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data, health, sex life or sexual orientation..

Yes	No	N/A
○	◉	○

When using Machine Learning: in case there are sensitive variables in the data set and their use is not justified by the purpose of the service (e.g., an application for medical diagnosis or monitoring), those variables have to be deleted during the training of the algorithm. In order to ensure a fair result, the sensitive variables can be used afterwards (once the training is done) with the sole objective to verify that the learned model is not discriminating with respect to those variables.
When using Big Data (not using Machine Learning): in case there are sensitive variables in the data set and their use is not justified by the purpose of the service (e.g., an application for medical diagnosis or monitoring), those variables have to be deleted.

It is important to notice that while privacy & security are part of the AI Principles of Telefónica, the company has specific dedicated processes and areas taking care of privacy and security. Indeed, privacy and security are relevant for any digital system, and not only for AI and Big Data applications. Therefore, regarding the principle on Privacy & Security, the questionnaire refers to the respective areas of the company: the DPO and the CISO.

Technical tools

Because some of the questions of the questionnaire are impossible to answer without specific tools, our methodology includes both in-house tools and external tools (mostly open source). Some of the in-house tools relate to privacy-enhancing technologies such as anonymization, a tool to detect algorithmic discrimination against protected groups, and a personal data transparency tool. We are constantly evaluating new external tools, especially related to bias/discrimination and explainability. This is a very active area of research with new tools appearing at a rapid pace. It is the responsibility of the Community of Experts to study and recommend the inclusion of new tools.

Further Telefónica references

Towards organizational guidelines for the responsible use of AI, Richard Benjamins, 24th European Conference on Artificial Intelligence - ECAI 2020 Santiago de Compostela, Spain, https://ecai2020.eu/papers/1347_paper.pdf

Responsible AI by Design in Practice, Richard Benjamins, Alberto Barbado, Daniel Sierra, Proceedings of the Human-Centered AI: Trustworthiness of AI Models & Data (HAI) track at AAAI Fall Symposium, DC, November 7-9, 2019, https://arxiv.org/html/2001.05375, https://arxiv.org/abs/1909.12838

Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Alejandro Barreto Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, Information Fusion, Volume 58, June 2020, Pages 82-115.

Towards a framework for understanding societal and ethical implications of Artificial Intelligence, Richard Benjamins, Idoia Salazar, In vulnerabilidad y cultura digital. Riesgos y oportunidades de la sociedad hiperconectada, Dykinson, pages: 2020, https://www.dykinson.com/libros/vulnerabilidad-y-cultura-digital-riesgos-y-oportunidades-de-la-sociedad-hiperconectada/9788413246475/