Taula d'entitats
del Tercer Sector Social
de Catalunya

m4Social

REPORT

RADAR OF ALGORITHMS AND AUTOMATED DECISION-MAKING PROCESSES FOR CITIZENS' ACCESS TO SOCIAL RIGHTS

JANUARY 2024

CREDITS

Study powered by:
Taula d'entitats del Tercer Sector Social de Catalunya

Coordination:
Roger Civit
Laura Castro
Xènia Carné
Gerard Sentís

Authors of the study [KSNET]:
Romén Adán González
Mireia Álvarez
Elena Costas

Peer reviewers [KSNET]:
Beatriz Belmonte Acero
Manuel González Noriega

Design: Toni Sánchez Poy

Barcelona, January 2024

CONTENTS

SUMMARY

Artificial intelligence (AI) has developed exponentially over the last decade. These developments have primarily been driven by breakthroughs in the field of deep learning. Generative AI, for text, images and audio, is at a level of development today that was difficult to imagine just a few years ago. Meanwhile, systems that use AI are increasingly a part of everyday life. These range from smart personal assistants, recommendation systems on platforms, bots, models that help to diagnose and treat diseases, to the translation and transcription of texts.

The possible applications of AI also have transformative potential, not only in the private sector, but also in the public sector and in the social field in particular. Their ability to simplify processes and provide improved public services can facilitate and speed up both the work of staff in government bodies and citizens' relationships with them. However, the use and application of AI systems also entails a series of risks as well as benefits.

In this context, **the Platform of Third Social Sector Institutions has launched the project *Radar of artificial intelligence algorithms and automated decision-making processes for citizens' access to social rights in Catalonia*.** This project has a twofold objective: first, to help society in general and Third Sector institutions in particular to understand the current context in order to establish a shared position. Second, to map all the artificial intelligence and automated decision-making systems that are currently being used by the Catalan Public Administrations in the social sphere.

For this reason, **this study consists of two documents**. The first, ***Artificial Intelligence in Public Administrations***, which was presented in September 2023, is an informative report on the origins, development, governance, benefits, risks and application of artificial intelligence by the Public Administration.

The second document, ***Radar of algorithms and automated decision-making processes for citizens' access to social rights***, which was completed in December 2023, covers a total of 12 identified systems in four different Catalan public administrations. A factsheet containing important infor-

mation for each of these algorithms has been produced. This Radar is the first repository of artificial intelligence systems and automated processes used in the social sphere used by Public Administrations in Catalonia that has been created.

In a context of profound transformations, it is essential to be able to identify what is relevant, the breakthroughs taking place and the risks they entail, and the expected achievements arising from the developments taking place. This study seeks to contribute to raising awareness of Artificial Intelligence as well as to public debate, primarily by increasing public knowledge of the artificial intelligence systems that are being used by government bodies.

DOCUMENT 1

ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATIONS

SEPTEMBER 2023

1. INTRODUCTION

'Artificial Intelligence' (AI) has become a term that is as widespread as it is (perhaps) unhelpful. It is an umbrella concept, and it is sometimes hard to determine what is being referred to when it is used. Although we have been living with scientific breakthroughs in the field of AI for some years, with its use in the private sector and its integration in many of the applications and websites which we use on a daily basis, it is only recently that the media and public opinion has begun to pay particular attention to it. Everything probably changed with the advent of ChatGPT, a user-friendly interface that lets anyone with access to the Internet interact directly and consciously with an Artificial Intelligence system. It is a system that takes just a few seconds to produce answers similar to those a human being could provide.

This has raised many doubts in the public debate. Will most people lose their jobs? Will AI put an end to critical thinking? Or even to humanity? However, it is important to create spaces for reasoned debate about Artificial Intelligence. **What exactly are we talking about when we talk about AI? What are its origins, and how has it developed? How has it been used in the private sector in the last decade? How could the Public Administration benefit from it, and what innovative applications are taking place in other countries? What benefits can AI bring to society? What are the risks, and how should we address and regulate them? There are of course many others.**

In 2022, the AI Impacts project completed the second edition of the *Expert Survey on Progress in AI*[1], in which they interviewed 738 researchers who were experts on machine learning on the impacts of Artificial Intelligence in the future. While it remains a value judgement, half of the Artificial Intelligence experts thought that the probability of the long-term effect of advanced AI on humanity being 'extremely bad (e.g. human extinction)' was only 5% or lower. Indeed, one in four of the experts rated it at 0%.

Although Artificial Intelligence has considerable potential for change in the private sector, the public sector is not immune. Several public authorities already use this technology to help them with different tasks. Indeed, the main type of technology currently used in the public sector is Automated Decision-Making Systems (ADMS), which may (or may not) use AI.

1 It can be consulted here.

Automated decision-making systems (ADMS) are systems that are able to make decisions using technological means without any human intervention. They may be based on any type of data, such as those provided by individuals in questionnaires, observed data such as geolocation data, or data inferred from certain characteristics[2]. According to the *Automating Society* report by AlgorithmWatch[3], ADMS may (or may not) use artificial intelligence techniques, i.e. they may use simple rule-based analysis procedures and automation, or implement sophisticated techniques such as natural language processing, predictive analytics, or computer vision. The authors advocate adopting a holistic approach - hence the use of the word systems rather than technologies. When referring to a system, the concept encompasses not only the algorithm or technique in isolation, but instead the entire decision-making process, the diagnosis that explains why the solution was adopted, the algorithm itself, the data, the code, how it was developed (by a public or private company) and finally implemented. Using a holistic approach, it is possible to understand the overall process, from the time a problem or need is identified, to the final implementation of a solution, by way of all the necessary intermediate steps.

This report therefore uses the notion of AI as a broad concept that includes both ADMS and artificial intelligence, given that ADMS are now increasingly prevalent within the Public Sector, and present opportunities and hazards similar to those of AI.

By considering AI as a resource for the Public Sector when allocating social welfare benefits to citizens, for example, it is essential that Third Sector institutions understand how these systems are being used in the field of social rights. **The objective of this report is therefore to help society in general and Third Sector institutions in particular to understand the current context in order to establish a common position**. If institutions want to be heard, it is essential that they share and convey a similar vision, in terms of both diagnosis and proposals.

From this starting point, the report is structured as follows: a series of concepts that appear throughout the report are briefly defined in the second section. The third chapter describes the origins and development of AI up to the present day. A short example clarifying the concept of training as it applies to AI is then provided. Sections 4 and 5 focus on AI governance, the

2 This definition is taken from the document *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679* prepared by the Data Protection Working Party, an advisory body of the European Union. It is available for consultation here.

3 Available here.

role of governments and the European Union's AI Act. Chapter 6 presents systems that use AI which are applied by Public Administrations (PAs), in various areas, internationally, in Catalonia and in Spain. Finally, chapter 7 reviews the main benefits and above all the risks of the application of AI by Public Administrations in the social sphere.

2. DEFINITIONS

Artificial Intelligence (AI) is a broad and complex field, with many subsections and associated techniques. This umbrella definition covers a wide variety of techniques: machine learning, natural language processing, voice recognition, expert systems, robotics and computer vision. Before starting, some of the definitions used throughout this report are defined below to avoid confusion.

2.1 Key definitions

- **Algorithm**: In general terms, an algorithm is a set of rules or instructions that solve a problem or achieve a goal in a step-by-step procedure. In the context of AI, algorithms often refer to the methods used to build and train machine learning models. There are many different algorithms, and when building an AI model, the data scientist chooses the most appropriate algorithm for the problem to be solved.
- **Data**: Artificial intelligence is impossible without data. Data are the source of information. A large proportion of the creation of a model is focused on collecting and cleaning data. Data may consist of images, text, or a data array (such as an Excel spreadsheet). Regardless of their original format, the information is always converted into numbers for the selected algorithm to learn.
- **Training**: This is the model's creation phase. After having collected and processed a certain amount of data, the data are passed on to a specific algorithm so that it 'learns' to recognize patterns. For example, with information about the location, characteristics and the price of homes, a model could be trained to learn to predict the price of a new home based on its location and characteristics.
- **Artificial Intelligence (AI)**: There is no consensus in the scientific community on how to define the term 'AI'. However, the European Commission has proposed an operational definition to establish legal liabilities which is based on a definition of the OECD. Article 3(1) of the proposed Artificial Intelligence Act states that an 'Artificial Intelligence system' is:

  «Software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.»

The techniques listed in Annex 1 of the proposal for the Artificial Intelligence Act include:

> *a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods, including deep learning;*
>
> *b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deduction engines, (symbolic) reasoning, and expert systems;*
>
> *c) Statistical approaches, Bayesian estimation, search and optimization methods.*

- ***Machine Learning***: A subfield of AI that focuses on designing systems that can learn from data. In other words, instead of directly programming all the actions that the model will carry out (as was the case in the 1960s and 1970s), machine learning systems are trained with a dataset, and learn patterns that are then used to make predictions or classifications. Machine learning algorithms include linear regression algorithms, logistic regression, k-nearest neighbours, classification trees and random forests, among many others.
- ***Deep Learning***: A subfield of machine learning that focuses on artificial neural networks with many layers ('deep'). Neural networks are another type of algorithm within the field of AI. These models are capable of learning complex and subtle patterns in data. They are used in a wide variety of applications, including speech recognition, computer vision, and natural language processing.
- ***Classification and prediction algorithms***: The terms 'classification' and 'prediction' are used in the field of machine learning to describe two different types of problems that these algorithms can tackle.
  - **Classification algorithms** are used to predict/identify the category or type of an object or event. An example of classification is spam detection, in which emails are classified as 'spam' or 'not spam'. Another example is a model that decides whether or not a person is entitled to social welfare benefits. There are various types of classification algorithms, including logistic regression, decision trees and k-nearest neighbour algorithms (often abbreviated to k-NN).
  - **Prediction algorithms** focus on predicting a continuous value. For example, they can be used to predict the price of a house based on its characteristics, or how much inflation will rise in the next month.

This group includes algorithms such as linear regression ad regression trees.

- **Generative AI or Foundational Models**: These are AI systems that can create new and original content. This can range from generating text and creating images, to music and other types of content. OpenAI's GPT-4 model, which follows ChatGPT, is an example of generative AI in the field of text.
- **Symbolic AI**: this term refers to the developments in the first wave (the 1960s and 1970s). Symbolic AI refers to approaches to developing intelligent machines involving encoding experts' knowledge and experience into sets of rules that can be executed by the machine.

2.2 An example of application

Having set out the definitions that will be used throughout this report, and to complete our understanding of how this technology works, here is a simple and hypothetical example that will shed some light on how most systems use machine learning algorithms, by answering the following question: how are these machine learning models trained?

Imagine that a Third Sector institution is carrying out a programme of mathematics and language support classes for students in difficult socio-economic situations. It works with more than 100 secondary schools, which are initially sent a single support teacher. When necessary, the institution sends another teacher to a secondary school to enhance the project's quality and impact.

However, the process of deciding whether to send an extra support teacher is expensive and time-consuming. The teacher working on the project has to complete a report, experts from the institution have to go to the secondary school and assess the case, etc. The institution has learned from experience that the process is so expensive that they are unable to provide answers in time. At this point, they decide to use an Artificial Intelligence system.

Fortunately, the institution's data team has been collecting information from the various projects, including information on the municipality where the secondary school is located, the district's per capita income, the students' average grade, and the school dropout rate. They have also included a column in their Excel spreadsheet that shows whether or not an extra support teacher has been sent to the project.

Project	Secondary school unicipality	Income per capita in the secondary school district	Students' average grade	Secondary school dropout rate	Support teacher
1	A	25	7	7	No
2	B	10	5,6	43	Sí
3	B	8	8	33	Sí
4	C	17	7,2	20	No
...	...	...	...	...	...
50	D	16	7,5	15	No

Based on this information, the data team decides to train a machine learning model that predicts whether or not an extra support teacher needs to be assigned to a project. The data analysts will not decide which criterion to follow, but will let the model 'learn' from the data. In other words, the model will identify the circumstances in which the institution decides to send a new support teacher. This learning process is called model training, as the model learns from the following data:

Explanatory variables				Outcome
Secondary school unicipality	Income per capita in the secondary school district	Students' average grade	Secondary school dropout rate	Support teacher
A	25	7	7	No
B	10	5,6	43	Sí
B	8	8	33	Sí
C	17	7,2	20	No
...	...	...	...	...
D	16	7,5	15	No

The idea behind the training is for the algorithm to learn to predict the most probable outcome based on the explanatory variables it uses. For example, projects in some municipalities with a drop-out rate below a given value and a specific average grade tend to have a support teacher assigned to them.

How do we know if the algorithm provides a good result? The data team will compare the actual results (whether or not the project is assigned a support teacher) with those estimated by the algorithm for each project.

If the model only gets 30% of the cases right, the model will be discarded as it is not useful for the institution. The data team will try using different models and adjustments until they find one that achieves a higher level of success.

After the model has been trained, the data team applies it to the remaining fifty projects for which they do not know whether to send a support teacher. The model has estimated that five schools would need a support teacher. However, instead of having analysed the fifty projects on a case-by-case basis to find out whether an extra teacher would be necessary, it will focus on the five identified as priorities by the model.

Explanatory variables					Outcome
Project	Secondary school unicipality	Income per capita in the secondary school district	Students' average grade	Secondary school dropout rate	Support teacher *(estimate)*
51	B	22	7,4	12	No
52	A	8	5	31	Sí
...	D	...	...	...	...
99	D	12	5,2	6	Sí
100	C	25	6,5	25	No

It is important to note that the data team will be able to explain why the model estimates that a support teacher is needed or not depending on which machine learning algorithm has been used. The most accurate models, i.e. those with the best results, tend to be more opaque 'black boxes', which means that it will be difficult for the data team to explain the reason for the result to the institution. On the other hand, simpler algorithms are more explainable, but can make many more mistakes.

In conclusion, what was the process used in this hypothetical example? First, the institution identified a shortcoming: some projects could ask for a support teacher, but they do not have sufficient resources to examine each case individually.

Second, based on information from other projects for which they decided at the time whether or not to send a support teacher, they created a machine learning model which was able to learn which are the most important characteristics when making the decision, and was therefore able to predict on its own whether an extra support teacher should be sent or not in other cases subject to the same variables. No team of experts at the

institution decided which characteristics were important when making the decision. The algorithm identified the patterns by itself, learning what the institution took into account when making this decision in the past.

This made it possible to estimate whether all of the projects need a support teacher in just a few seconds, without having to carry out an individual and very expensive assessment of each school. The institution increases its efficiency and effectiveness by being able to focus on the projects that need it most.

3. THE ORIGINS OF ARTIFICIAL INTELLIGENCE AND THE LATEST DEVELOPMENTS

The history of the development of Artificial Intelligence is a story of vicissitudes and fluctuations, from the most basic models to the sophisticated deep learning systems. This non-linear progress has 'exploded' over the last five years. So how have we reached this point?

The origins of AI date back to the middle of the twentieth century, when Alan Turing proposed the idea of 'thinking machines' **in the 1950s**. The Dartmouth Conference, where the term 'Artificial Intelligence' was coined, was held in 1956. The organizers of the conference, John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, are considered some of the founding fathers of AI.

One of the first machine learning models was the perceptron, which was a supervised learning algorithm that was used to perform simple classification tasks. Although it initially sparked a great deal of enthusiasm, its limitations when solving more complex problems soon became apparent.

AI experienced what is known as its first 'golden age' in the 1960s and 1970s, when breakthroughs in areas including natural language processing and expert systems took place. During this period, researchers believed they were on the verge of creating machines that could understand and respond to natural language, solve complex problems, and learn from their experiences. However, these systems did not 'learn' in the same way as modern machine learning techniques, like neural networks. They were based on rules explicitly encoded by humans. They did not learn from the data, or improve their performance with experience.

Those limitations led once again to the initial enthusiasm beginning to fade in the **mid-1970s**, paving the way for the 'AI winter'. This led to cuts in research funding, and the field became less popular in academia.

The 1980s saw an important milestone in the development of neural networks and by extension, of Artificial Intelligence. Until that time, the application of neural network algorithms had been limited by the difficulties involved in training them effectively, and that was particularly true of multi-layer or 'deep' networks. An efficient solution for training these models had not been found.

The situation changed with a crucial breakthrough in how these networks were trained: the backpropagation algorithm, which was published by Rumelhart, Hinton and Williams in 1986. This algorithm paved the way for effective training of deeper and more complex neural networks, which eventually led to the development of the deep learning techniques we see today.

However, although the backpropagation algorithm was an important step forward, the limitations in computing capacity at that time and the lack of large datasets prevented neural networks from becoming widespread. In other words, AI algorithms with considerable potential already existed in the 1980s, but good data and computers that were powerful enough were in short supply.

This limitation continued during **the 1990s**, although this did not mean that no significant progress was made. Machine learning, and supervised learning in particular, began to become more important during these years. A number of algorithms were developed and refined, including Support Vector Machines (SVM), which provided robust levels of performance in a wide variety of tasks.

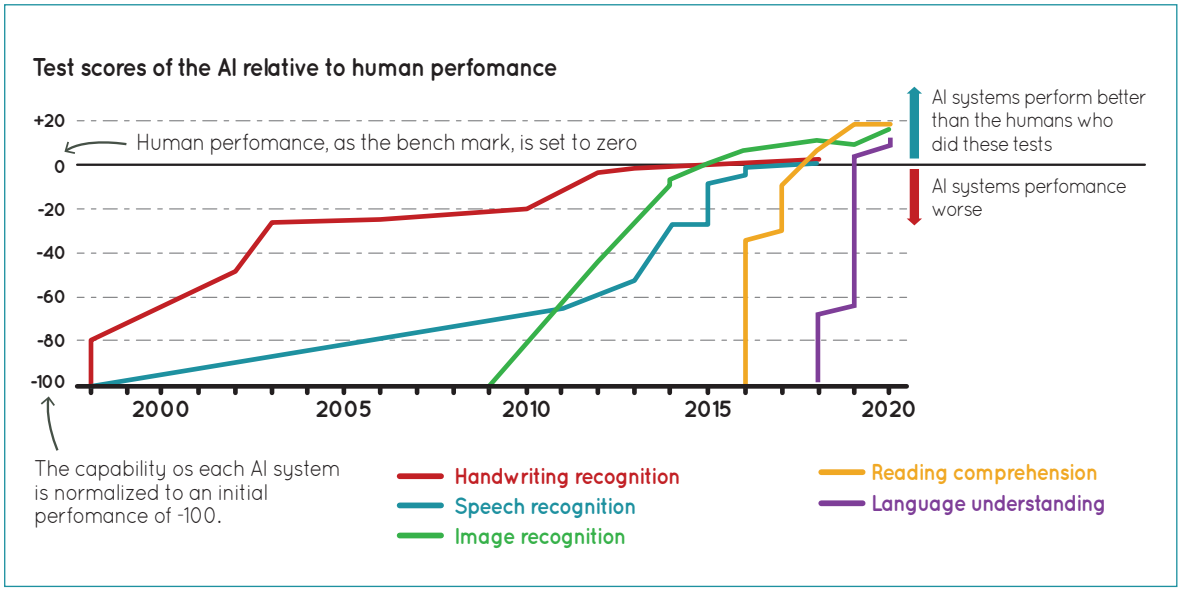
However, it was not until **the first decade of this century**, and the advent of high performance computing and Big Data, when the right conditions were established to enable neural networks, and thus deep learning, to begin to show their true potential. The widespread use of the Internet and the proliferation of digital devices created a large amount of data, providing the fuel that machine learning and deep learning algorithms required.

Geoffrey Hinton published a paper discussing how to train deep neural networks in 2006, and this marked the beginning of the modern era of deep learning. Breakthroughs in AI have been swift and constant ever since.

The decade after 2010 was a period of great progress in the field of Artificial Intelligence. Convolutional neural networks (CNNs), developed by Yann LeCun, transformed the field of image recognition by becoming the standard tool for computer vision tasks such as object recognition in images. Recurrent neural networks (RNNs) proved to be very effective in the field of natural language processing (NLP) and led to significant breakthroughs. In addition, there was a turning point in 2017: Google Research published the architecture of its Transformer neural network. Transformers were used to build the GPT and LaMDA models, which are behind OpenAI's ChatGPT and Google's BARD interfaces. These models are capable of producing answers and solutions at a level that is the same as or higher than a human.

The graph below is probably the best possible summary of the rapid development of AI in recent years. Published by *Our World in Data* in the article 'The brief history of artificial intelligence: The world has changed fast – what might be next???, shows the maximum score achieved by an AI model performing various tasks over time. A score equal to 0 is equivalent to the score obtained by a human, so when an AI model achieves a result higher than 0, this means that it has exceeded human performance levels.

Figure 1. Language and image recognition capabilities of AI systems have improved rapidly..



Source: Roser (2022b)

By 2020, various AI models were able to score as well as or better than a human in tasks including speech, writing and image recognition, reading comprehension and language comprehension. Progress has been surprisingly fast in these latter areas, and has surpassed humans in a few years.

AI has long been used for a wide spectrum of activities. It sets the price of aeroplane tickets, monitors our behaviour at airports, and assists the pilot of the plane during a flight. There are AI models that determine whether people are given a loan, a grant or a job. They are also used to transcribe and translate text. Virtual assistants like Alexa are present in thousands of homes. The series, films and videos that Netflix and YouTube recommend to us are based on AI called recommendation systems. **AI is far from being a technology of the future - it is a technology of the present that has already impacted our lives, and has been doing so for some time.**

According to the report by the Spanish National Observatory of Technology and Society (ONTSI) on how Spanish companies use artificial intelligence, 11.8% of companies with more than 10 employees used the technology in 2022 (50% more than the previous year), and 4.6% micro-enterprises did so (an increase of approximately 30%). Spain was at the EU27 average level in 2021, although the long way behind the level of adoption of the leading countries such as Denmark (24%), Portugal (17%) and Finland (16%).

The country's companies adopted AI to automate workflows or help in decision-making (46.2% of companies) and to identify objects or people based on images (39.7%). More than 30% use natural language processing techniques, more than 20% use technologies to generate written or spoken language, and around 30% apply machine learning techniques. Moreover, these percentages will continue to grow. The Government's Digital Spain 2026 strategy sets the goal of 25% of Spanish companies using artificial intelligence and big data by 2025.

However, **it was not until the emergence of generative AI and the famous ChatGPT application in particular, that AI entered the realm of public opinion**. ChatGPT registered 1,000,000 users in just five days. Instagram took two and a half months to reach the same figure, while it took Spotify five months, Facebook ten months, and Twitter three and a half years. It was a historic achievement.

The capacity of the GPT model, and its easy interaction with the user through the ChatGPT interface, has surprised the entire world. It is an AI model that is capable of solving tasks for which it was not trained. It provides reasonably good translations, summarizes information, resolves doubts, generates code in almost any programming language, and does a great deal else besides. Furthermore, generating text is not the only thing that artificial intelligence can do. Generative AI can generate text, images, sounds and videos. OpenAI's DALL-E leads the field in image generation, and can generate and/or edit images from text, or Stable Diffusion, which is also open source.

All these systems are currently subject to some limitations. The images they generate may contain errors. ChatGPT can 'hallucinate' information[4] or make errors in mathematical calculations. However, we should not lose sight of the timeframe: this is the first generation of these models, created

4 In the context of language models, 'hallucinating information' refers to the possibility of the model generating or providing details that are not supported by real data or real information it has learned. This may be due to the nature of the model's training, as it is not focused on 'memorizing' data or facts, but instead the model looks at vast amounts of text and learns to predict the next word in a sequence.

in just a few years. These complex limitations may well be overcome with time. At that point, we will probably be able to assess the true potential of AI systems.

So in the past few decades, AI has gone from being a dream to a reality that affects almost every aspect of our lives. Image recognition, audio transcription, text translation, performing prediction tasks and generating text and videos are areas that have been dominated by Artificial Intelligence models.

Of course, as we will see in later chapters, AI can also be applied in the context of Public Administration, providing opportunities to improve the efficiency, accuracy and scope of public services, but also creating risks and challenges. **Understanding the current situation is crucial for planning future actions in the Third Sector.**

4. GOVERNANCE AND NATIONAL STRATEGIES IN THE FIELD OF ARTIFICIAL INTELLIGENCE

How are decisions made? Are they aligned with ethical values? Who makes them? How are they implemented and who oversees them? In short, why is it important to talk about AI governance? First, Artificial Intelligence has the potential to transform our societies. **Applying the right governance at the right time can help make the transformation positive, preventing and minimizing risks and enhancing benefits**. On the other hand, any AI system that is being used by a company or public institution should be subject to a universal framework which governs an appropriate use of it.

For these reasons, after having considered the development and evolution of AI, it is essential to understand how the interaction between governments, companies, civil society and citizens is taking shape, and how the regulatory environment and discourse around AI is being shaped as a result.

The governance of artificial intelligence takes place at different levels and in different areas: it can happen at an international and a national level. It can come from and be directed exclusively at the private sector, the public sector or both. It encompasses aspects including the definition of ethical standards, the enactment of laws and regulations, and the implementation of supervision and oversight systems to prevent and mitigate potential harm or abuse.

There are various types of global governance initiatives in the field of AI. Some of these initiatives have come from private companies. For example, Microsoft has established the Office of Responsible AI (ORA) and the AI and Ethics in Engineering and Research Committee (Aether) to provide an internal safeguard that AI systems are developed responsibly. There is also the Partnership on AI founded in 2016 by Amazon, Facebook, Google, DeepMind, Microsoft, IBM and Apple (which entered in 2017) which includes more than 100 partners from academia, civil society and industry, in a non-profit organization that advocates for research into best practices and the responsible use of AI. However, there have been some recent changes. Companies like Microsoft, Twitter and Twitch are dispensing with workers who used to study the negative aspects of AI.[5]

5 Information from the article in The Washington Post, *As AI booms, tech firms are laying off their ethicists*, available at: https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/.

These initiatives can also come from international bodies such as the OECD's Recommendation on Artificial Intelligence of 2019, or UNESCO's Recommendation on the Ethics of Artificial Intelligence of 2021. In both cases, they are still non-binding regulations that basically cover ethical and moral principles.

However, the sector has considerably high levels of concentration of power. A small number of companies control the resources that enable the creation of AI systems, including physical, epistemic, and computing resources. Furthermore, the way in which AI is distributed, usually in the form of platforms, means that they are key players. This means that major firms have a key role, and can try to shape expectations, the frameworks for debate and the narratives about the present and future of AI, the vision involved, interests and propositions.

The main risk in this situation is of technologies being implemented which create more risks than benefits due to a lack of a critical perspective. **For this reason it is essential that other actors, such as the Third Sector, organize, coordinate and establish a common position**. This position must be in favour of the breakthroughs and improvements that technology can bring, but anticipate its risks and provide safeguards in its application.

Two technology giants have clearly taken the lead to date. Microsoft has partnered with OpenAI. This lab is the forerunner of the most famous tool, ChatGPT, as well as others such as DALL-E 2 (which is capable of generating images based on a few words) and Whisper (human-level English audio transcription). After investing €10,000 M in OpenAI, Microsoft is integrating AI into all its products: Windows, Word, etc. Meanwhile, Google has responded by teaming up with DeepMind, a company specializing in AI which is based in London. Google has released the Bard tool, a direct competitor to ChatGPT. The two tech giants and two AI research labs are leading the competition (for now).

However, Meta has also taken significant steps in recent months. It has published models for language generation, image generation, text-to-speech generation and vice versa, and music generation. The main difference between Meta and its competitors is its decision to be open source, so that any programmer in the world can access the model and the code, and contribute to its improvement and development. This greatly increases its ability to create models that are more efficient, more robust and more secure.

4.1 The role of States in AI governance

So what role are States adopting? One of the most important changes that is taking place is the role that countries are adopting. The member states of the European Union are leading the way in this new role, in terms of both for their joint push to regulate AI, and the publication of their national strategies in this area.

A recent study[6] classified countries according to their approach and actions in their national strategies:

- **Development' countries**: these countries prioritize a role for the State as a facilitator and initiator of AI projects through innovation and the direct allocation of resources. This approach is predominant among the countries in the post-Soviet bloc and in East Asia, where the State plays a leading role in development.
- **'Promotion' countries**: these countries assign leadership to the private sector. Countries such as the United States, the United Kingdom and Ireland take a decentralized approach and delegate the key role to the private sector, do not involve the State, and prefer the industry to self-regulate. Although they have different approaches, these countries prioritize innovation over protection.
- **'Control' countries**: the countries in the European Union focus on oversight, reflecting the Union's strong belief in regulation. These States want to ensure that society is protected from the risks of AI by thorough regulatory frameworks, and prioritize protection over innovation.

The role of the State is important, because it opens or closes windows of opportunity for actors such as Third Sector institutions. It will be more difficult for the vision and perspective of institutions to be influential in a context in which the Government delegates leadership to the private sector. However, this does not apply to the European context, which **potentially provides an opportunity for institutions to both voice their concerns and make meaningful contributions.** Not only in legislative terms, but also in terms of discourse.

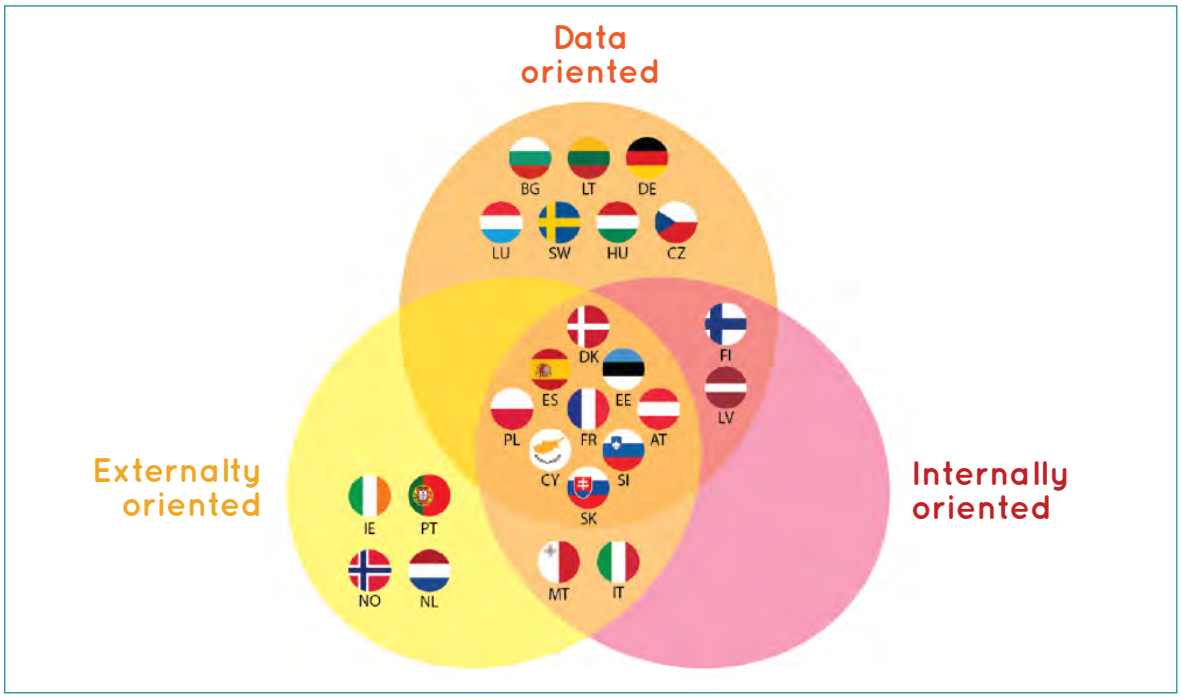
6 Gleb Papyshev & Masaru Yarime (2023) The state's role in governing artificial intelligence: development, control, and promotion through national strategies. *Policy Design and Practice*, 6:1, 79-102, DOI:10.1080/25741292.2022.2162252.

4.2 The role of States in the application of AI in the Public Sector

The report *European Landscape on the Use of Artificial Intelligence by the Public Sector*, produced by AI Watch, the European Commission's knowledge service to monitor the development, uptake and impact of AI, analyses 24 national strategies in European member states (plus Norway) and classifies them in three categories based on the public sector's level of engagement, actions and initiatives. These exercises are very useful as they enable governments to be compared in terms of the actions of others. Knowing the role our country is adopting is essential for understanding the role the Third Sector can play in AI governance.

The first type of States are ***externally oriented***. These governments focus on public-private cooperation for the development of AI in the public sector. They believe that the state does not have sufficient capacity and competences, and that the systems are too complex for dealing with AI autonomously. As a result, they place greater emphasis on creating an ecosystem of start-ups and companies that develop AI for the public sector, promoting public-private collaboration. This group includes countries such as Ireland, Portugal, Norway and the Netherlands.

Figure 2. Clustering exercise on the national strategies.



Source: Excerpt from Tangi et al (2022)

Meanwhile, **data oriented countries** prioritize enabling access to data and improving its quality, removing barriers and improving infrastructure for the development of AI in the public sector. Making more databases available, facilitating data sharing between public institutions and improving data collection and governance are their primary measures. Germany and Sweden are among the countries that fall into this category.

States with an **internally oriented** strategy are those that focus on improving the internal capacity of the State as an instrument to foster the development of AI. They favour methods such as creating new public bodies or institutions specializing in AI, improving the human capital of civil servants by training, and acquiring internal knowledge about AI through events. However, no country is exclusively committed to these measures as the only solution. For example, Finland and Latvia combine them with a data-oriented strategy.

Finally, nine of the States studied, including Spain, include all three approaches in their national strategies. **Most countries are therefore clearly committed to working with the private sector to promote AI in the public sector, which is reasonable since that is where most of the human and technological capital currently is.**

4.3 Spain's National Strategy

In the specific case of Spain, the Spanish Government published its National Artificial Intelligence Strategy in December 2020, in order to create a political framework that defines the various actions that government administrations will take to facilitate the development and deployment of AI in the country's economy and society.

The national strategy contains objectives linked to training and research, such as enhancing the development of human capital in AI through training, attracting talent and qualified employment, and developing scientific excellence to make Spain a leading country.

It also aims to promote the deployment and use of AI technologies in both the public and private sectors, and to position Spain as a leader in the development of tools, technologies and applications for the projection and use of the Spanish language in AI.

While promoting the development of human and technological capital, it also establishes ethical objectives, such as ensuring an ethical framework that defines individual and collective rights and builds an environment of

trust, and ensuring inclusion in the AI-based economy, including gender, digital and territorial inclusion.

In terms of measures directly related to and with an impact on the public sector, Spain's National Strategy presents political initiatives in all the areas analysed by AI Watch:

(1) Increase public sector employees' knowledge of AI;

(2) Improve internal capacity, provide training in AI functions, and create new specialized departments.

(3) Measures to improve data quality and access.

(4) As regards ethics and the legal framework, it seeks to develop both an ethical framework and to reform the laws for data sharing.

(5) In terms of funding and bureaucratic processes, there is a commitment to finance AI projects, foster start-ups based on GovTech and review bureaucratic processes in order to streamline them.

(6) In the field of learning by doing, Spain aims to regulate and lead the implementation of sandboxes.

As a result, regardless of the final translation into policy initiatives and regulatory frameworks, and provided there is no change in priorities or governing parties do not change, **Spain has made a clear initial commitment to be a leader in the development and application of Artificial Intelligence, in both the private and public sectors.**

5. THE HUMAN-CENTRIC APPROACH AND THE EUROPEAN UNION'S ARTIFICIAL INTELLIGENCE ACT

Chapter 3 discussed the historical development and evolution of artificial intelligence since the mid-twentieth century. With this perspective in mind, Chapter 4 focuses on how governments, businesses and civil society are interacting with each other, laying the groundwork for future AI governance, and in particular, how AI is being applied in Public Administrations by States.

The development of any economic sector is heavily influenced by the regulatory framework in which it operates. In the field of artificial intelligence, the law that is stimulating the most debate due to its potential to boost (or undermine) the development of AI in European countries is the European Artificial Intelligence Act.

This Act is based on a very specific perspective: the human-centric approach. This perspective calls for progress that is 'useful' for the individual. In other words, it sees AI as a means to improve well-being, and not as an end in itself. This chapter sets out this approach in order to explain the keys to the draft Act that has already been approved, having explained the starting point.

5.1 The human-centric approach

In 2018, the European Commission created the High-Level Expert Group on Artificial Intelligence in order to produce two reports: (1) Ethics Guidelines for Trustworthy AI and (2) Policy and Investment Recommendations for Trustworthy AI.[7]

The group published the first draft of the ethics guidelines in December 2018, which was submitted to stakeholders and Member States for consultation. The final document, Ethics Guidelines for Trustworthy AI, was published in April 2019. This document, which also sets out the position of the European Commission, sends a clear message: AI is not an end in itself, but rather a means to serve people with the ultimate goal of improving well-being. In other words, **AI must be developed and imple-**

7 The reports by the group of experts can be consulted at: https://digital-strategy.ec.europa.eu/es/policies/expert-group-ai

mented in a human-centric way. Trust in AI is a prerequisite for achieving this.

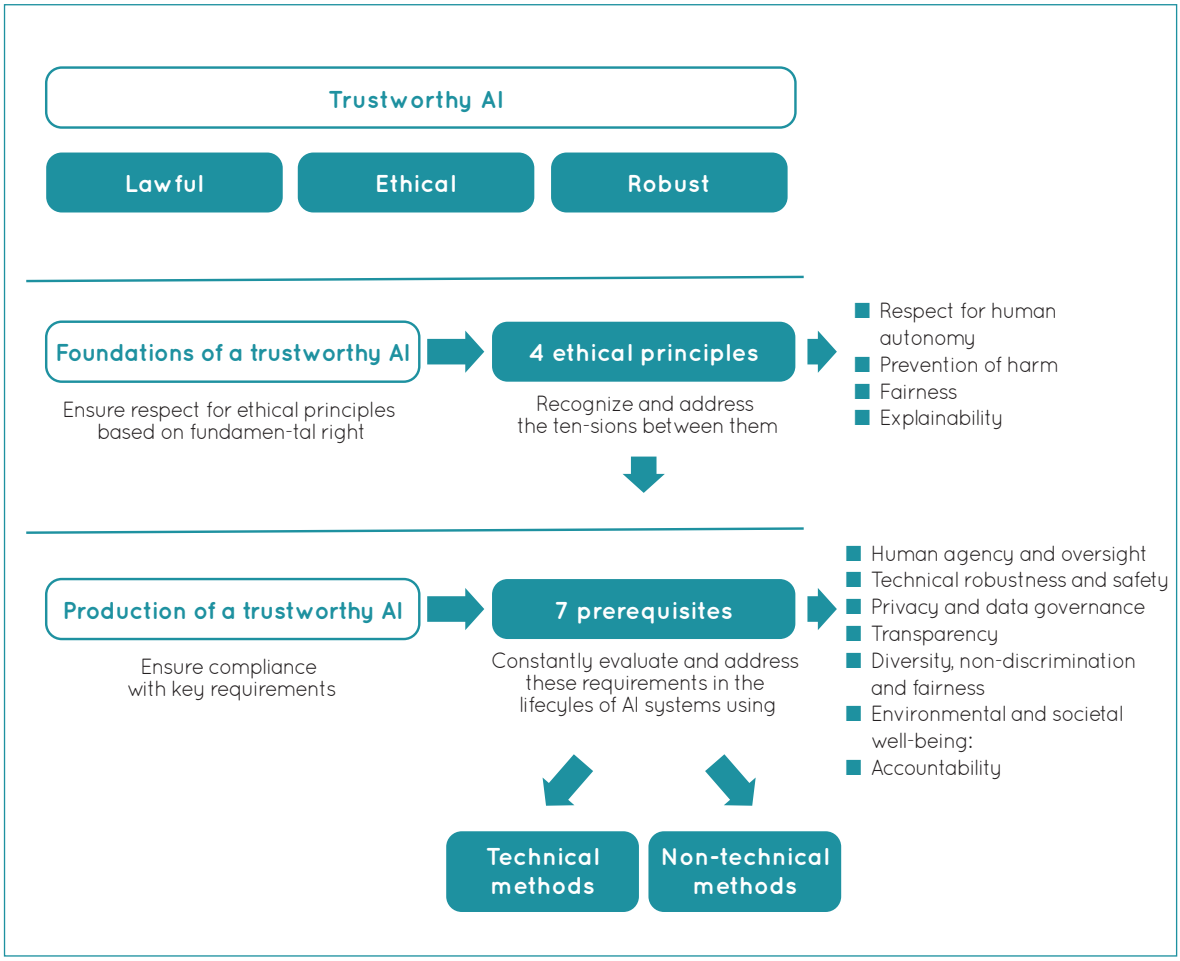
AI has the potential to significantly transform society, improving individual and social well-being. But for this to happen, AI systems must be human-centric, and based on a commitment to their use in the service of humanity and the common good. We are currently within a window of opportunity for shaping the future (and present) development of Artificial Intelligence, which involves both trying to maximize its benefits and preventing and minimizing its risks.

But how can we ensure the development of human-centric AI? The working group's proposal begins by establishing the fundamental rights and ethical principles that any system using Artificial Intelligence should respect. They include respect for individual freedom, democracy, equality and the Rule of Law. Based on all the rights and principles identified, the working group decides that an AI is trustworthy if it fulfils three essential requirements. It must be:

- **Lawful:** it complies with all applicable laws and regulations.
- **Ethical:** it adheres to ethical principles and values.
- **Robust:** it is robust from both a technical and social perspective.

A lawful, ethical and robust AI will be a trustworthy AI that respects the fundamental rights and ethical principles mentioned above. These three abstract components are embodied in seven requirements that every AI system must meet, which are summarized in the following chart and described below:

Figure 3. Framework for a trustworthy AI



Source: excerpted and adapted from HLEGAI (2019)

(1) Human agency and oversight: the user's well-being must be central to the system's functionality, and human oversight must be guaranteed.

(2) Technical robustness and safety: trustworthy AI requires safe, reliable and robust algorithms that deal with the errors and inconsistencies that may arise, as well as potential attacks.

(3) Privacy and data governance: respect for privacy and the quality and integrity of data must be guaranteed.

(4) Transparency: this includes traceability, explainability and communication.

(5) Diversity, non-discrimination and fairness: this includes the prevention of unfair bias, accessibility and universal design, and stakeholder participation.

(6) Environmental and societal well-being: the impacts must be considered in social and not merely individual terms, so both their social and environmental impacts must be taken into account.

(7) Accountability: there must be mechanisms to ensure the accountability of the results, both before and after their deployment. Audits (external and internal) are fundamental in this respect. Potential negative impacts must also be identified, as well as trade-offs and redress.

The European Commission's vision can therefore be summarized as follows: **given the transformative capacity of Artificial Intelligence systems, we now have a window of opportunity to shape their development, so that we ensure that future AI systems will always be human-centric, guiding progress towards improving individual and social well-being, and henceforth preventing its potential uses and designs with negative effects**. This vision is the basis for the development of the European Union's Artificial Intelligence Act.

5.2 The European Union's Artificial Intelligence Act

The European Parliament approved the draft of the European Union Artificial Intelligence Act (EU IA Act), which had been in preparation for two years. Negotiations between the States have now begun, during which Spain could play a key role with its presidency of the Council of the European Union, and the final text is expected to be published at the end of the year.

This legislative text is not only relevant because of the direct impact it could have in Spain, but also because it is an international benchmark. Other countries will take note of the successes and errors in the legislation that is being produced by the European Union (as well as China).

The regulation of AI has been a central political issue in the Union in recent years, and this draft is the result of this effort. After 2019's non-binding 'soft' regulation with the Ethics Guidelines for Trustworthy AI, in 2021 the Commission shifted to a legislative approach, and called for the adoption of a new regulatory framework. Based on a human-centric approach, i.e. one which guarantees the development of a trustworthy AI, it seeks to create a regulation that fulfils two parallel objectives: (1) to promote the development of AI and (2) to manage the associated risks.

In other words, this Act seeks to regulate the uses of AI rather than itself, endeavouring to ensure that technological breakthroughs do not render

the legislation obsolete. Now, before considering this issue, how does it define Artificial Intelligence? Point 1) of Article 3 defines an 'Artificial Intelligence system' as:

software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

The AI systems specified in Annex I are:

a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning.

b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems.

c) Statistical approaches, Bayesian estimation, search and optimization methods.

Private companies' commercial products and applications of AI by the Public Administration that use of some of these strategies will therefore be regulated (in principle) by this law.

The specific objectives of the Act are to:

1) Ensure that AI systems placed on the European Union market and used are safe and respect existing law on fundamental rights.

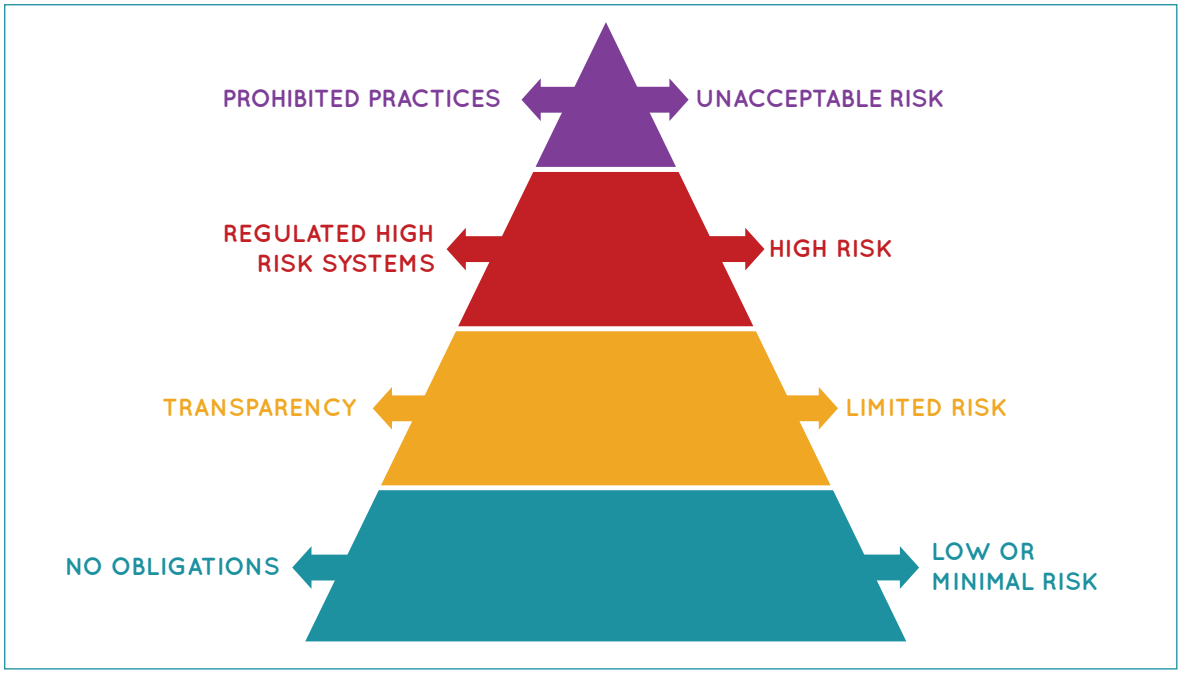
2) Ensure legal certainty to facilitate investment and innovation in AI.

3) Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems.

4) Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

As mentioned above, this legislation is based on a classification of the risks that Artificial Intelligence could pose to the safety, health and fundamental rights of a person in each case, and sets out obligations for providers and users depending on the risks posed by the uses of AI.

Figure 4.



Source: EU Artificial Intelligence Act.

First, **low or minimal risk AI system**s will not be subject to any obligation, and may be developed and used in the EU without any further legal requirements. However, the Act calls for the creation of codes of conduct.

Second, there are **limited risk AI systems**. Those responsible for these systems are subject to general transparency obligations. For example, they will have to present information to report the use of an AI system when it interacts with humans, or if it is used to generate or manipulate images, audios or videos, there will be an obligation to disclose that they have been generated by automated means.

Third, there are **high risk AI systems**. In these cases the draft legislation explicitly states that 'the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI'.

The systems considered high risk are specified in point 1 of article 6 in Title III, and in Annex III. Without listing all the possible cases, some examples include:

- Systems used as safety components of a product or as a product falling under harmonised Union health and safety legislation (toys, aircraft, cars, etc.).

- 'Real-time' and 'post' remote biometric identification of natural persons
- Systems used for the management and operation of critical infrastructure (road traffic, and the supply of water, gas, heating and electricity).
- Systems used to assess students and determine access to educational institutions.
- Systems used to select or screen workers or evaluate candidates.
- Systems used to assess a person's risk of committing a criminal offence.
- Systems used by public authorities to verify the authenticity of travel documents.

All these systems are divided into eight areas that could be updated if necessary: biometric identification and categorization of people; the management and operation of critical infrastructure; education and training; employment and workers management; access to essential private services and public services; law enforcement; migration, asylum and borders; and the administration of justice.

These systems will be subject to a whole range of new rules. The suppliers of these systems will have to register them in an EU database managed by the Commission before marketing them in order to undergo an *ex ante* conformity assessment. In addition, they will have to meet a series of requirements and obligations, specifically related to risk management, testing, technical robustness, training and transparency, human oversight and cybersecurity (Articles 8 to 15).

Finally, there are **AI systems that pose an unacceptable risk**. These systems are completely prohibited, as their use is considered unacceptable due to contravening Union values. The following uses are prohibited:

1) The use of subliminal techniques that transcend a person's consciousness to alter their behaviour or cause physical or psychological harm.
2) AI that exploits any of the vulnerabilities of a specific group of people due to their age or physical or mental disability to alter their behaviour.
3) The use of AI systems by public authorities in order to engage in social scoring of natural persons.
4) The use of AI systems for 'real-time' remote biometric identification of natural persons in publicly accessible spaces (except in specific

cases such as the search for victims of a crime or prevention of a threat).

In addition, as a measure to support innovation, article 53 includes the establishment of a controlled environment for AI tests (*sandboxes*) which will be established by one or more Member States or by the European Data Protection Supervisor. This environment will facilitate the development, testing and validation of AI systems before they are introduced to the market. Spain aims to become a leading country in this area, as will be discussed in the next section.

Finally, the Act calls for the creation of a European Artificial Intelligence Board to act as an advisory board to the Commission in order to contribute to effective cooperation between the national authorities and the Commission, to coordinate and contribute to an analysis on internal market issues, and to assist the national authorities and the Commission on the application of this legislation. This Board will be composed of representatives of the national authorities. These authorities must be established or designated by each State.

5.3 Actions by Spain

So far, in addition to policy documents such as the National Strategy, Spain has issued two decrees of particular interest in terms of the steps the country is taking in this area. Two important decrees were issued in the summer of 2023: first, the decree regulating the application of the *sandbox* in Spain, and second, the statutes of the Supervisory Agency.

Spain's sandbox Royal Decree

In June 2022, the Government of Spain and the European Commission presented the pilot project to launch the EU's first AI regulatory sandbox, which will receive 4.3 million Euros of Recovery and Resilience Funds until 2025.

A year later, in June 2023, the Government published the draft Royal Decree regulating the Artificial Intelligence sandbox. The aim is to provide companies, and especially small and medium-sized companies and start-ups, with a controlled testing environment to ensure the development of responsible and human-centric technology, mitigating potential risks to health, safety and fundamental rights.

Participation in this controlled testing environment is limited to AI systems that are classified as high risk, general purpose AI systems, foundational

models, and generative AI systems. However, the participation of systems for scientific, defence or national security purposes is prohibited.

The ultimate goal of the sandbox is **for the companies developing this type of application to have an environment in which to test and evaluate them, in order to ensure compliance with all the requirements that the future European AI Regulation may demand in order to be able to market their product.**

The Statute of the Spanish Agency for the Supervision of Artificial Intelligence (SASAI)

The Royal Decree approving the Statute of the Spanish Agency for the Supervision of Artificial Intelligence (SASAI), affiliated to the Spanish Ministry of Economic Affairs, was approved in August 2023. The creation of this agency made Spain the first European country to have an institution with these characteristics. The agency will be in charge of supervising the application of European Union regulations on Artificial Intelligence.

6. ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATIONS

6.1 What do we mean by AI in the context of Public Administrations?

The previous sections, which discussed what artificial intelligence is, examined its development, governance and the draft European Union Artificial Intelligence Act, provided an understanding of what we are talking about when we refer to artificial intelligence, focused the debate and placed our current political and social circumstances in context.

At this point, the next step is to focus on an aspect of particular interest to the Third Sector: how is artificial intelligence being applied in Public Administrations? In other words, which public processes and services are incorporating AI in order to improve their effectiveness? This section will use real cases to explain how AI is being used in the public sector and in the field of social rights in particular.

As mentioned above, artificial intelligence has also reached the Public Administrations. However, it is important to distinguish between the digitization process that the public sector is undergoing, and the introduction of artificial intelligence in this area. The digital transformation of Public Administrations consists of a number of technological mechanisms that may or may not include artificial intelligence.

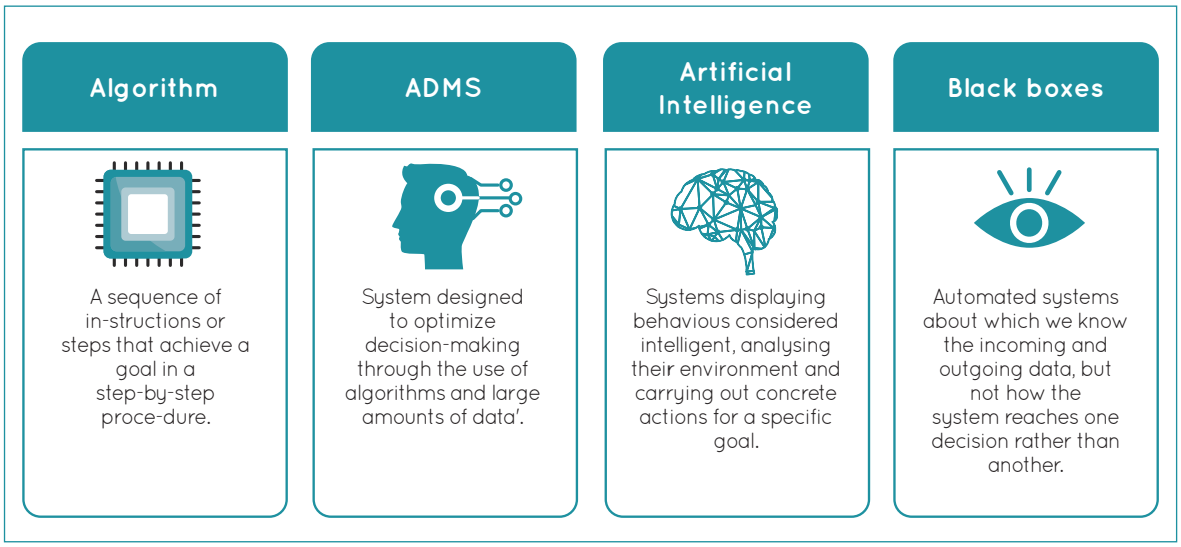
The main type of technology used by the public sector is automated decision systems (ADMS), which are defined as 'processes designed to optimize decision-making through the use of algorithms and large amounts of data'.[8] These systems may use AI (as defined above), although others that do not.

Finally, another concept linked to ADMS is the 'black box', which refers to the situation in which the procedures and variables used by ADMS algorithms are not easy to observe, and therefore not transparent to humans.

As mentioned in the introduction, in this report the concept of AI is therefore used in broad terms to include both ADMS and artificial intelligence, as ADMS are currently far more widespread in the public sector, and present opportunities and risks similar to those offered by AI.

8 *AlgorithmWatch*, 2019, cited by Jiménez Arandia, P., 2023.

Figure 5.



Source: Compiled by the authors based on Jiménez Arandia, P. (2023).

6.2 Applications of AI in Public Administrations.

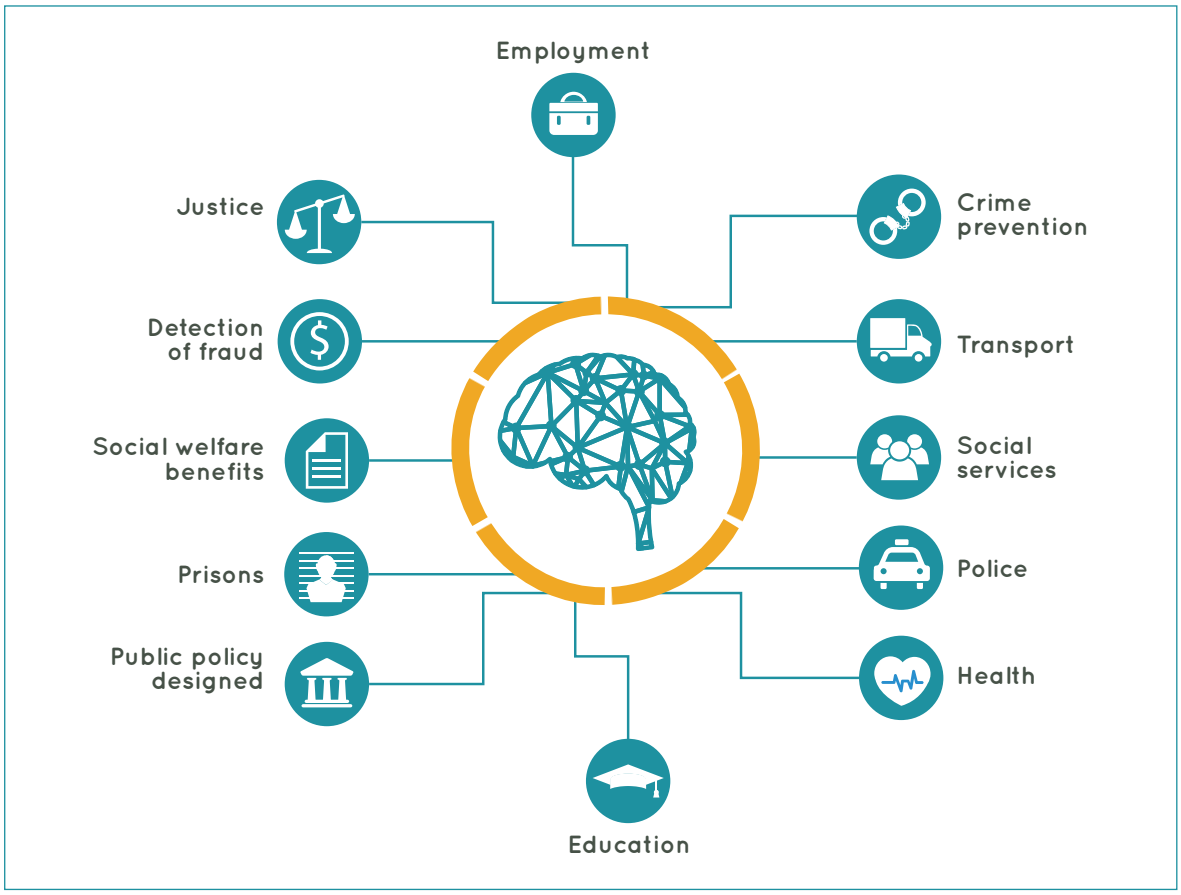
The use of ADMS and AI in Public Administrations has increased in recent years. As mentioned above, the government is linked to AI in two areas of action: first, it works to establish a legislative framework that guides its development; and second, it works to define the tasks in which its implementation can be useful in terms of efficiency and effectiveness. This section will therefore set out the tasks in which Public Administrations have begun working with AI.

International experiences with the initial applications of AI within the public sector show how it can have a major impact on improving internal operations, decision-making and public services, as well as trust in government. This evidence therefore points to the vast potential of AI within the framework of Public Administrations.

As long as ethical principles and those of a democratic State are maintained, artificial intelligence makes the processes currently carried out by the Public Administration more efficient and accurate. As will be discussed in greater detail in Section 7, the many benefits of AI in the public sector can help design better public policies, improve communication and engagement with citizens, and improve the quality and speed of public services. For this reason, it has begun to be implemented to automate public grants (in the case of BOSCO), to detect possible cases of welfare fraud (SyRI, Robodebt), to help the police (VioGén, VeriPol, PredPol), to prevent crime (RisCanvi, COMPAS, LSI-R, OASys), to promote employment, to ac-

cess university (Ofqual) and by the Social Services to identify vulnerable minors (Gladsaxe, Allegheny), among other uses.[9]

Figure 6.



Source: compiled by the authors

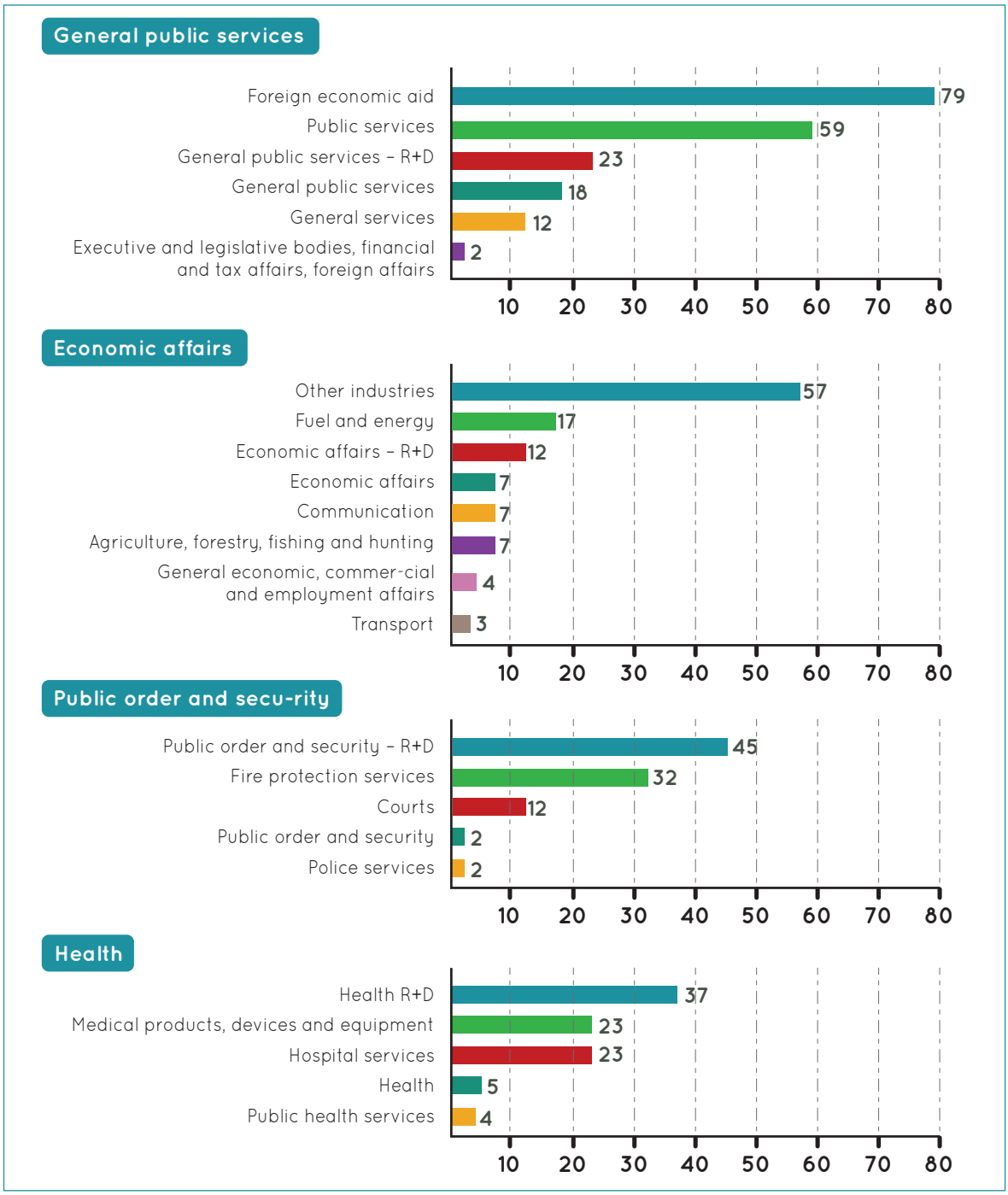
More specifically, according to the information compiled by the *Joint Research Centre* of the European Commission,[10] Ithe main tasks on which work with Artificial Intelligence have begun are in the field of transport; in executive and legislative bodies, financial and fiscal affairs, and foreign affairs; in general public services; in public health services; for police services; for public order and security; and finally, in the field of social welfare. However, as can be seen in Figure 1, the areas where artificial intelligence currently has the strongest presence are foreign economic aid, public services, economic affairs of other industries and R+D in public order and security.

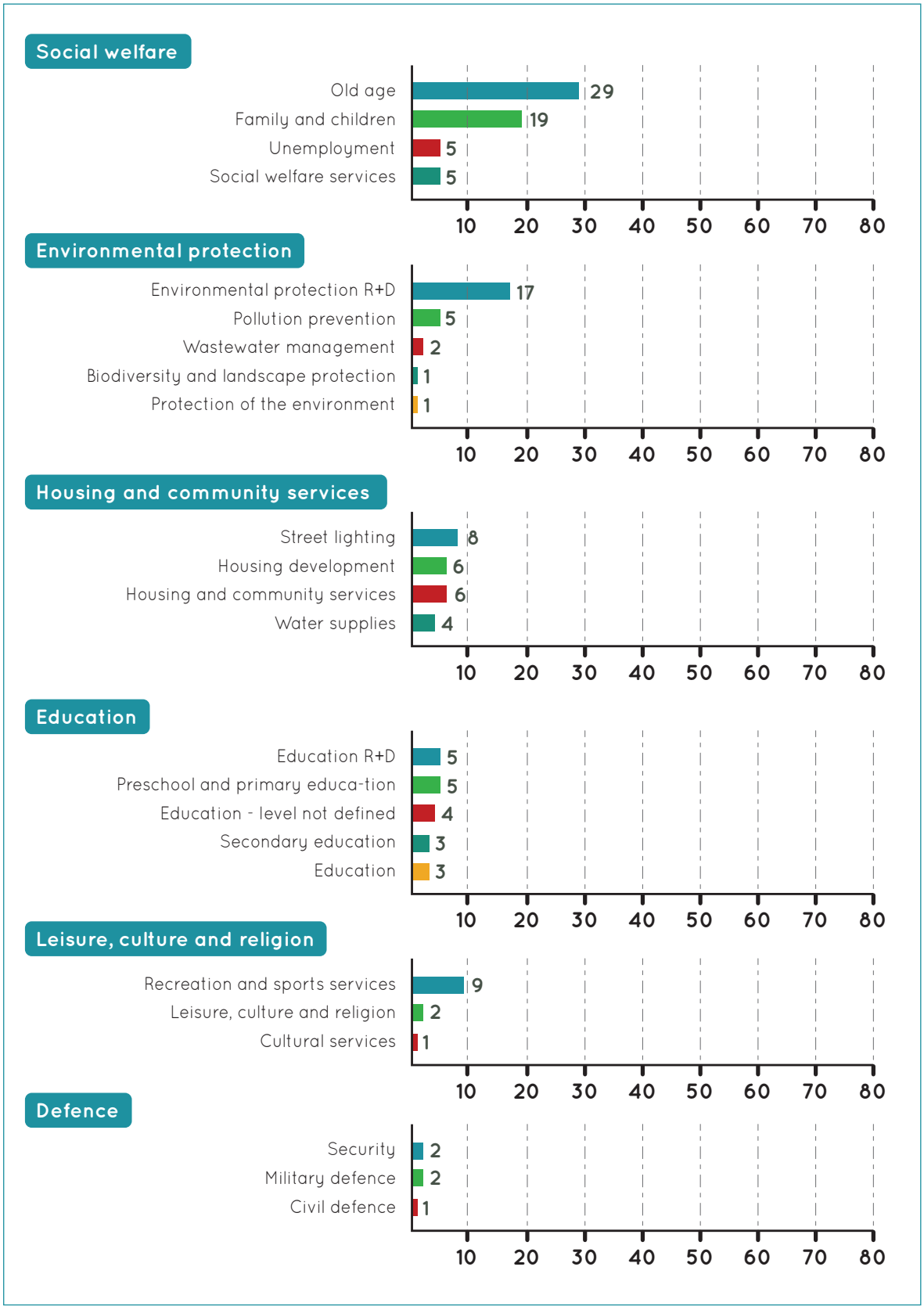
9 The details of the country that has implemented each of the algorithms mentioned are listed below:

Spanish: BOSCO, VioGén, VeriPol, RisCanvi; The Netherlands: SyRI; Australia: Robodebt; United States of America: PredPol, COMPAS, LSI-R, Allegheny; Canada: LSI-R; United Kingdom: LSI-R, OASys, Ofqual; Denmark: Gladsaxe.

10 Available at: https://data.jrc.ec.europa.eu/dataset/7342ea15-fd4f-4184-9603-98bd87d8239a

Figure 7. Number of artificial intelligence algorithms in use in the European Union by field.



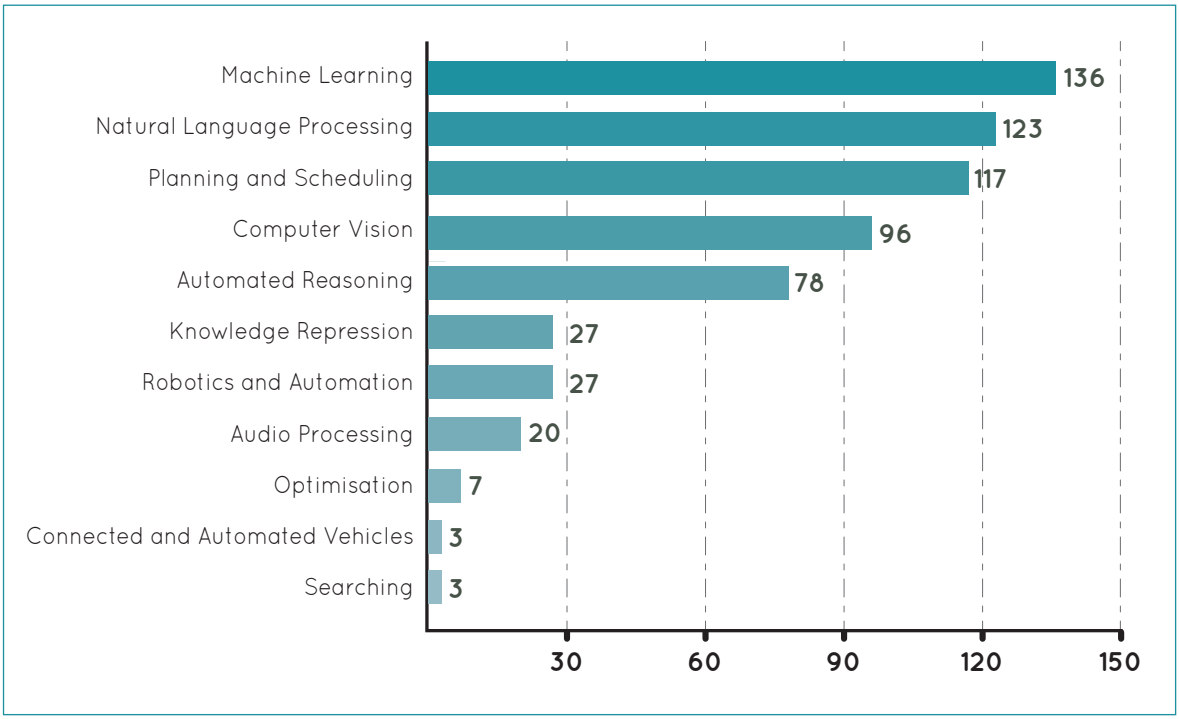


Source: compiled by the authors based on data from the Joint Research Centre.

The most common type of AI in the public sector is machine learning, followed by natural language processing and planning and scheduling algorithms (see Figure 8).

These algorithms are mostly used at a national level, (371 out of a total of 686, or 54.1%), 26.7% are used at a local level, and only 10.1% and 9.2% are used for regional tasks or tasks between different countries respectively.

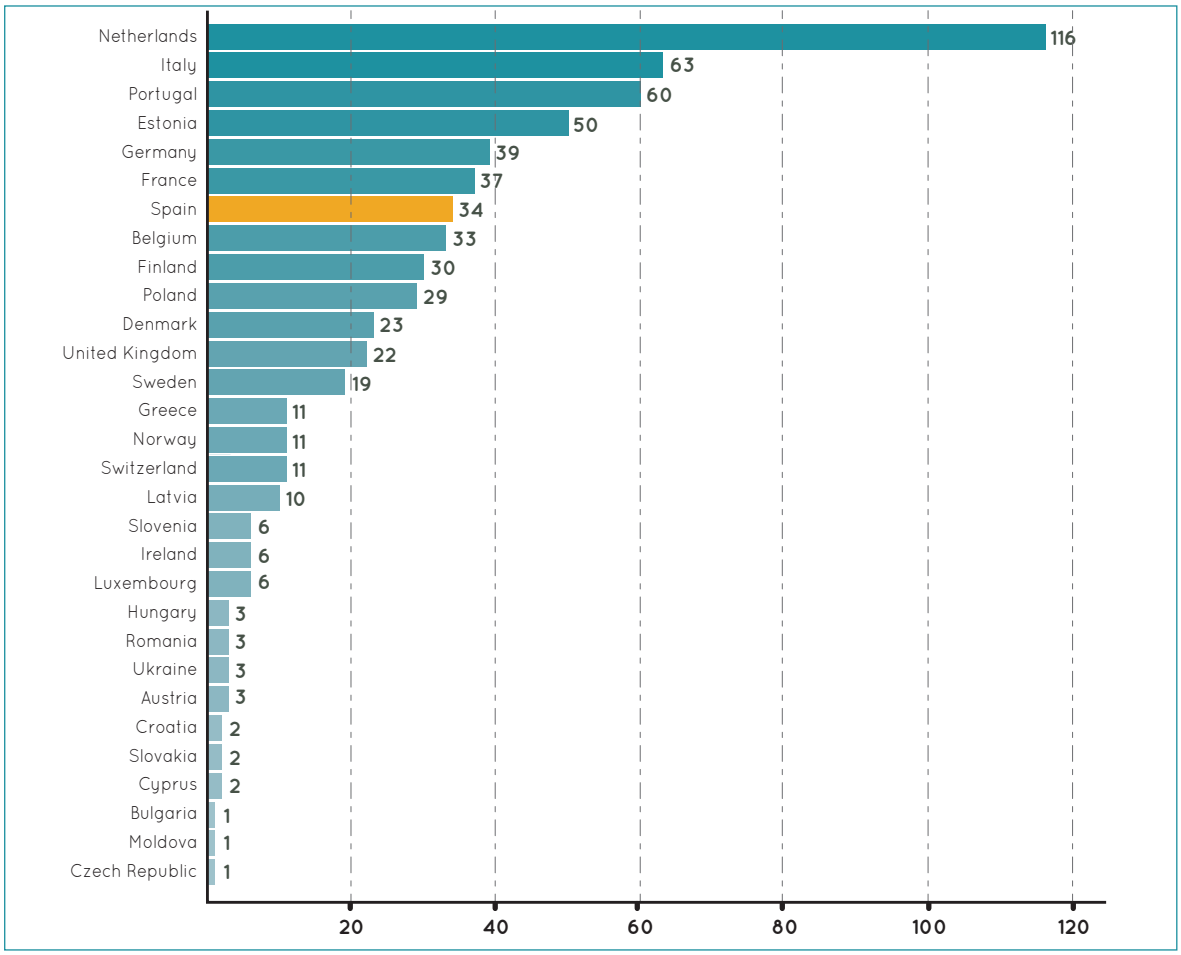
Figure 8. Number of Artificial Intelligence algorithms used in the Public Sector in the European Union, by type..



Source: compiled by the authors based on data from the Joint Research Centre.

In addition, the figure below shows how the three countries with the most algorithms implemented, in development or in the pilot test phase in the public sector are the Netherlands (with 116 algorithms), Italy (with 63 algorithms) and Portugal (with 60 algorithms). Spain is in sixth place, well above the average.

Figure 9. Number of artificial intelligence algorithms used in the European Union by countrys.



Source: compiled by the authors based on data from the Joint Research Centre.

However, the ranking varies when only currently implemented algorithms are taken into account. As a result, and taking into account only the 260 algorithms implemented in the territory of the European Union, the countries that currently work with the most algorithms are Italy (with 38 algorithms), Estonia (with 27) and the Netherlands (with 26).

As a result, AI can be implemented in specific areas including health, public transport, national defence, education, the administration of justice and the social services (see Section 6.4). AI has therefore has great potential to improve the work of Public Administrations - a capacity that will have to be demonstrated with the results of the applications undertaken in the present and the future. However, as will be discussed in more detail in Section 7, the introduction of AI is also associated with various challenges that the public sector must attempt to minimize in order to be able to maximize its benefits. Hence **the importance of establishing a step prior to the implementation of AI that determines whether AI is the best solution for**

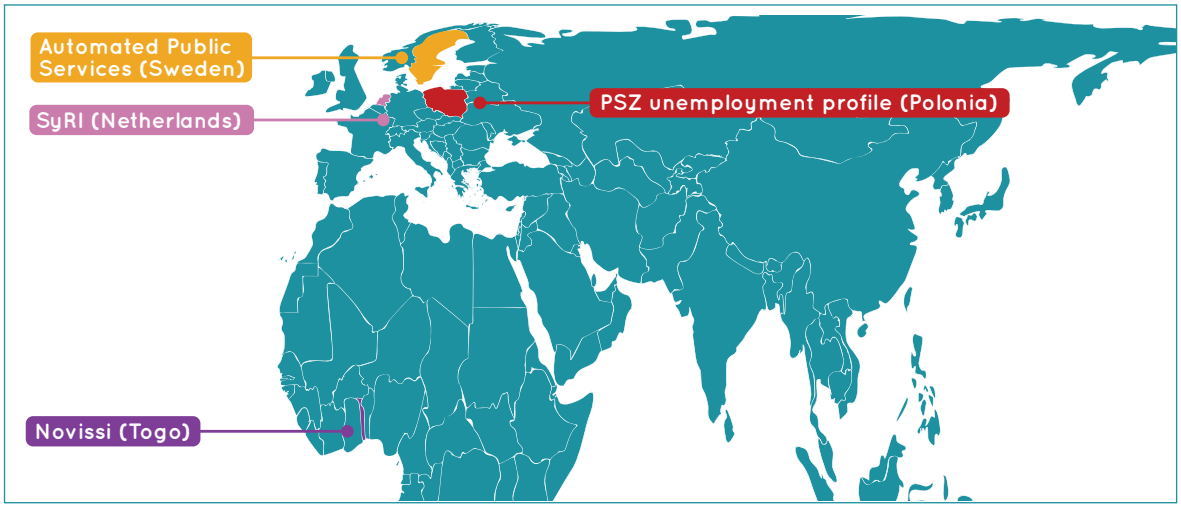
that specific area, by means of an analysis that assesses both the potential benefits and the risks it may pose.

The following cases are international and national experiences that provide examples of good and bad practices and present both the benefits and the risks of AI in Public Administrations. At the same time, these cases highlight the need to make additional efforts to ensure that AI is used in a trustworthy, ethical and manner and as a result, does not have a negative impact on the public's trust in the government. In fact, a study by the Boston Consulting Group research notes that support for governmental AI is correlated with trust in government, and that 'trust in institutions is essential for governments to have the support needed to deploy capabilities of AI'[11]. This trust does not depend solely on the efficiency and effectiveness of public service, but also on the level of satisfaction and perceived justice in its operations.

6.3 International cases

Four different cases are presented below as examples of how AI systems are being implemented in Public Administration processes and services. They have been selected due to their popularity and the volume of information available. Two cases with very positive results (Sweden and Togo) and two others where problems have arisen (Poland and the Netherlands) are also discussed.

Figure 10.



Source: compiled by the authors

11 Carrasco et al., 2019, cited in Berryhill, J., 2019

Sweden (Automated Public Services)

Automation of Public Services in Trelleborg

Country	**Sweden**
Name	**Automation of Public Services**
Implementing Public Administration	Social welfare at a local level
Expected impact	Shorter waiting times, increased efficiency, improvement of citizens' experience
Description of the innovation	The social services in the municipality of Trelleborg (Sweden) have used a programme that uses AI for the automation of various social welfare benefits since 2016. Faced with a situation in which citizens had to wait for an average of 8 days (and sometimes as many as 20 days), and which overloaded the professionals who had to answer their queries and deal with the requests manually, robotic processes automation (RPA) was proposed as the best solution to reduce waiting times and delays in payments to citizens. However, supervision by a professional has been maintained in cases where the application is rejected. By 2020, the system was able to process applications for home care, sickness benefits, unemployment benefits and taxes.
Result and impact	This automated system has significantly reduced waiting times for grant applications and the time taken to deal with cases involving financial vulnerability, and all decisions concerning financial assistance are made within 24 hours. Furthermore, a study on the impact on the professionals shows a positive effect, as artificial intelligence has helped them work more effectively and provided them with legal certainty. This shows the potential of technology and human collaboration to redefine practices in social welfare. However, another report highlighted the importance of ensuring that this technology is trusted by the professionals involved, since they are otherwise reluctant to change. This case therefore shows how process automation and artificial intelligence systems can be very beneficial, provided that the relationship between professionals and these systems is not neglected.

Togo (Novissi)

Novissi	
Country	**Togo**
Name	**Novissi**
Implementing Public Administration	Government of Togo, state level
Expected impact	Access to government emergency measures, money transfers
Description of the innovation	Faced with the emergency created by Covid-19, the Government of Togo wanted to deploy financial aid for families in the greatest need and thereby mitigate the pandemic's negative impact. However, the Government of Togo did not have a complete social register that would enable it to identify the poorest people, and the pandemic made it impossible to compile a new register. The lack of information increased the exclusion of people eligible for the grants. Accordingly, an alternative had to be sought. This alternative, called Novissi, was developed and implemented using machine learning algorithms, satellite data and mobile phone networks. The grants were initially distributed to families that met the following three requirements: 1) They had to be registered on the Novissi platform and have provided basic information from their mobile phone, 2) They had to be registered to vote in a specific region (Greater Lomé), 3) They had to be self-declared as working in an informal occupation in their register of voters. The aim was subsequently to roll out this service to people living in poverty in the country's rural areas.
Result and impact	An evaluation of this programme shows the potential of using new data sources that complement traditional sources to reach a larger amount of eligible people, especially in crisis situations where traditional data may be missing or may not up to date. It highlights the potential of machine learning for processing large amounts of data and combining various data sources, which would not otherwise be possible. In this specific case, the evaluation shows how the flexibility of the machine learning system reduced errors and enabled it to reach more people who needed help. It also reduced the number of people who would have been wrongly excluded from welfare benefits in comparison with the geographic alternative.

Netherlands (SyRI)

SyRI	
Country	**The Netherlands**
Name	**SyRI (Systeem Risico Indicatie)**
Implementing Public Administration	Social welfare at central and municipal levels
Expected impact	Enhanced inspection capabilities, improved child welfare,reduced misuse of public funds
Description of the innovation	In 2012, the Dutch Tax Authority began using self-learning algorithms to create profiles of fraud risk in order to prevent childcare benefit fraud.
Result and impact	After operating for several years, this system was withdrawn due to its obviously negative effects. The algorithm had been developed in such a way that it classified families who had filled out the application documents incorrectly as having made fraudulent benefit claims. Having dual nationality also influenced this profiling. Coming from an unfavourable socioeconomic background, being an immigrant and belonging to an ethnic minority were characteristics that led the algorithm to disproportionately penalize these sectors of the population. The professionals responsible for monitoring the cases also ignored many false alarms and continued to demand the return of the money. As a result, more than 10,000 people fell into poverty, others committed suicide after receiving bills for debts that they were unable to pay, and more than 1,100 children were separated from their families and placed in social care centres. A total of 30,000 families were affected by this algorithm. This led to legal proceedings to investigate the algorithm, which led to the Tax Authority being fined 2.75 million Euros and the resignation of the acting Government.

Poland (Unemployed Profiling)

PSZ (Publiczne Sł uł by Zatrudnienia) unemployment profile	
Country	**Poland**
Name	**PSZ (Publiczne Sł uł by Zatrudnienia)**
Implementing Public Administration	Economic affairs, central and municipal levels
Expected impact	Personalized public services, reduced unemployment, improved efficiency
Description of the innovation	In 2012, Poland's Ministry of Labour and Social Policy wanted to reform its employment offices, which at that time were short-staffed and inefficient. However, the Government of Poland did not have sufficient funds to increase the workforce. As a result, it used an automated profiling system to manage cases of unemployed people more efficiently. The system was designed with the aim of placing unemployed people in one of three categories, taking individual characteristics into account. Each of these three categories is associated with the type of programme that the person is eligible for, i.e. it defines people's eligibility in terms of the range of services available (e.g. employment placement programmes, professional training programmes, and even unemployment benefit). Data such as age, sex, and duration of unemployment were collected during an initial interview, and were then used to categorize each individual. This allocation system therefore defined the level of support and the funds allocated to each person, and one of the categories was no state benefits.
Result and impact	After seven years, this classification system was abolished in 2019 as it was found to be inefficient and patterns of discrimination were identified: being a single mother, having a disability or living in a rural area placed people in a worse category, and they received less assistance from the state's professional services as a result. In addition, 44% of the staff working in the offices said that it was not useful in reducing their day-to-day workload, and 80% said that the system should be changed.

6.4 Applications in the social sphere

As seen above, numerous cases have experimented with using artificial intelligence in Public Administrations, to perform very varied tasks. However, when these innovations take place in the social sphere - and more specifically in social welfare - they are subject to greater interest and need to move forward carefully, as they can have an obviously greater impact on citizens. Due to its nature, this area is particularly sensitive, given that the people who are eligible for these welfare benefits are usually in circumstances involving socio-economic vulnerability. Indeed, social welfare is 'the most visible face of Governments', and as such AI must be introduced very carefully, as there may be considerable risks if it is not implemented correctly.

With an effective implementation of AI (as defined by the European Union, in terms of being human-centric and trustworthy), Public Administrations working in the social sphere can use this tool to assess eligibility and needs, make decisions concerning registration, for the allocation of benefits and to monitor them. At the same time, AI enables the appointments with professionals that the system provides to be those that best meet the citizen's needs. It can also help improve the geographical distribution of users, manage waiting lists and determine the evolution of demand in real time.

The report "Intel·ligència Artificial: Decisions Automatitzades a Catalunya" [Artificial Intelligence: Automated Decisions in Catalonia] by the Catalan Data Protection Authority makes for interesting reading in this regard. Although it presents examples of the application of algorithms in various fields such as health, the legal system and education, it also lists seven applications in the social sphere. They will not all be discussed here, although they will be mentioned. Barcelona City Council deals with an average of 50,000 people a year at its 40 social service centres. These people have very diverse problems, which are addressed by a staff of 700 people. The City Council decided to exploit the information stored in the 300,000 interviews that had been carried out, and applying machine learning techniques, it developed a model capable of analysing new cases, categorising the requests, and providing possible answers.

Finally, to enable oversight in the interventions, the makes the attention provided for each user to be more personalized and comprehensive (e.g. with the Single Electronic Social History of the 'Cohesiona System' of the Andalucia Regional Government). It also anticipates future demands, and enables preventive action to be taken. Two cases of applications of

algorithms in Spain which have had a national impact due to the sensitivity in the area where they work are discussed below.

Spain (VioGén)

VioGén	
Country	**Spain**
Name	**VioGén**
Public Administration implementing it	Ministry of the Interior, at a national level
Expected impact	The ability to monitor and protect women who are victims of gender violence and their sons and daughters anywhere in Spain.
Description of the innovation	In 2007, the Ministry of the Interior of the Government of Spain developed the Comprehensive Monitoring System for cases of Gender Violence (known as VioGén). VioGén was designed in order to determine the level of danger of suffering further attacks, to choose the most appropriate protocols for protecting victims and their children, and to establish a follow-up system to determine and adjust safety measures. However, the final decision is subject to oversight by a professional, who can raise the level of risk but can never lower it.
Result and impact	The system was used to assess 6,047,700 cases between its implementation and May 2023, with 77,213 active cases. The system has also been modified five times in order to better adapt it to new situations. However, this system is not error-free, and cases persist of women who are killed by their partners or ex-partners even after they have reported the problem. At the same time, the algorithm is also opaque and the risk scoring system it uses is unknown.

Spain (BOSCO)

BOSCO	
Country	Spain
Name	BOSCO
Implementing Public Administration	Social Services, at a national level
Expected impact	Improve the regulation of awards of social discount vouchers for payment of electricity bills
Description of the innovation	The BOSCO algorithm was introduced by the Spanish Government and implemented by the country's electricity companies in order to regulate the social discount vouchers applied to payment of electricity bills. Holders of these vouchers are eligible for a discount on their electricity bill.
Result and impact	Despite being totally opaque to the population, an aspect which has been condemned by the Civio Foundation, this algorithm also seems to be inefficient in that eligible people who applied for the voucher ultimately did not receive it. As a result, despite its opacity, concern has been expressed regarding the possibility that this algorithm is discriminating against groups in the population who are entitled to receive it, such as pensioners and widows.

7. BENEFITS AND RISKS OF APPLYING ARTIFICIAL INTELLIGENCE TO PUBLIC ADMINISTRATIONS IN THE SOCIAL SPHERE

In this report we have seen the progress made by governments - with particular emphasis on Spain - in order to ensure that Public Administrations and society as a whole can maximize the benefits and minimize the risks that a future with artificial intelligence may entail. While some of this progress has been discussed, we have not yet considered the benefits and risks we are referring to when we talk about AI. This section therefore presents the benefits and risks that have been identified to date.

As mentioned above, the implementation of artificial intelligence by Public Administrations has a series of associated benefits and risks that must be taken into account, as they can have a direct impact on the exercise of social rights.

Benefits

The **benefits** of this technology are that it provides the Public Sector with numerous opportunities to design better policies and improve decision-making, enhance communication and engagement with citizens, improve speed and quality in the provision of public services, and to replace routine tasks performed by public officials and enable them to carry out tasks with greater added value. AI therefore allows the design and provision of public services to be improved, as well as improving the internal administration of state institutions.

As for **improvements to the design and provision of public services**, artificial intelligence can help Public Administrations to identify interests, concerns and perceptions of various stakeholders so that they can be included in the government's agenda. It can also help to identify problems with access to services, provide a better understanding of citizens' behaviour, and determine how various public institutions impact on specific groups in the population.

At the same time, this technology makes it possible to manage large amounts of data generated by citizens in their interaction with mobile devices and social networks. Public Administration professionals can use this processing to design more personalized services adapted to citizens' real

circumstances and thereby provide a better service for institutions, with greater savings on time and resources.

AI also offers an opportunity to make interactions with citizens more efficient. Citizens can use chatbots to resolve doubts and queries flexibly, improving their levels of satisfaction, provided that they know that there is an algorithm behind the process.

In the second area, **improving the internal administration of state institutions**, artificial intelligence has prediction, optimization and control techniques that can be used to help allocate and manage economic resources, detect and prevent fraud, and prevent these resources from being used inefficiently. It can also help generate new regulations, update them and guarantee compliance. Thanks to its automation capacity, this technology can also be used for repetitive and routine tasks, enabling Public Administration professionals to perform more complex activities that require more emotional skills, creativity and a human perspective. One of the examples where the public service can be automated is data entry through automatic handwriting recognition, voice recognition and natural language processing.

By reducing the workload by eliminating repetitive and mechanical tasks, the administration thereby gains time that can be devoted to faster, more efficient and personalized care, and to improving other processes that need it, or to providing new services.

Finally, the precision and prediction capabilities can be used by Public Administrations to perform diagnoses or define areas of action when dealing with emergencies.

Risks

The **risks** of applying AI in the Public Sector are primarily algorithms' **lack of accountability, transparency and explainability.**

The OECD defines the principles of *accountability*, transparency and explainability as follows:

- **Principle of accountability:** 'AI actors must ensure that their AI systems are reliable'.
- **Principle of transparency and explainability:** 'Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art.'

The major criticisms of this technology are based around its **opacity to be understood by citizens**, in which the complexity of its process plays a significant role. Various cases all over the world have shown how these AI algorithms are often not open source (and therefore not transparent). The lack of transparency consequently makes auditing the algorithms difficult, therefore hindering compliance with the principle of accountability.

At the same time, several algorithms are also not easily understandable by citizens, in breach of the principle of transparency and explainability. According to this second principle, **people who use these solutions must be able to understand how the result was arrived at, and both they and people who do not use them but are affected by them must be able to exercise their right to complain if they are not satisfied**. However, opacity in code like that found in black boxes prevents compliance with this principle. Furthermore, this aspect is closely linked to the principle of accountability, as without a guarantee of transparency and explainability, it is much more difficult to prove that an AI system is trustworthy.

These principles are therefore essential in order to ensure that algorithms do not act against the common good, and that the social rights of citizens are not being violated as a result. Understanding how a result was obtained is essential for identifying biases and patterns of discrimination. As a result, when public actors do not guarantee compliance with these principles, public oversight is reduced, and this particularly affects citizens belonging to minorities and disadvantaged groups, as they cannot demand that their social rights are guaranteed. To prevent this, initiatives based on creating algorithm registries are appearing in cities like Barcelona[12], and aim to make these systems more transparent and trustworthy.

12 More information is available here.

Apart from the failure to comply with these principles, another problem associated with the **amplification of biases and discrimination** is the lack of good quality data. These patterns of discrimination often arise as a result of the algorithms' learning, based on the data they are trained with and which they use to learn. This means that if the data entered is of poor quality and contains errors and discrimination that already exist in society, the algorithms will simply replicate those biases. This means that it is very important to have high quality data that does not perpetuate inequalities by means of the Matthew effect, 'where the privileged gain advantages, while those who are already disadvantaged suffer even more'[13]. At the same time, this bias can foster a process of social polarization and lead to radical behaviour towards certain groups in the population.

Another risk to consider is the **security and privacy of the information used**. The information used by Public Administrations is extremely sensitive, and any leakage of information or cyberattacks could have serious consequences for both the public institutions and for citizens. In this respect, it should also be noted that the Data Protection Law[14] can be an impediment to training some models if the security and privacy of the information used is not guaranteed.

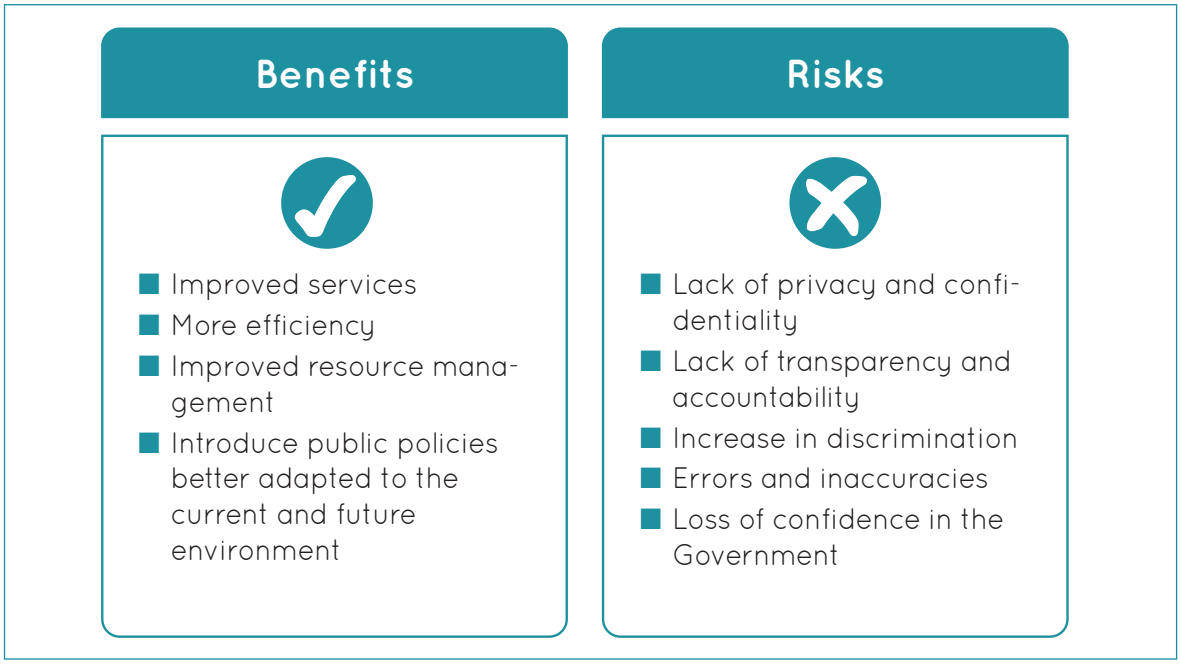
As a consequence, all these risks may affect the effectiveness of public policies, as possible situations involving exclusion of the target population could reduce their impact. All these risks ultimately have a negative impact on the trust that citizens place in Public Administrations. These errors and the perpetuation of inequalities end up eroding the population's trust in the effectiveness of the work of the Public Administrations, with direct implications for the willingness of citizens to participate and ultimately for the state of democracy.

In short, the main benefits of AI include: **improvements in resource efficiency** (both financial and personal), **an improved provision of public services and a better adaptation of the Agenda's public policies and programmes**. However, the risks are primarily linked to the current **lack of transparency in these processes, the lack of adequate data to prevent bias and discrimination, the sensitivity of the data processed and safeguards related to its privacy, situations of false positives** and errors that have serious consequences for the people affected and as a result, ultimately lead to **citizens' loss of confidence in the work done by the Public Sector**.

13 Herzog, 2021, cited in OECD, 2023, p.8

14 Spanish Organic Law 3/2018, of 5 December, on personal data protection and the guarantee of digital rights.

Figure 11.


Benefits
Improved services
More efficiency
Improved resource management
Introduce public policies better adapted to the current and future environment
Risks
Lack of privacy and confidentiality
Lack of transparency and accountability
Increase in discrimination
Errors and inaccuracies
Loss of confidence in the Government

8. CONCLUSION

Artificial intelligence has the potential to significantly transform our society. An increasing number of companies and public administrations will integrate systems that use Artificial Intelligence into their products and processes as they seek to become more efficient, effective or beneficial.

The European Union and national governments have been working for years on regulations to limit the potential for development, in the hope of minimizing and preventing risks without affecting innovation and progress in the field of AI. This regulation was anticipated to reach its final form in late 2023.

In this context of the present and future, it is essential that institutions in the Third Sector establish a shared position and discourse, and convey it to both the political and public spheres. Most regulations will probably be enacted in the next few years, but the worldview of artificial intelligence will also be consolidated. **It is therefore essential for the Third Sector to be an actor with a voice (and a vote) if it wants to participate in shaping future society.**

Third Sector institutions are currently at a crucial point in terms of adopting a proactive strategy and making their voice heard during the drafting of upcoming legislation. At the same time, institutions must be able to act as a bridge between citizens and Public Administrations to raise awareness of the developments that are available to them, and inform them of how they can benefit while avoiding all the risks.

Furthermore, having a common vision as a Third Sector reinforces the work done in the oversight of these applications in the field of social rights and at the same time, enables new agreements to be reached with various agents in order to take advantage of the opportunities that this technology offers.

In short, **the Third Sector is facing a unique opportunity to adopt a common position, and by doing so reinforce its voice and continue working to guarantee social rights in this changing new environment.**

REFERENCES

— Ahn, M. J., & Chen, Y.-C. (2020). Artificial Intelligence in Government: Potentials, Challenges, and the Future. In S.-J. Eom & J. Lee (Eds.), Dg.o '20: The 21st Annual International Conference on Digital Government Research (pp. 243–252). Association for Computing Machinery. https://doi.org/10.1145/3396956.3398260

— AI Watch (2022). Spain AI Strategy Report. Comisión Europea. Consultado el día 21/06/2023. Available at: https://ai-watch.ec.europa.eu/countries/spain/spain-ai-strategy-report_en

— ADB (2020). AI in Social Protection – Exploring Opportunities and Mitigating Risks. Recuperado de: https://www.adb.org/publications/ai-social-protection-exploring-opportunities-mitigating-risks

— Abu Elyounes, D. (10 de febrer de 2021). Why the resignation of the dutch government is a good reminder of how important it is to monitor and regulate algorithms. Medium. Retrieved on 25 July 2023, from https://medium.com/berkman-klein-center/why-the-resignation-of-the-dutch-government-is-a-good-reminder-of-how-important-it-is-to-monitor-2c599c1e0100

— Algorithm Watch; Bertelsmann Stiftung. (January 2019). Bertelsmann Stiftung. Automating Society. Taking Stock of Automated Decision-Making in the EU. Berlin: AW AlgorithmWatch gGmbH, January 2019. Retrieved from: https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

— Belinchón, F. (2023). La primera ley europea sobre inteligencia artificial se amplía para incluir a ChatGPT y entra en su recta final. El País. Available at: https://cincodias.elpais.com/economia/2023-06-14/la-primera-ley-europea-sobre-inteligencia-artificial-se-amplia-para-incluir-a-chatgpt-y-entra-en-su-recta-final.html

— Berryhill, J. *et al.* (2019), 'Hello, World: Artificial intelligence and its use in the public sector', OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, https://dx.doi.org/10.1787/726fd39d-en.

— Big Data Magazine (2022). Algoritmos de la Administración Pública espanyola: poco conocidos, opacos y sin auditar. Retrieved on 25 July 2023, from: https://bigdatamagazine.es/algoritmos-de-la-administracion-publica-espanola-poco-conocidos-opacos-y-sin-auditar

— Blazquez, L. (20 February 2023). What is VioGén and how does it work?. Rtve. Retrieved on 25 July, from: https://www.rtve.es/noticias/20230220/objetivo-igualdad-funcionamiento-viogen/2424326.shtml

— Boucher, P. N. (2020). Artificial intelligence: How does it work, why does it matter, and what can we do about it?. Scientific Foresight Unit (STOA), European Parliament.

— Carrasco, M., Mills, A. Whybrew and A. Jura (2019), 'The citizen's perspective on the use of AI in government', BCG, 1 March, www.bcg.com/publications/2019/citizen-perspectiveuse-artificial-intelligence-government-digital-benchmarking.aspx

— De Schutter, O. (2022). Non-take-up of rights in the context of social protection. Human Rights Council. Obtained from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/322/20/PDF/G2232220.pdf

— European Commission (2019). Building Trust in Human-Centric Artificial Intelligence. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. COM(2019) 168. Brussels. Available at: https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence

— Felton, J. (11 May 2022). The ai that led to children being re-homed and the fall of an elected government. IFL Science. Retrieved on 25 July 2023, from https://www.iflscience.com/the-ai-that-led-to-children-being-rehomed-and-the-fall-of-an-elected-government-63622

— Gleb Papyshev & Masaru Yarime (2023) The state's role in governing arGleb Papyshev & Masaru Yarime (2023) The state's role in governing artificial intelligence: development, control, and promotion through national strategies, Policy Design and Practice, 6:1, 79-102, DOI: 10.1080/25741292.2022.2162252

— Heikkilä, M. (29 March 2022). Dutch scandal serves as a warning for Europe over risks of using algorithms. Politico. Retrieved on 25 July 2023, from https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

— High-Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

— Jimenez Arandia, P. (2023). Govern Obert: Transparencia algorítmica en el sector público. (rep.) Vol.9. Generalitat de Catalunya. Generalitat de Catalunya. Retrieved from: https://governobert.gencat.cat/web/.content/01_Que_es/04_Publicacions/colleccio_govern_obert/GovernObert_9/Govern-Obert-9-cast.pdf

— KOSA AI (2021). The importance of AI governance and 5 key principles for its guidance. Medium. Available at: https://kosa-ai.medium.com/the-importance-of-ai-governance-and-5-key-principles-for-its-guidance-219798c8f407

— Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications policy, 44(6), 101976.

— La Moncloa (19 May 2023). El Sistema VioGén supera los seis millones de valoracions de riesgo a víctimes de violència de genero. Retrieved on 25 July 2023, from: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2023/190523-sistema-viogen-victimas-violencia-genero.aspx

— Lokshin, M., & Umapathi, N. (23 de febrero de 2022). AI for social protection: mind the people. Brookings Institution. Retrieved on 28 June 2023, from https://www.brookings.edu/articles/ai-for-social-protection-mind-the-people/

— Marilín Gonzalo (2023). Ley de IA en la Unión Europea: estas son las claves de la propuesta. Newtral. Available at: https://www.newtral.es/ley-ia-union-europea-inteligencia-artificial-europa/20230520/

— Minguijon J. y Serrano-Martinez C. (2022). La Inteligencia Artificial en los Servicios Sociales: estado de la cuestión y posibles desarrollos futuros. Cuadernos de Trabajo Social, 35(2), 319-329. https://doi.org/10.5209/cuts.78747

— Misuraca, G., & Van Noordt, C. (2020). AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. JRC Research Reports, (JRC120399).

— Misuraca, G., and van Noordt, C. (2020). Overview of the use and impact of AI in public services in the EU, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399

— Spanish Ministry of Economic Affairs and Digital Transformation (2020). National Artificial Intelligence Strategy. Available at: https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx

— Niklas, J. (16 April 2019). Poland: Government to scrap controversial unemployment scoring System. Algorithm Watch. . https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/

— NL Times. (17 December 2020). Parents faced 'unprecedented injustice' for years in childcare subsidy scandal. NL Times. Retrieved on 25 July 2023, from https://nltimes.nl/2020/12/17/parents-faced-unprecedented-injustice-years-childcare-subsidy-scandal

— OASI (s/d). Discriminación algorítmica. Retrieved from: https://eticasfoundation.org/oasi/social-impact-algorithms/

— National Observatory of Technology and Society (2023). Uso de inteligencia artificial y big data en las empresas españolas. State Secretariat for Digitalization and Artificial Intelligence, Spanish Ministry of Economic Affairs and Digital Transformation. Available at: https://www.ontsi.es/es/publicaciones/uso-de-inteligencia-artificial-y-big-data-en-las-empresas-espanolas

— OECD (2023). Global Trends in Government Innovation 2023, OECD Public Governance Reviews, OECD Publishing, Paris https://doi.org/10.1787/0655b570-en.

— OECD (2021). 'Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems', OECD Digital Economy Papers, No. 312, OECD Publishing, Paris, https://doi.org/10.1787/008232ec-en.

— OECD (s/d). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449

— ONTSI (2022). Uso de inteligencia artificial y big data en las empresas españolas. Available at: https://www.ontsi.es/es/publicaciones/uso-de-inteligencia-artificial-y-big-data-en-las-empresas-espanolas

— Our World in Data (2022). Artificial intelligence is transforming our world — it is on all of us to make sure that it goes well. Available at: https://ourworldindata.org/ai-impact

— Our World in Data (2022). The brief history of artificial intelligence: The world has changed fast – what might be next?. Available at: https://ourworldindata.org/brief-history-of-ai

— Ranerup, A., & Henriksen, H. Z. (2022). Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden's Social Services. Social Science Computer Review, 40(2), 445–461. https://doi.org/10.1177/0894439320980434)

— Roser, M. (2022a). Artificial intelligence is transforming our world — it is on all of us to make sure that it goes well. Our World in Data. Available at: https://ourworldindata.org/ai-impact

— Roser, M. (2022b). The brief history of artificial intelligence: The world has changed fast – what might be next?. Our World in Data. Available at: https://ourworldindata.org/brief-history-of-ai

— Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008

— Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F. (2022). AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Publications Office of the European Union, Luxembourg , ISBN 978-92- 76-53058-9, doi:10.2760/39336, JRC129301.

— Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., & Pignatelli, F. (2022). AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector (No. JRC129301). Joint Research Center.

— Turkish Policy Quarterly. (7 March 2022). AI and democracy: Towards better public oversight. Turkish Policy Quarterly. http://turkishpolicy.com/article/1108/ai-and-democracy-towards-better-public-oversight

— Vélez, M. I., Gómez Santamaría, C., & Osorio Sanabria, M. A. (2022). Conceptos fundamentales y uso responsable de la inteligencia artificial en el sector público. Report 2. Caracas: CAF. Retrieved from: https://scioteca.caf.com/handle/123456789/1921

— Veale, M., Matus, K., & Gorwa, R. (2023). AI and Global Governance: Modalities, Rationales, Tensions. Annual Review of Law and Social Science, 19.

— Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. International Journal of Public Administration, 42(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

DOCUMENT 2

RADAR OF ALGORITHMS AND AUTOMATED DECISION-MAKING PROCESSES FOR CITIZENS' ACCESS TO SOCIAL RIGHTS

DECEMBER 2023

1. INTRODUCTION

The first document in this project focused on the preparation of the report 'Artificial Intelligence in Public Administration'. This report was produced in order to provide Third Sector institutions with a better understanding of the phenomenon of artificial intelligence (AI), in terms of its historical, technical and governance aspects and in its current applications, especially in the public and social spheres. It was therefore an informative report aimed above all at the Third Sector and non-specialist readers.

In this second document in the project, the priority has been to identify artificial intelligence systems and automated processes in the social sphere, which are being used by the Catalan Public Administrations or are in the pilot phase.

AI is at the centre of contemporary debate. The most recent launches, and those that are yet to come, provide a glimpse of a potential transformation in social relations and the labour market that will be global in its scope. **Since 2022 we have undoubtedly seen a breakthrough in terms of launches of amazing models and tools, and have had little time to adapt to them or to consider their consequences.**

The aim of this project is to produce a collection - also known as a radar or repository - of all the algorithms, artificial intelligence models and/or automated decision processes that are being used in Catalonia in the field of access to social rights.

The Radar will be aimed at citizens, public administrations and third sector institutions in order to provide clear and comprehensive information about the impact of this technology and access to social rights in Catalonia. This is very important, since **the intended use of the repository will determine decisions such as the information to be collected and the language to be used, among other aspects.**

This **Radar** was created in three steps. First, the radar itself was designed. In other words, the variables it would include and which information would therefore be collected for each algorithm. This exercise was carried out using other international repositories as benchmarks.

Second, the fieldwork was conducted. After designing an interview questionnaire and launching an online form, public administration officials who were considered likely to have information on the subject were contacted directly. Those people were given the alternative of filling out the form di-

rectly, or having an interview. In order to increase the likelihood of answers, the decision was taken to make the form public and disseminate it using social media.

Third, after compiling the responses, the information was processed, cleaned and completed to create the Algorithm Radar.

The major difficulty in creating this radar was the lack of knowledge of the population universe. In other words, **the total number of AI systems currently implemented or in the development process which are being used by Catalan public administrations is unknown**. However, in the interviews with managers of various administrations, they clearly reported that in the field of social rights, there are few cases in which these algorithms are being used. **A total of 12 algorithms from 4 different administrations were mapped.**

However, **the Algorithms Radar is the first phase in a more extensive project**. In this phase, the primary objective was to map all the algorithms that were being used.

This document is therefore divided into three sections:

(1) An explanation of the methodology used to compile all the information about the algorithms used by the Public Administration in Catalonia.

(2) An explanation of the variables that make up the Radar, i.e. the fields in the social sphere for which information was collected.

(3) A presentation of the records prepared for each of the 12 algorithms collected.

2. METHODOLOGY

This phase consisted of mapping the various AI algorithms and automated decision processes that are used by the Catalan Public Administrations. Given that they are fairly recent and little-known systems, a combined methodology had to be used when carrying out this mapping. In specific terms, the research team created a form including all the fields and variables of interest, while interviews were conducted with people identified as relevant.

A total of 28 people were initially identified. These people were considered relevant due to meeting two requirements: first, they work in Catalan Public Administrations; and second, they have direct knowledge or have participated in the design or implementation of an algorithm that has been or is in use in the Public Administration. These people are therefore the Radar's primary source of information.

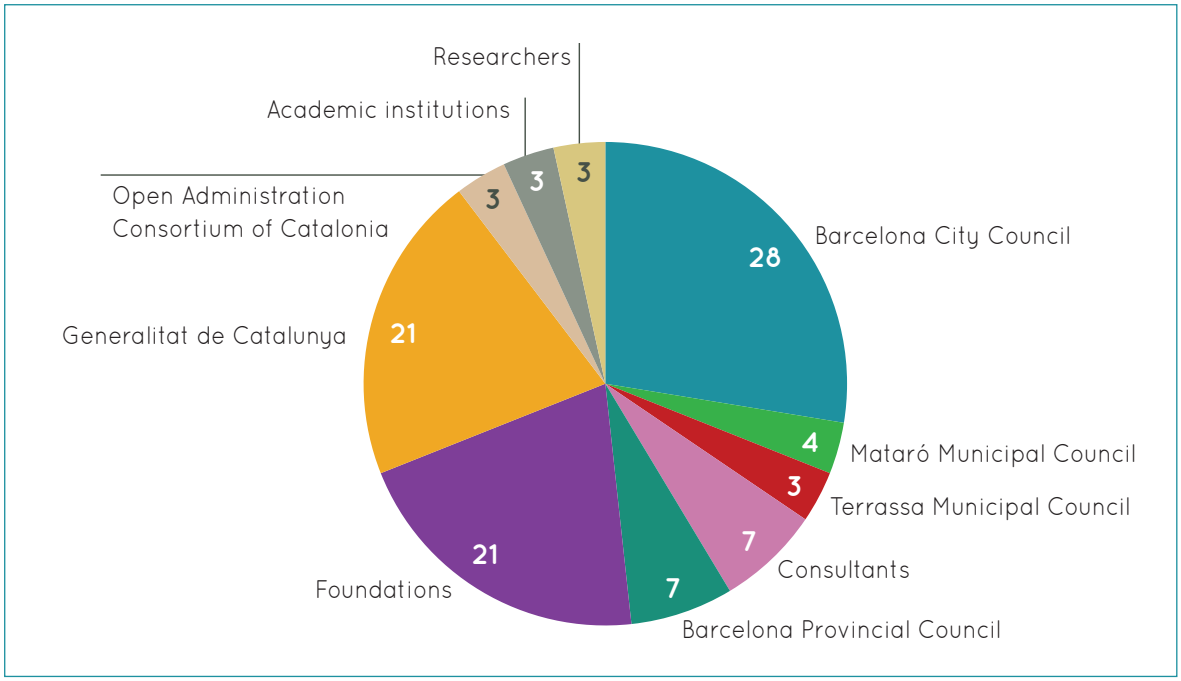
They were largely identified thanks to the on-the-ground knowledge of the Platform of Third Social Sector Institutions, which was used to draw up a list of people who could potentially know of use cases. Some of these people provided contact details of other people in management positions who had not previously been identified. This snowball dynamic provided the basis for creating the final list of people contacted. As can be seen in the figure below, the majority of those contacted work at Barcelona City Council (28%), the Generalitat de Catalunya (21%), and various foundations (21%).

After these individuals had been identified, they were contacted and asked to cooperate. When they were contacted, they were offered two alternative ways to provide information to the research team: either filling in the form and/or arranging an interview to resolve their doubts or to discuss specific issues related to the information to be collected.

At the same time as this group of individuals identified as relevant was compiled, the form was disseminated by means of social media in order to maximize the number of algorithms collected, and thereby create as comprehensive a repository as possible.

A total of thirteen responses to the form were obtained, and several interviews with different people and institutions were carried out. Interviews were organized with professionals working at Barcelona City Council, the Barcelona Municipal Institute of Social Services, the Generalitat de Catalunya, the Open Administration Consortium of Catalonia (AOC) and the ICT Social Health Foundation.

Figure 12. Organizations where the contacted persons work



The information compiled in the Radar was provided by the majority of the people contacted who filled out the form. However, the DPR system was recorded by three different experts, and their responses were therefore combined into one. The record for RisCanvi was very brief, and as a result it was recreated in its entirety based on the information that was publicly available. Modifications and minor adjustments were made to the wording for the other systems.

3. THE DESIGN OF THE ALGORITHMS RADAR

The radar or repository was designed taking both international standards and the decisions of the KSNET research team into account. The former include the AI-WATCH repository Selected AI cases in the public sector at the European Commission's Joint Research Centre[15] and the Ethics Foundation's OASI Register[16]. However, the Amsterdam and Helsinki AI registries were also taken into account to a lesser extent[17].

The first two cases mentioned above are databases used for recording the information collected for the various algorithms, and the first – the AI-WATCH repository of the European Commission's Joint Research Centre – focuses on Europe, and the second – the Ethics Foundation's OASI Register, includes international cases outside Europe.

As a result, comparability with these two databases was sought with the inclusion of shared fields such as the objective, status and benefits (AI cases) and risks and audit (OASI).

The table below lists the fields included in the radar, which were completed by the experts contacted through the online form, or by KSNET based on the information provided by the experts during the interview.

3.1. Fields in the Algorithms Radar

The fields in the Radar are listed below. Information for each field was collected using a specific question in the questionnaire.

Section 1. Basic information

1. Namer

Internal or public name of the algorithm, if any.

2. Description

Context in which this automated process/algorithm/AI system has been developed and applied. In other words, the problem that had been iden-

15 Link: https://data.jrc.ec.europa.eu/dataset/7342ea15-fd4f-4184-9603-98bd87d8239a.

16 Link: https://eticas.tech/oasi.

17 Amsterdam: https://algoritmeregister.amsterdam.nl/. Helsinki: https://ai.hel.fi/en/ai-register/.

tified, the solution that was chosen, how it was implemented and what exactly this consisted of, and the results to date.

3. Objective

What is the objective of the automated process/algorithm/AI system? What tasks was it developed for?

3.1 Objective

Select objectives that the algorithm fulfils. The options are taken from *Selected AI cases in the public sector* by the European Commission's Joint Research Centre.

Automate tasks	Automated performance of a set of tasks that would take a person much longer to perform.
Collect personal data	Systematic or predetermined collection of data about individuals and/or groups for publicly known or unknown purposes and based on publicly known or unknown criteria.
Evaluate human behaviour	Generate assessments of the way individuals and/or groups behave based on publicly known or unknown criteria applied to publicly known or unknown data.
Recognize facial features	Identify specific facial features in images of people, such as the shape of the eyes while a person is smiling, based on publicly known or unknown criteria applied to publicly known or unknown data.
Identify images of faces	Compare images of individual people's faces to images of faces previously registered in a database based on publicly known or unknown criteria applied to publicly known or unknown data.
Predict human behaviour	Generate possible future scenarios in which individuals and/or groups may behave based on publicly known or unknown criteria applied to publicly known or unknown data.
Produce profiles and classify people	Generate profiles of individuals and/or groups and classify and order them based on publicly known or unknown criteria applied to publicly known or unknown data.
Simulate human speech	Generate speech that closely resembles the way people speak for publicly known or unknown purposes.
Image recognition	Identify the content of digital images, e.g. whether it is an image of a cat or a dog, based on publicly known or unknown criteria applied to publicly known or unknown data.
Generate automatic translations	Automatically translate texts or written speeches from one language to another.
Generate online search results	Produce an ordered list of websites or other online resources in response to a search query, usually as written or spoken search terms.

Sound recognition	Identify the content of speech or other sounds, e.g. whether it is a person speaking or a specific animal or object, based on publicly known or unknown criteria applied to publicly known or unknown data.
Chatbot	Provide automated responses to inquiries from the public.

Section 2. Development and implementation

4. Administration responsible

Public Administration that is using or will use the algorithm.

5. Developing institution

The institution developing the algorithm at a technical level. The "developing institution" refers to the company or individual that created the AI system or automated process.

6. Status

What is the status of the implementation?

- Being defined: a need has been identified and the function (and design) of the algorithm is being defined
- Planned: the system has been designed, but is not yet being developed
- In development: the system is being developed
- Pilot: the system is being tested
- Implemented: the system has been implemented and is in use
- No longer in use
- Other: ...

7. Start date

If the algorithm is being used, the year it was first used. If it is in the pilot phase, when the test started. If planned or in development, the estimated date when it will begin to be used.

7.1 End date

If the algorithm is no longer used, in what year was it discontinued? If it is being pilot tested, when does the test end?

Section 3. Group, benefits and risks

8. Group

Group which the algorithm is aimed at, or where the algorithm has the greatest direct impact. For example, the algorithm could be implemented in a process that only involves minors.

9. Benefits

Main benefits for both the Administration and for citizens obtained by implementing this AI algorithm/system.

9.1 Benefits

Select the benefits provided. The options are grouped into three categories, and are taken from *Selected AI cases in the public sector* by the European Commission's Joint Research Centre.

[a] Improved public services

- Personalized services (the range of services is better suited to the population's needs) [a]
- Public services (citizen-centred: increased public accessibility to public services)
- Better quality of information and public services
- More responsive, efficient and cost-effective public services
- New services or channels

[b] Improved administrative efficiency

- Reduced costs
- Response capacity of government operations
- Improved management of public resources
- Enhanced quality of processes and systems
- Improved co-operation and improved communication
- Reduced or eliminated risk of corruption and abuse of the law by public officials
- Greater fairness, honesty and equality

[c] Open government capabilities

- Greater transparency in public sector operations
- Increased public participation in government actions and policy formulation
- Improved public oversight and influence on government actions and policies

10. Risks

In your opinion, what are the main risks for both the Administration and the citizen of the implementation of this AI algorithm/system?

10.1 Risks

Select the risks that the use of this AI algorithm/system presents or could entail. Categories taken from the Ethics Foundation OASI repository.

- Gender discrimination
- Racial discrimination
- Religious discrimination
- Socio-economic discrimination
- Another type of discrimination
- Social polarization / radicalization
- State surveillance
- Threat to privacy
- Create addiction
- Manipulation / changes in behaviour
- Spread wrong information
- Other: ...

Section 4. Supervision and transparency

11. Human intervention

What role does human supervision play?

- The result of the algorithm is supervised by a person who has the final decision
- A person looks at the result of the algorithm, but that person cannot change the decision

- The result of the algorithm is monitored/consulted by a human, but they do not have the knowledge/ability to evaluate it
- The result is not monitored/queried
- Other: ...

11.1 Human intervention (details)

If the answer given is the third option, explain why it is not supervised. If the answer given is an alternative option, provide more details about the level of human intervention.

12. Auditing

Has the algorithm passed an audit system? Select an option if it has undergone an audit:

- An internal audit has been or is currently being conducted
- An external audit has been or is currently being conducted
- An external audit will take place soon (planned)
- An external audit will take place soon (planned)
- Other: ...

13. Transparency

Degree of accessibility to the information in this algorithm. Is there a public document and/or website that sets out the following information? (Tick as appropriate)

- Data used to train the algorithm
- Type of algorithm used
- Problem to which a solution is sought
- How it is being implemented
- Objectives and population interacting with the algorithm
- Results obtained to date
- Other: ...

14. Public visibility

Are citizens who are directly or indirectly affected by the decision/result of the algorithm aware that this algorithm is being used? This does not mean that the citizen actually knows about it, but rather whether the information has been made public, is easily accessible and is clearly visible.

- The citizen is informed directly (active communication)
- The information is public and accessible (passive communication)
- The information is public but not easily accessible (it depends on the citizen's proactivity to find it)
- The citizen is not informed

15. Links and sources

Is there a public document and/or website that contains information on the use of this algorithm?

4. RESULTS

The case files, i.e. all the information collected for the **12 artificial intelligence systems identified**, are presented below. The information from each system is summarized in a table in which each row represents one of the fields explained above.

The following table lists the names of the algorithms together with the administration that used them:

	ALGORITHM	ADMINISTRATION
1	**System for identifying demands, problems and responses (DPR)**	Barcelona City Council and the Municipal Institute of Social Services (belonging to the City Council)
2	**Incident classification support system – MARIO**	Barcelona City Council
3	**IRIS Case Processor**	
4	**Chat systems, infoCanal Alzheimer and the 0-16 Childhood Fund**	
5	**Transcription of social emergency reports (Barcelona Social Emergency Centre - CUESB)**	
6	**RisCanvi**	Generalitat de Catalunya - Information and Communication Technologies Department
7	**Social Benefits Regulation Engine**	
8	**Benefits Advisor**	
9	**Identification of Social Intervention needs**	
10	**Conversational chatbots**	Open Administration Consortium of Catalonia (AOC)
11	**Video identification service to obtain the idCAT Mòbil identification document**	
12	**Energy poverty reports automation service**	

Six of the twelve AI systems mapped operate at the local level, and six at the regional level. The only municipality with AI systems is Barcelona. Furthermore, seven have been implemented and the rest are in the pilot phase.

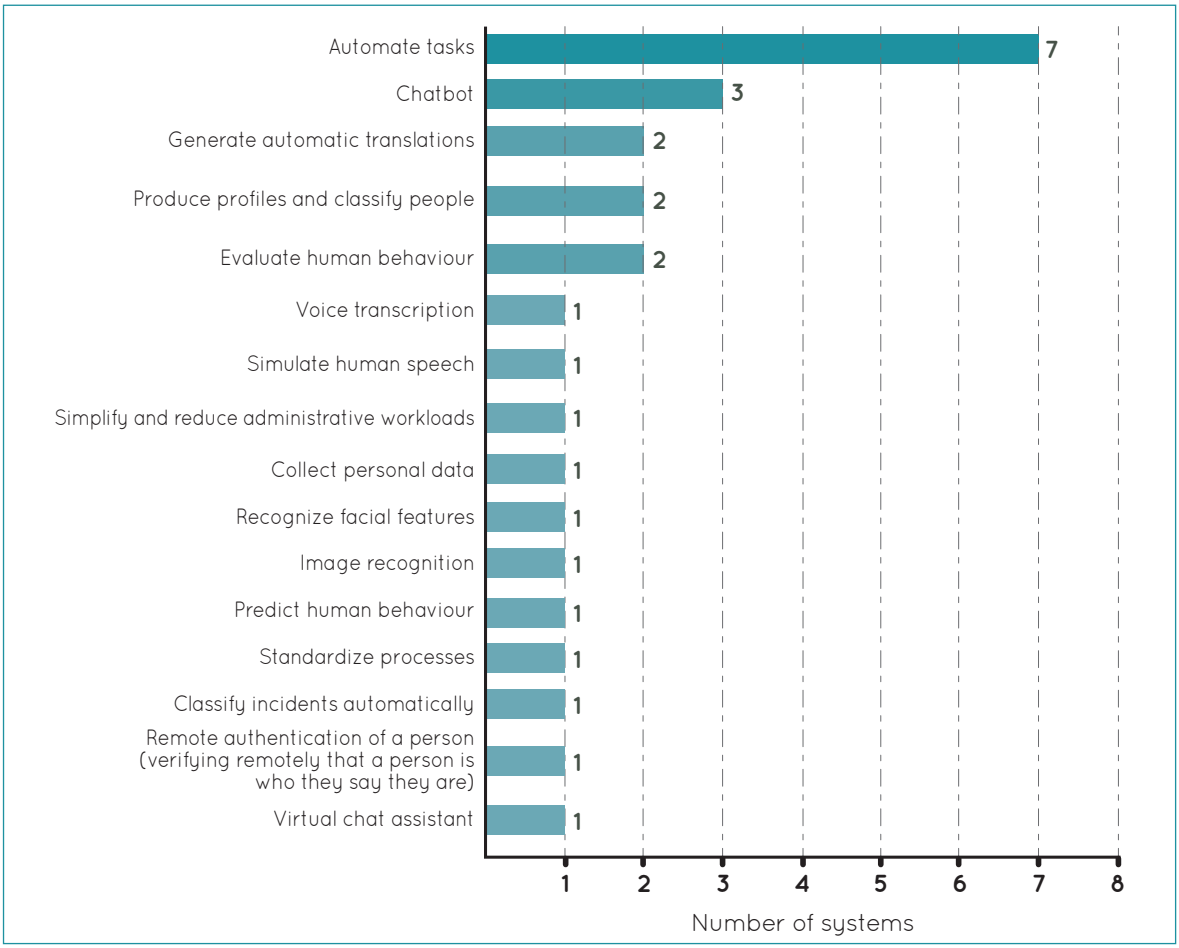
As regards the degree of human intervention, only one system is not subject to any kind of human supervision (the MARIO system), but considering that it is limited to giving users recommendations about how to classify their incident within a series of topics, it is reasonable that no human supervision is

required. In the other cases, either the results are monitored periodically or a human has the final decision, depending on the use. Furthermore, eight of the systems have already undergone or will soon undergo an audit, and public information is available about the main characteristics of the systems in nine cases.

In general terms, and despite these technologies being in a very recent phase of development, implementation and use by the Catalan Public Administration, **the degree of security (considered to be oversight of the system by means of audits and supervision, as well the dissemination of these tools) is therefore considerable, which is a positive starting point.**

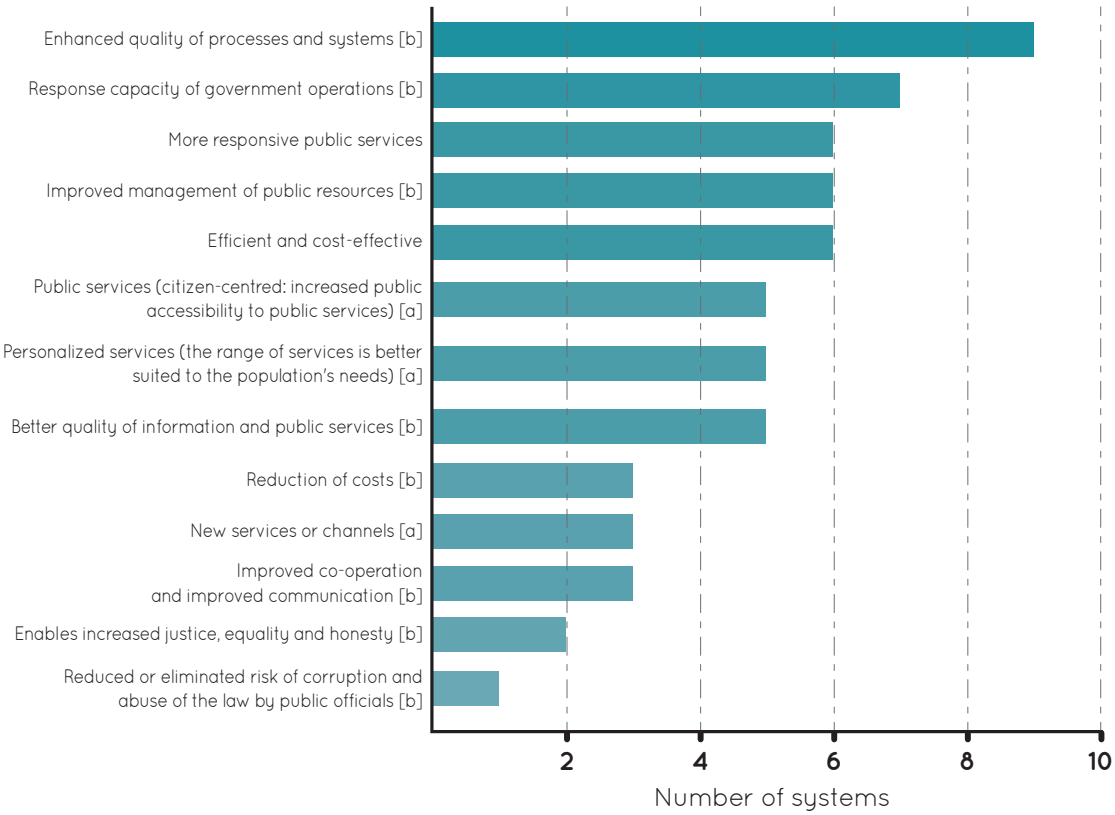
The objectives of these AI systems, classified according to the items used by the Joint Research Centre (JRC), vary. However, seven of the twelve systems included are used for the automation of tasks. Gains in efficiency and time therefore seem to be an important factor in the decision to implement these systems.

Figure 13. Objectives fulfilled by Artificial Intelligence systems



As for the benefits, and also in accordance with the categories taken from the JRC, most of the experts highlighted the increase in the quality of the processes and systems in nine of the twelve algorithms studied, followed by the response capacity of the government's operations

Figure 14. Benefits provided by Artificial Intelligence systems



Finally, the experts identified few risks. It is also important to note that more than one interviewee mentioned that any process implemented by the Public Administrations must have safeguards. This significantly reduces the range of systems that can be used. Only three of the systems considered were deemed to be able to pose a threat to privacy, while two systems could carry a risk of spreading erroneous information. Discrimination on the grounds of race, socio-economic status or other types of discrimination has also been identified in one or two systems.

Risk identified by the expert	Number of AI systems that comply
Threat to privacy	3
Spread wrong information	2
Other types of discrimination	2
Socio-economic discrimination	1
Racial discrimination	1

4.1. Factsheets

Factsheets for the twelve systems identified are presented below.

1. Collective Intelligence - Demand, Problem and Response Identification System (Automatic DPR)

FIELD	RESPONSE
1. Name	Collective Intelligence - Demand, Problem and Response Identification System (Automatic DPR)
2. Description	The automatic DPR project is part of a broader initiative to implement a support system for decisions for professionals based on advanced artificial intelligence tools, which compile past experience (collective intelligence), to make proposals for diagnoses and resources. This project uses natural language processing (unstructured data) to respond to the need to register data more quickly in the information system used by social care professionals. Barcelona City Council's Social Rights Department deals with an average of fifty thousand initial appointments every year. The people who visit the forty social service centres located all over the city have financial problems, dependency issues, mental illness, alcoholism, and may need psychological help, assistance with adaptation, may be experiencing gender violence, etc. These very varied problems are dealt with by a workforce contained more than seven hundred professionals, including social workers, psychologists and social educators. When the person arrives at the centre, they are seen in private booths. The social worker records the conversation, and when it has finished, they transcribe the problem, as well as the help or service which the person concerned has been referred to. In the internal system, information is classified using three letters: demand (D), problem (P), and resource (R). The City Council currently has hundreds of thousands of interviews, many of which end up being repetitive because the problems are quite similar. The automatic DPR provides support for the process for registration of DPR: Demands, Problems and Responses (what the person being dealt is asking for, the problems that the professionals identify and the responses that are proposed to deal with them). According to the professionals, they encode the DPR because it is mandatory, but it provides little added value in terms of service. However, this data is very useful to the organization (for extracting indicators); although they could also be used to provide information for the professionals involved (to see how the case involved, transferring cases, etc.).

FIELD	RESPONSE
2. Description	The project began with a pilot test at three Social Service Centres in 2018. The test was successful, both qualitatively (the suitability of the proposals made by the system) and quantitatively (the use made of them): its effectiveness and its level of consistency was very high (between 75% and 95% depending on the category) and a larger volume of multiple proposals was detected (more than one D, P, R per case) than in manual registration. After the test period ended, the City Council decided to maintain the system in these three centres in the city before rolling out the tool elsewhere. The coding of the dictionaries of problems and resources used by the social service centres changed in October 2020. This meant that the system had to be retrained, as it had to make new associations. With the new dictionaries in use, the centres were required to record demands, problems and responses over a period of time from June 2021 onwards, and the system was retrained. As a result of the positive reassessment of the success rate of the automatic proposal, this new functionality was deployed in all the centres in the city of Barcelona in October 2021, as well as in the dependency care service and the Vulnerable Childhood and Adolescence Study Demands Team. A process to internalize the Collective Intelligence System within the City Council's corporate infrastructure was carried out in 2022. Quality measurement and automatic and manual retraining processes were defined during this process, which also enabled the algorithm's degree of success to be monitored. Use of the automatic DPR functionality is now part of the daily routine of the teams of professionals, and considered a consolidated tool within the system and part of the career journey of care professionals.
3. Objective I	The automatic DPR is a tool integrated into the information system of social service centres that automates the coding process for the demands received, the problems detected and the proposals for the prescription of resources made by professionals in the system. The primary objective is therefore to support the social worker's decision. The system has been trained with machine learning techniques using a repository of three hundred thousand interviews, which are primarily the records of the conclusions of the appointments made by the professionals. Based on the conclusions of the interview entered by a professional, the system, which has been previously trained, analyses all the content and looks for matches. Based on these matches, it then looks for what the professionals have reported in similar situations, and proposes Demands, Problems and Responses. The professional can then validate these proposals, and after validating them and reporting the D again, can then ask the system to refine the P and R again.

FIELD	RESPONSE
	As a result, the professionals do not have to search through dictionaries of defined demands, problems and responses, as the system offers a proposal based on its interpretation of the conclusions of the appointments and the records of information entered into the system.
3. Objective I	This project makes an action that previously had to be performed manually into an automatic task, which only requires the professionals' validation. This tool reduces the hours allocated to administration by the teams of professionals, in order to be able to increase the time spent on social care, while increasing the registration and systematization of information.
4. Objective II	◗ Automate tasks ◗ Evaluate human behaviour ◗ Preparation of profiles and classification of people
5. Body responsible	Municipal Social Services Institute - Barcelona City Council Social Rights Department - Barcelona City Council
6. Developing institution	In its pilot phase, the project was led by the Innovation Directorate of the Social Rights Department, in collaboration with the Municipal Institute of Social Services and implemented by the company INNOVA. The algorithm's internalization phase was administered by the Municipal Computer Institute in partnership with the company Accenture.
7. Status	Implemented: the system has been implemented and is in use
8. Start date	2018 in pilot format, 2019 in the social rights department and finally for all centres in Barcelona in October 2021,
9. End date	
10. Group	Social workers
11. Benefits I	**1)** Help for social workers in categorizing interviews with citizens **2)** Systematization of the prescribed services/benefits based on a neutral and objective view of the demands. The system provides assistance by guiding the professional with proposals based on previous experience (collective intelligence), so that the DPR brings added value to the care process. **3)** Efficiency of resources allocated: it facilitates the DPR registration process, reduces the time spent on this task and facilitates multi-registration.

FIELD	RESPONSE
12. Benefits II	◗ Increased quality of information and public services [a] ◗ Personalized services (the range of services is better suited to the population's needs) [a] ◗ Public services (citizen-centred: increased public accessibility to public services) [a] ◗ More responsive, efficient and cost-effective public services [a] ◗ Improved management of public resources [b] ◗ Enhanced quality of processes and systems [b] ◗ Response capacity of government operations [b]
13. Risks I	There may be some bias, as social workers are humans who have their own bias, but this is minimized by the large volume of interviews and their limited impact on people. The interviews are part of the municipal corporate system/cybersecurity standards. The model, but not the data, is hosted on an external server. Since the system is used in the second phase - after a response has already been given - the risk may be associated with the poor management of the service/process, rather than with the algorithmic system. The process itself, and the way the system is used, mitigates the risks.
14. Risks II	
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	The system offers a proposal to the Demands, Problems and Responses professional. The professional can then validate this proposal, and after validating it and reporting the D again, can then ask the system to refine the P and R again. The professional therefore always takes the final decision.
17. Auditing	An external audit has been or is currently being conducted
18. Transparencia	◗ Problem to which a solution is sought ◗ Objectives and population interacting with the algorithm ◗ How it is being implemented ◗ Results obtained to date
19. Visibilidad	The information is public but not easily accessible (it depends on the citizen's proactivity to find it)
20. Links	https://ajuntament.barcelona.cat/mesames/noticia/posem-en-marxa-una-prova-pilot-dintel%C2%B7ligencia-col%C2%B7lectiva-a-tres-centres-de-serveis-socials/ https://ajuntament.barcelona.cat/dretssocials/es/innovacion-social/inteligencia-colectiva https://ajuntament.barcelona.cat/premsa/wp-content/uploads/2021/02/Mesura-de-Govern-dInnovacio-Social.pdf

2. IRIS Case Processor

FIELD	RESPONSE
1. Name	IRIS Case Processor
2. Description	An internal management tool to speed up the process of categorizing complaints, incidents and suggestions that reach the Social Rights Department through the IRIS platform. The system also provides information trends of the most recurring topics based on various criteria.
3. Objective I	The aim is to improve the extraction of relevant data from queries/complaints that arrive through the IRIS system, either by using automated classification systems or by creating natural language queries.
4. Objective II	Automate tasks
5. Body responsible	Barcelona City Council - Social Rights Department
6. Developing institution	Social Innovation Technical Secretariat - Social Rights Area
7. Status	Pilot: the system is being tested
8. Start date	2022
9. End date	In progress
10. Group	Managers of the Social Rights Department
11. Benefits I	Improvedefficiency, more specific answers and more in-depth knowledge of the organization of queries/complaints that arrive
12. Benefits II	◗ More responsive, efficient and cost-effective public services [a] ◗ Response capacity of government operations [b] ◗ Enhanced quality of processes and systems [b]
13. Risks I	There are no associated risks since the classification system is for internal use and for facilitating the exploitation of data.
14. Risks II	
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	The application makes a suggestion of the topics covered in the complaint, incident or suggestion sent by the citizen. They may be edited or rejected.
17. Auditing	An internal audit has been or is currently being conducted
18. Transparency	
19. Visibility	The citizen is not informed
20. Links	There is currently no public information available, as it is in the pilot phase.

3. Incident classification support system - MARIO

FIELD	RESPONSE
1. Name	Incident classification support system – MARIO
2. Description	IRIS is the service that enables citizens to report incidents or send complaints to Barcelona City Council to be resolved. Citizens can use IRIS to submit information and queries, as well as complaints and suggestions for improvement. In this process, the citizen who reports the incident must classify it using a thematic tree provided by the software application. This classification is important, because it is used to send the incident straight to the appropriate department, thereby speeding up the response process. Errors in thematic classification lead to inappropriate responses and delays in resolving incidents, which affects the quality of the service provided. A module called MARIO has been developed within the IRIS service upgrade project, which is based on machine learning algorithms and natural language processing (one of the technologies within AI) to simplify the process of classifying incidents for the public. Based on an analysis of the free text that describes the incident, MARIO suggests the most likely categories for the incident so that the citizen can choose the most appropriate category.
3. Objective I	Minimize the error rate in the initial classification of incidents and thereby reduce manual classification processes
4. Objective II	Automatic classification of incidents
5. Body responsible	Information services and public attention office - Barcelona City Council
6. Developing institution	
7. Status	Implemented: the system has been implemented and is in use
8. Start date	April 2021
9. End date	
10. Group	Citizenship with access to digital media
11. Benefits I	Reduces the manual work of public officials by reclassifying incidents and speeding up the administration's response.
12. Benefits II	Better quality of information and public services [a], Enhanced quality of processes and systems [b]
13. Risks I	
14. Risks II	
15. Human intervention	The result is not monitored/queried

FIELD	RESPONSE
16. Human intervention (details)	Not supervised because in principle, the algorithm guides the citizen towards an appropriate classification for the incident that is being processed.
17. Auditing	
18. Transparency	Problem to which a solution is sought
19. Visibility	The information is public and accessible (passive communication)
20. Links	https://media-edg.barcelona.cat/wp-content/uploads/2023/08/03111619/drets-digitals.pdf

4. Chat systems, infoCanal Alzheimer and the 0-16 Childhood Fund

FIELD	RESPONSE
1. Name	Chat systems, infoCanal Alzheimer and the 0-16 Childhood Fund
2. Description	This is a new information channel on the municipal websites used to provide information for citizens about Alzheimer's disease and about the 0-16 Childhood Fund. The answers in the chat systems are built based on validated in-house sources in both cases.
3. Objective I	Offer people curated information through natural language conversations.
4. Objective II	*Chatbot*
5. Body responsible	Barcelona City Council - Social Rights Department
6. Developing institution	Social Innovation Technical Secretariat - Social Rights Area
7. Status	Implemented: the system has been implemented and is in use
8. Start date	2022, the InfoCanal and the 0-16 Childhood Fund in 2023
9. End date	In progress
10. Group	General public
11. Benefits I	Expand information channels and be able to provide answers more easily
12. Benefits II	Increased quality of information and public services [a]
13. Risks I	The main risk is that the chat system does not give adequate answers.

FIELD	RESPONSE
14. Risks II	Spread wrong information
15. Human intervention	All the interactions between people and the chat system are recorded and can be evaluated.
16. Human intervention (details)	
17. Auditing	An internal audit has been or is currently being conducted
18. Transparency	Information is provided in the environments where they are implemented.
19. Visibility	The information is public and accessible (passive communication)
20. Links	InfoCanal Alzheimer: https://ajuntament.barcelona.cat/sanitatisalut/ca/canal/barcelona-lalzheimer Fons 0-16: https://ajuntament.barcelona.cat/serveissocials/ca/canal/ajuts-durgencia-social-families-amb-infants-de-0-16-anys

5. Transcription of social emergency reports (Barcelona Social Emergency Centre - CUESB)

FIELD	RESPONSE
1. Name	Transcription of social emergency reports (Barcelona Social Emergency Centre - CUESB)
2. Description	Automated transcription processes using AI are identified as a significant part of the search for solutions to speed up tasks of low added value for the professionals in the Social Rights Department. The emergency services at the Barcelona Social Emergency Centre (CUESB) write reports after each intervention, which currently means that the staff have to return to the office and fill out unsystematized forms. The aim of this pilot test is for the professionals to be able to dictate reports 'on site' or at any time during the intervention and for the report to be transcribed.
3. Objective I	Streamline the process involved in creating reports through transcription systems.
4. Objective II	Voice transcription
5. Body responsible	Barcelona City Council - Social Rights Department
6. Developing institution	Social Innovation Technical Secretariat - Social Rights Area
7. Status	In development: the system is being developed
8. Start date	2024
9. End date	The pilot test will run for six months and will be evaluated
10. Group	Social emergency service workers, Barcelona City Council
11. Benefits I	Reduced report preparation times, streamlining of tasks with low added value
12. Benefits II	Enhanced quality of processes and systems [b]
13. Risks I	The system has no associated risks, as the user is offered the transcript as a supplement and it can be reviewed and edited.
14. Risks II	
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	
17. Auditing	An external audit will take place soon (planned)

FIELD	RESPONSE
18. Transparency	An external service is used for transcription. The information is public on its website.
19. Visibility	The citizen is not informed
20. Links	External service: https://www.speechmatics.com/

6. RisCanvi

FIELD	RESPONSE
1. Name	RisCanvi
2. Description	The RisCanvi risk assessment and assessment project dates back to late 2007, and the recommendations of the Committee to study measures to prevent recidivism in serious crimes (Resolution JUS/2363/200, Official Journal of the Government of Catalonia no. 4937, of 24 July 2007). The first point in the recommendations mentioned '(...) the establishment of a specific technical procedure for assessing the degree of risk that is shared by all competent institutions - administration of justice, prison services and police forces' The RisCanvi assessment is universal and continuous: it assesses the risk of all inmates of all prisons and parolees, and these assessments must be periodically updated. As a general rule, the assessments are valid for a maximum of six months. The RisCanvi scales (RisCanvi-Screening and RisCanvi-Completa) have been designed to predict the risk of five types of behaviour: self-directed violence, intra-institutional violence, general recidivism, violent recidivism and failure to comply with sentence conditions. A total of 30,832 prisoners have been assessed with a total of 116,680 scales since its implementation in 2008. The third version of the algorithm, RiscCanvi v3, is currently in use

FIELD	RESPONSE
3. Objective I	The objectives of the RisCanvi assessment protocol can be divided into general and specific objectives **General:** 1. Improve individual predictions of risk of future violence 2. Roll out the application of risk assessment tools as a work procedure for professionals working in prison facilities 3. Introduce risk management as an intervention tool 4. Enhance institutional coordination in identifying and monitoring cases **Specific:** 1. Make dynamic predictions adapted to internal and external changes. 2. Identify inmates at risk, regardless of the crime committed. 3. Introduce techniques for screening and detecting risks of violence quickly and easily. 4. Roll out the risk assessment to the entire prison population. 5. Assess the effect of the intervention with specific programmes, in relation to risk management. 6. Standardize criteria among professionals and rate their technical training. 7. Organize systematic procedures for the assessment of specific risks. 8. Perform high quality retrospective analyses of decisions taken in order to improve the assessment process.
4. Objective II	◗ Evaluation of human behaviour ◗ Prediction of human behaviour ◗ Preparation of profiles and classification of people
5. Body responsible	Secretariat for Penal Measures, Rehabilitation and Assistance for Victims. Ministry of Justice of the Government of Catalonia
6. Developing institution	
7. Status	Implemented: the system has been implemented and is in use
8. Start date	2008
9. End date	
10. Group	The prison population
11. Benefits I	

FIELD	RESPONSE
12. Benefits II	◗ More responsive, efficient and cost-effective public services [a] ◗ Improved management of public resources [b]
13. Risks I	
14. Risks II	◗ Socio-economic discrimination ◗ Racial discrimination ◗ Other types of discrimination
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	The algorithm assigns a level of risk that can be corrected by the person.
17. Auditing	An internal audit has been or is currently being conducted
18. Transparency	◗ Problem to which a solution is sought ◗ How it is being implemented ◗ Objectives and population interacting with the algorithm ◗ Results obtained to date
19. Visibility	The information is public but not easily accessible (it depends on the citizen's proac-tivity to find it)
20. Links	https://justicia.gencat.cat/web/.content/home/ambits/reinsercio_i_serveis_peni/manual-aplicacio-protocol-avaluacio-riscanvi.pdf https://digitalfuturesociety.com/es/podcasts/capitulo-1-riscanvi-i-el-algoritmo-de-la-carcel/ https://eapc.gencat.cat/web/.content/home/biblioteca/fons-coleccions/docs_eapc/tic/2018/03_bones_practiques_adm_electronica/10179_eRis-Canvi-EAP-marc-2018-PRE-1.pdf

7. Social Benefits Regulation Engine

FIELD	RESPONSE
1. Name	Social Benefits Regulation Engine
2. Description	The Regulation Engine is a part of the new eSocial platform (a new technological platform to meet the management needs of the Ministry of Social Rights) and is de-signed to apply the criteria used to grant a social welfare while it is being processed.
3. Objective I	Confirm whether a citizen or institution meets the criteria to receive a social welfare benefit. Five algorithms have currently been developed: Basic Needs, Household Ex-penses, Non-Contributory Retirement Pension, Supplement to the Non-Contributory Retirement Pension and Benefit for birth, foster care and adoption (families)
4. Objective II	Automate tasks
5. Body responsible	Generalitat de Catalunya - Ministry of Social Rights
6. Developing institution	Telecommunications and Information Technologies Centre (CTTI)
7. Status	Implemented: the system has been implemented and is in use
8. Start date	December 2018
9. End date	
10. Group	5 algorithms have been developed: Basic Needs, Household Expenses, Non-Contributory Retirement Pension, Supplement to the Non-Contributory Retirement Pension and Benefit for birth, foster care and adoption (families). It benefits all citi-zens who need this benefit, i.e. senior citizens, families and citizens in a situation of social exclusion.
11. Benefits I	Efficiency in processing. Reduced processing times.
12. Benefits II	◗ Public services (citizen-centred: increased public accessibility to public services) [a] ◗ Response capacity of government operations [b] ◗ Improved management of public resources [b] ◗ Enhanced quality of processes and systems [b]
13. Risks I	None detected
14. Risks II	
15. Human intervention	The algorithm is monitored periodically to ensure that it is functioning correctly
16. Human intervention (details)	
17. Auditing	An external audit will take place soon (planned)

FIELD	RESPONSE
18. Transparency	◗ Type of algorithm used ◗ Problem to which a solution is sought ◗ How it is being implemented, goals and population interacting with the algorithm ◗ Results obtained to date
19. Visibility	
20. Links	https://ctti.gencat.cat/ca/detalls/detallNoticia/Pla_Transformacio_Digital_Drets_Socials_noticia_2023

8. Benefits Advisor

FIELD	RESPONSE
1. Name	Benefits Advisor
2. Description	Facilitate access to social benefits information
3. Objective I	Implement a chatbot that helps citizens find out what kinds of assistance they can receive from the administration.
4. Objective II	◗ Simulate human speech ◗ Generate automatic translations ◗ Chatbot
5. Body responsible	Generalitat de Catalunya - Ministry of Social Rights
6. Developing institution	Telecommunications and Information Technologies Centre (CTTI)
7. Status	Pilot: the system is being tested
8. Start date	July 2023
9. End date	
10. Group	All citizens
11. Benefits I	It creates another channel for the administration to communicate with the public.
12. Benefits II	◗ Personalized services (the range of services is better suited to the population's needs) [a] ◗ Increased quality of information and public services [a] ◗ New services or channels [a] ◗ Response capacity of government operations [b]

FIELD	RESPONSE
13. Risks I	Sharing of personal data
14. Risks II	
15. Human intervention	The result of the algorithm is monitored periodically
16. Human intervention (details)	
17. Auditing	An external audit will take place soon (planned)
18. Transparency	
19. Visibility	
20. Links	

9. Identification of Social Intervention needs

FIELD	RESPONSE
1. Name	Identification of Social Intervention needs
2. Description	Subject to the citizen's consent, the public official can enable a function of the eSo-cial social intervention tool during the social intervention. This function uses the tran-scription of the essential part of the conversation with the citizen to automatically identify situations of need.
3. Objective I	Facilitate the work of the social worker.
4. Objective II	◗ Automate tasks ◗ Collect personal data ◗ Generate automatic translations
5. Body responsible	Generalitat de Catalunya - Ministry of Social Rights
6. Developing institution	Telecommunications and Information Technologies Centre (CTTI)
7. Status	Pilot: the system is being tested
8. Start date	December 2023
9. End date	
10. Group	All citizens

FIELD	RESPONSE
11. Benefits I	Administrative efficiency, Standardization of the administration's criteria.
12. Benefits II	◗ Personalized services (the range of services is better suited to the population's needs) [a] ◗ Public services (citizen-centred: increased public accessibility to public services) [a] ◗ Response capacity of government operations [b] ◗ Enhanced quality of processes and systems [b] ◗ Improved co-operation and improved communication [b]
13. Risks I	
14. Risks II	
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	
17. Auditing	
18. Transparency	
19. Visibility	The citizen is informed directly (active communication)
20. Links	

10. Open Administration Consortium of Catalonia (AOC) chatbots

FIELD	RESPONSE
1. Name	Open Administration Consortium of Catalonia (AOC) chatbots
2. Description	**Problem detected: :** Need for constant assistance: The public employees and citizens required continuous support and information 24 hours a day, 7 days a week, but face-to-face or telephone support was unable to provide this constant availability. Increase in queries and incidents: The increase in digital procedures and the use of the transparency portal led to an increase in the demand for information and assistance. **Solution adopted and implementation:** Development of chatbots: The decision was taken to create chatbots with the ability to interpret and answer questions by using AI algorithms in order to address this issue. Training with existing data: The algorithms were trained with already existing information, such as FAQs from the AOC support portal, structured into questions/answers and guided chat flows. Implementation in several specific areas: the chatbots were integrated in different areas related to the digital identification process, for obtaining, using and renewing the idCat Certificate and browsing the Transparency and e-TRAM 2.0 portal. **Results to date** Improved user experience: Users can access information and resolve their doubts more quickly and immediately, reducing the dependence on in-person or telephone support. Increased self-resolution: Chatbots have enabled users to resolve inquiries and incidents without the need for human interactions, increasing efficiency and freeing up resources. Supervision and continuous improvement: Analysis of the chat logs and user feedback have helped improve the algorithms, add new conversation flows, and expanded content to better match user needs.

FIELD	RESPONSE
3. Objective I	The main objective of the system is to facilitate digital identification processes and access to information within the transparency portal, to improve the user experience and increase self-resolution of incidents and queries by citizens and public officials. The system is developed for three specific tasks: 1. 1. Assistance with identification: users are provided with support in the identification and identity validation processes required to carry out a digital procedure. 2. Support in applying for and administering certain documents: help with resolving the most common doubts and queries related to obtaining, using and renewing the idCAT Certificate. 3. Application of chatbots in specific areas, such as the Transparency portal and e-TRAM 2.0, providing specific information and facilitating browsing in these areas.
4. Objective II	◗ Automate tasks ◗ Chatbot ◗ Conversational virtual assistant
5. Body responsible	Open Administration Consortium of Catalonia (AOC)
6. Developing institution	ONE MILLION BOT, S.L.
7. Status	Implemented: the system has been implemented and is in use
8. Start date	21 January 2021
9. End date	
10. Group	A support service for public employees and citizens
11. Benefits I	**Benefits for the Administration:** Optimized resources: Reduced workload for public employees as the chatbots handle frequent and repetitive queries, freeing up the employees to carry out more complex and specialized tasks. Improved efficiency: Increased efficiency in information management and resolution of queries, enabling an immediate and accurate response to a large volume of simultaneous queries. Data analysis: Analysis of the chat logs provides a valuable source of data for the Administration, allowing it to identify patterns, areas for improvement and users' needs to adapt future services. Reduced costs: Reduced costs associated with face-to-face and telephone customer service, as chatbots provide digital assistance without the need for human resources.

FIELD	RESPONSE
11. Benefits I	**Benefits for the citizen:** 24x7 access to information: Constant access to information and support without any restrictions of hours or days, improving convenience and the user experience. Immediate response: Quick and immediate answers to queries and doubts obtained, avoiding waiting time and speeding up procedures and processes requiring specific information. Self-resolution of queries: Ability to self-resolve queries and incidents with no need to resort to personnel at the citizens' service centres, giving citizens more autonomy in their interaction with the Administration. Improved user experience: The users' experience in interacting with public services has been improved by guided browsing and precise answers to specific questions, facilitating understanding and helping to clear up doubts.
12. Benefits II	◗ Personalized services (the range of services is better suited to the population's needs) [a] ◗ Public services (citizen-centred: increased public accessibility to public services) [a] ◗ Increased quality of information and public services [a] ◗ More responsive, efficient and cost-effective public services [a] ◗ New services or channels [a] ◗ Reduction of costs [b] ◗ Improved management of public resources [b] ◗ Enhanced quality of processes and systems [b] ◗ Improved co-operation and improved communication [b]
13. Risks I	**Risks for the Administration:** Dependence on technology and service outages: Any mistakes by the chatbots or any outages interruptions in the operation of the system can interrupt the service to the users and create mistrust in the reliability of the service. Data security and privacy: Despite the security measures taken, there is a risk of data protection vulnerabilities, especially due to the conversations being recorded, which could endanger the users' privacy (this may happen if the user inputs unnecessary personal data). Quality of answers: If the training material and continuous improvement process is not administered properly, the chatbots may provide inappropriate responses, undermining the trust in the information provided and perceptions of the system's effectiveness.

FIELD	RESPONSE
13. Risks I	**Risks for citizens:** Trust and understanding of the support: citizens' trust in the system may decline if they believe that the answers are not satisfactory or are not suitable for helping them, especially in complex cases or specific situations. Data protection and management of personal information: There is a concern that users' personal data may be vulnerable if the system does not manage the collected information correctly, or if errors in security protocols occur. Communication and limitations on the assistance: Chatbots may have limitations in terms of understanding of natural language or in their ability to provide appropriate responses for all situations, which may affect the quality of the assistance and users' satisfaction.
14. Risks II	◗ Threat to privacy ◗ Spread wrong information
15. Human intervention	The result of the algorithm is consulted by specialists at the citizens' service centres (first or second level) and by experts in monitoring the algorithm, but they cannot modify the answer given.
16. Human intervention (details)	The specialists at the citizens' service centres (first or second level) deal with the questions that could not be answered, and produce the conversation flows that will permit the right answer to be given the next time. Monitoring experts regularly su-pervise the comment and chat logs, and analyse the service's usage statistics. Based on this data, the experts identify the issues that interest users, and develop new chat flows and expand the conversational corpus with new keywords.
17. Auditing	
18. Transparency	◗ Data used to train the algorithm ◗ Type of algorithm used ◗ Problem to which a solution is sought ◗ How it is being implemented ◗ Objectives and population interacting with the algorithm ◗ Results obtained to date ◗ Risk management, contact information for inquiries and the developer of the algorithm
19. Visibility	The citizen is informed directly (active communication)

FIELD	RESPONSE
	The initiative of publishing the Open Administration Consortium of Catalonia's algorithmic transparency factsheet for chatbots in the Transparency Portal is one of the recommended actions referred to in the proposed AI Regulation and the EU Guidelines for Trustworthy AI.
20. Links	The primary motivation that has led us to make this proposal is to ensure the transparency of the AI algorithms that we apply to our digital administration services, and so that people are informed and empowered to use them when they interact with them. Algorithmic Transparency Factsheet: https://www.aoc.cat/ia-transparencia-xatbot/

11. idCAT Mòbil identification document video identification service

FIELD	RESPONSE
1. Name	idCAT Mòbil identification document video identification service
2. Description	The initial problem lies in people's difficulty in obtaining a digital identity document (eID) in Catalonia, especially during the pandemic. With restrictions on mobility and the growing need to interact with administrations from home, the absence of a means of digital identification created obstacles and limitations in transactions and communications with public institutions. The solution was the development of an AI video identification system for the idCAT Mobile. This solution allows people to obtain an officially valid digital identity document to interact with the Catalan administrations from any location which has Internet access. This system uses facial recognition algorithms and the validation of official documents to ensure reliable remote authentication comparable to a physical presence. The implementation involves a simple process for the user, who can perform a video identification using their mobile phone. The system verifies the match between the official document presented and the person's 'selfie' video, ensuring the integrity of the process. This verification is subsequently validated by specialized human operators to ensure it is correct, and to eliminate potential risks or problems. This system has successfully provided more than 110,000 digital identities to date, solving the initial problem of accessing a digital identity during restrictions such as those in place during the pandemic. The results show high levels of efficiency and acceptance by users, providing a viable and secure solution for their need to interact with administrations without having to make a physical journey.

FIELD	RESPONSE
3. Objective I	**Objective of the AI System:** The main objective is to allow anyone over the age of 16 to obtain a digital identity resource in order to interact with the Catalan administrations from their home, using only a mobile phone and an official document containing a photograph, such as a passport or national identity document identity card. **Function of the AI System:** The main function of the system is to provide a video identification service that uses AI algorithms to remotely verify that a person is who they say they are, with high levels of accuracy and security. This process leads to an individual being registered in the IDCAT Mobile electronic signature and identification system, thereby obtaining a digital identity to be able to deal with any Catalan administration with security and privacy safeguards. The system is designed to be accessible 24x7 with a high level of availability, and provides benefits in terms of convenience, inclusion and enhanced security, while maintaining human supervision to ensure the process is carried out correctly and in compliance with data protection and security legislation.
4. Objective II	◗ Automate tasks ◗ Recognition of facial features ◗ Image recognition ◗ Remote authentication of a person (verifying remotely that a person is who they say they are)
5. Body responsible	Open Administration Consortium of Catalonia (AOC)
6. Developing institution	DELOITTE ADVISORY, S.L. which provides Deloitte OBA technology integrated with the VERIDAS solution
7. Status	Implemented: the system has been implemented and is in use
8. Start date	May 2020
9. End date	
10. Group	Citizens over 16 years of age

FIELD	RESPONSE
11. Benefits I	**Benefits para la Administración:** ◗ **Eficiencia:** reducción significativa del tiempo y de los costes administrativos asociados a la verificación de identidad y la gestión presencial. **Benefits for the Administration:** ◗ **Efficiency:** significant reductions in the time and administrative costs associated with identity verification and face-to-face management. ◗ **Security:** substantially increased security in the authentication process, with human oversight that reduces the risks of fraud or errors. ◗ **Regulatory compliance:** alignment with data protection legislation and legal certifications for issuing qualified certificates, ensuring the reliability and legality of the documents issued. However, there are no specific regulations applicable to idCat Mòbil and the Open Administration Consortium of Catalonia has decided to voluntarily follow the certificate model. **Benefits for the citizen:** ◗ **Convenience:** an easy and convenient means of access to digital proof of identity without the need to travel, and the ability to complete the process from home using a mobile phone. ◗ **Inclusion:** improved accessibility, with a simple process aimed at all types of people regardless of their digital skills. ◗ **Security and trustworthiness:** guarantee of an accurate and reliable authentication process, with a high level of security in identity verification.
12. Benefits II	◗ Personalized services (the range of services is better suited to the population's needs) [a] ◗ Public services (citizen-centred: increased public accessibility to public services) [a] ◗ More responsive, efficient and cost-effective public services [a] ◗ New services or channels [a] ◗ Reduction of costs [b] ◗ Response capacity of government operations [b] ◗ Improved management of public resources [b] ◗ Enhanced quality of processes and systems [b] ◗ Enables increased justice, equality and honesty [b]
13. Risks I	**Risks for the Administration:** ◗ **Protection of sensitive data:** dealing with biometric and personal data during video identification can expose the user to security risks, especially if there are outages in the encryption process or if the data is susceptible to being compromised during the procedure. ◗ **Human supervision and control:** despite supervision by the operators, there is the potential for human error in validating and verifying processes, which could result in incorrect identifications or failure to detect attempts at fraud. ◗ **Identity theft:** despite the controls, there is a risk that some people may use methods to cheat the system and impersonate others. ◗ **Dependence on technology:** a high level of reliance on this system could create vulnerabilities in the event of system failures or technology outages affecting the identification service.

FIELD	RESPONSE
13. Risks I	**Risks for the citizen:** ◗ **Privacy and data security:** the use of biometric data in the process may raise concerns about user privacy and security, especially if this data is vulnerable to cyberattacks or abuse by third parties. ◗ **Possible biases in facial recognition:** despite the low rejection rate, there is still a possibility of biases in facial recognition that could lead to errors in identification or cases of false positives or negatives, affecting different groups of people to different extents. ◗ **Dependence on connectivity and technology:** the video identification process is dependent on an Internet connection and the device's technology, which could restrict access for people with poor connections or limited technology.
14. Risks II	◗ Threat to privacy ◗ Other types of discrimination
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision.
16. Human intervention (details)	The specific actions carried out by the operators who monitor the transactions are: **a.** Watch the video and check that the person concerned has performed the whole process without any coercion, that they have performed the process according to the requirements stipulated (they are not wearing a head covering or wearing sunglasses, their face is visible, etc.) and check that the watermarks and authenticity holograms on the document displayed are visible. **b.** Review the result of the documentation evaluation parameters that the system has evaluated automatically: the percentage match between the photo on the identity document and the selfie photo, the expiry date of the identity document, the date of birth, the document number and that it is legible and contains watermarks.
17. Auditing	**1)** To identify the risks associated with the availability and security of the system, on 7/7/2020 the Open Administration Consortium of Catalonia carried out a security analysis of the video identification service following the guidelines of the Catalan Cybersecurity Agency and the Spanish National Security Framework (ENS). **2)** To identify the risks of the system from the point of view of fundamental rights, underlying principles and values, on 17/12/2020 the Generalitat de Catalunya car-ried out a data protection impact assessment (DPIA) and an assessment of the im-plementation of the video identification process for obtaining the Mobile idCat. 3) Regular quality tests are carried out to detect inaccuracies and prevent biases to guarantee the impartiality of the applied algorithms. **3)** National Institute of Standards and Technology (NIST) Facial Recognition Vendor Test (PDF latest report, 16/6/2023)

FIELD	RESPONSE
18. Transparency	◗ Data used to train the algorithm ◗ Type of algorithm used ◗ Problem to which a solution is sought ◗ How it is being implemented ◗ Objectives and population interacting with the algorithm ◗ Results obtained to date
19. Visibilidad	The citizen is informed directly (active communication)
20. Links	Innovation factsheet: 'Remote identification of citizens by video identification': https://www.aoc.cat/projecte-innovacio/identificacio-remota-dels-ciutadans-mitjancant-videoidentificacio/ An algorithmic transparency factsheet to be published on the Open Administration Consortium of Catalonia's Transparency Portal in the near future. [https://www.seu-e.cat/ca/web/consorciaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234]

12. Energy poverty reports automation service

FIELD	RESPONSE
1. Name	Energy poverty reports automation service
2. Description	The energy poverty reports automation service is a solution that was developed to automate the production of energy vulnerability social reports. These reports are essential for ensuring people and households at risk of residential exclusion have access to basic supplies, according to Law 24/2015 on urgent measures to address the housing and energy poverty emergency (article 6). For this reason, the various Social Services Basic Areas (SSBA) have to write the aforementioned reports on a monthly basis to certify economic vulnerability in order to prevent electricity, gas and drinking water companies (more than 50 different companies in the country) cutting off supplies due to a lack of payment. The main problem detected was the complexity and dispersed nature of the data required to produce energy poverty reports. The existence of multiple data sources and heterogeneous formats created an error-prone and inefficient manual process for verifying the energy vulnerability of individuals and family units at risk of residential exclusion.

FIELD	RESPONSE
2. Description	The solution adopted was the implementation of an AI system that uses data processing algorithms and technologies such as RPA and BPM. This system has been developed to automate the collection and verification of data, to produce energy poverty reports, and to notify energy supply companies of the result. Its implementation entails integration with multiple systems, such as the Via Oberta interoperability services and the Hèstia social services information system. In specific terms, the implementation consists of creating a cloud platform provided by a private company (ConsultorsBPM), which is hosted on Microsoft Azure in the European Community. This platform enables users to upload the energy suppliers' records, obtain the socio-economic data of the account holders, and take other steps necessary to calculate the coefficients and automatically generate vulnerability reports. The reports are also passed on to the supply companies via e-Notum. The results have a success rate of 90% to date, although with legal and organizational changes this figure could reach almost 100%. The implementation has been tested in six different municipalities: El Prat de Llobregat, Amposta, Granollers, Sant Boi de Llobregat, Palau Solità i Plegamans and Santa Coloma de Gramenet. The main results obtained to date are: 1. Significant savings on time and administrative tasks: speed and efficiency in the generation of energy poverty reports has been improved, allowing a better service to be provided to citizens in vulnerable situations. 2. Greater interoperability and improved data management: the links between companies, municipalities and other administrations have been enhanced, providing a more accurate perspective of the energy vulnerability situation. There has also been a significant reduction in administrative burdens for citizens due to interoperability. 3. Support for municipalities with limited resources: the system provides support for municipalities that do not have the funds needed to undertake large digital transformation projects, giving them access to this service through a cloud-based service with no need for a local installation.

FIELD	RESPONSE
3. Objective I	The main objective of the AI system is to automate and standardize the creation of energy poverty reports to improve management, data verification and decision-making in energy vulnerability situations, while ensuring compliance with data protection legislation regulations and ensuring human oversight to ensure accuracy of the results. As regards its functions, this service: 1. Standardizes and automates the generation of energy poverty reports. 2. Improves data management and interoperability between multiple sources and systems. 3. Optimizes municipalities' resources for confirming situations of energy vulnerability. 4. Ensures compliance with data protection legislation and current regulations. Human supervision is essential for validating the results of the system before they are sent to the supply companies, ensuring accuracy and preventing possible errors in specific cases that may circumvent the logic of the system. Likewise, the system is designed and implemented according to the requirements of data protection legislation, and in particular the General Data Protection Regulation (GDPR) and other local regulations. Compliance with these standards is crucial, as the work involves personal data and transparency, limitation in data the collection and the correct use of the consent of the affected persons must be guaranteed.
4. Objective II	◗ Automate tasks ◗ Standardization of processes, simplification and reduction of administrative workloads
5. Body responsible	Open Administration Consortium of Catalonia (AOC)
6. Developing institution	BPM consultants
7. Status	Pilot: the system is being tested
8. Start date	27 January 2023
9. End date	
10. Group	Staff in the social services basic areas (SSBA)

FIELD	RESPONSE
11. Benefits I	**Benefits for the Administration:** Savings on time and resources: significant reduction in the time and administrative tasks required to generate energy poverty reports. This improves the management of the Administration's human and economic resources, optimizing its performance. Standardization and improvement of data management: a more precise and standardized view of the energy vulnerability is achieved with the integration and automation of multiple data sources. This improves decision-making and facilitates communication between companies, municipalities and other administrations. Support for municipalities with limited resources: the cloud-based service provides an opportunity to access advanced technology with no need for large investments in infrastructure or internal resources. **Benefits for Citizens:** Improved and more flexible service: the automation and standardization of the process enables a faster and more accurate response to cases of energy vulnerability, improving the quality of social services and attention to people at risk. Ensure access to basic supplies: this system expedites decisions on vulnerability, ensuring that people and households at risk receive the services and basic supplies they need. Reduction of administrative workload and improvement of the user's experience: the automated process avoids unnecessary procedures and simplifies interaction with the Administration, reducing bureaucracy and improving the experience of vulnerable citizens.
12. Benefits II	◗ More responsive, efficient and cost-effective public services [a] ◗ Reduction of costs [b] ◗ Response capacity of government operations [b] ◗ Improved management of public resources [b] ◗ Enhanced quality of processes and systems [b] ◗ Improved co-operation and improved communication [b] ◗ Reduction or elimination of risk of corruption and abuse of the law by public officials [b] ◗ Enables increased justice, equality and honesty [b]

FIELD	RESPONSE
13. Risks I	Despite its advantages, the implementation of this AI system can also present some risks, for both the Administration and for citizens. **Risks for the Administration:** ◗ Dependence on technology: If something goes wrong in the platform or in the automated process, this could lead to delays or errors in the generation of the reports, placing additional pressure on human resources to solve them. ◗ Data vulnerability and security: integrating multiple data sources and using cloud services can lead to security vulnerabilities. If the system is not sufficiently protected, it could be vulnerable to cyberattacks or data losses, compromising the privacy and security of information about citizens. **Risks for citizens:** ◗ Possible errors in the assessment of vulnerability and loss of confidence: despite human supervision, there is a possibility that the system will make errors in certain cases. This could lead to inaccurate reports on the energy vulnerability of individuals or households, affecting their ability to access basic supplies and leading to a loss of confidence in the Administration. The loss of citizens' confidence in the Administration is a critical factor that may arise if there are frequent errors in vulnerability assessments. These errors can lead to distrust in the reliability of the reports produced, affect the perception of transparency and fairness in decision-making, and have a negative emotional impact on the people affected. If access to basic supplies is denied due to incorrect reporting, this can lead to distrust and despair among people who are already in vulnerable situations. For this reason, it is crucial to manage this risk by means of careful supervision, transparent processes and effective communication in order to retain citizens' trust in the services provided by the Administration in situations of energy vulnerability. ◗ Privacy and consent issues: Working with personal data involves a risk of privacy violations if they are not managed properly. The consent of the people affected must be obtained and used in a transparent manner and in accordance with data protection legislation.
14. Risks II	Threat to privacy
15. Human intervention	The result of the algorithm is supervised by a person who has the final decision
16. Human intervention (details)	
17. Auditing	

FIELD	RESPONSE
18. Transparency	◗ Problem to which a solution is sought ◗ How it is being implemented ◗ Results obtained to date
19. Visibility	The information is public but not easily accessible (it depends on the proactivity of the citizen to inquire)
20. Links	◗ Post: 'Pilot test for the automation of the energy poverty report to avoid power cuts' (AOC blog 27/1/2023): https://www.aoc.cat/blog/2023/pilot-pobresa-energetica/ ◗ Article 'Automating the energy poverty report in Catalunya' (10/10/2023, Joinup Platform, European Commission): https://joinup.ec.europa.eu/collection/public-sector-tech-watch/automating-energy-poverty-report-catalunya ◗ Presentation: 'Automated assessment of energy poverty' (18/10/2023, Semic 2023: Interoperable AI in the Age of AI): https://www.linkedin.com/posts/miquelestape_20231018-semic-2023-trustworthy-ai-for-activity-7122219700981256194-f9es/?utm_source=share&utm_medium=member_desktop ◗ The algorithmic transparency record will be published on the Open Administration Consortium of Catalonia's Transparency Portal in the near future. [https://www.seu-e.cat/ca/web/consorciaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234]

5. FINAL CONSIDERATIONS

In this second document in the project, **the priority has been to identify artificial intelligence systems and automated processes in the social sphere which are being used by Catalan Public Administrations or are in the pilot phase of implementation.**

In order to achieve this, it was necessary to clearly define the information to be collected. In other words, the radar fields. It was also necessary to carry out thorough fieldwork, identifying both the people and institutions and departments that could be aware of or using these systems. As a result of this effort **a total of 12 algorithms were mapped**, which are probably most (if not all) of the AI systems currently in use or in the pilot phase in Catalonia.

The cooperation of Public Administration professionals in this respect must be emphasised. Experts at the Generalitat de Catalunya, Barcelona City Council, the Barcelona Municipal Institute of Social Services and the Open Administration Consortium of Catalonia answered the questionnaire, and provided a remarkable degree of detail about the AI systems, which has enabled the Radar to be built.

The specific characteristics of the field of social rights also became evident during the conversations with the specialists. The need to design processes with safeguards when providing services to the public considerably limits the range of artificial intelligence techniques that can be applied. For example, if an AI system were to be used to decide whether to grant a benefit to a person, that system must not be a black-box algorithm (such as a neural network or a random forest) as at present, the explainability of these models is low, which means the ability to explain the model's decision is also limited. For this type of application, it would be preferable to use automation (strictly based on regulation) or explainable algorithms (such as regressions, which are in widespread use in economics and political science).

The specialists were at pains to point out that the obligation to give citizens clear answers about the reason for a result was a key factor in the decision as to whether or not this type of system was adopted. This is evident in the Radar, where many of the algorithms are for 'self-consumption', i.e. they are for automating procedures or helping experts in the

administration, but are not used for making decisions that directly affect citizens. It is reasonable to assume that the types of solutions that Public Administrations adopt in the social sphere will in many cases differ from those that may be used in other areas.

A final point to consider is that we are probably witnessing a paradigm shift. And in any process of profound social transformation, it is normal for part of society to be divided between fear of future uncertainty, and excessive and uncritical optimism about future changes. Finding the middle ground that contributes to effective and fair development, or in other words, a perspective that undertakes a thorough case-by-case analysis without making any prior assumptions, is not an easy task. **The key value of this study is that it has provided important knowledge about how Artificial Intelligence systems are currently being applied in the social sphere.**

In particular, it has provided a series of important conclusions. First, **the total number of systems identified is small - a total of twelve**. Various factors may provide some understanding of why fewer systems are in use in this area than in others such as health. The availability of funds is probably one of these factors. However, the relationship between the Public Administration and the citizen in the social sphere is based on a system of guarantees, and as such the range of opportunities within which Artificial Intelligence applications can be applied is limited.

Second, and related to the above, **many of these systems are for *internal consumption***, i.e. the people interacting with them are administration staff and not the public. This means that these systems are **more focused on improving the efficiency of administrative processes**.

Third, the systems identified carry very low or non-existent levels of risk. The algorithms applied are focused on solving problems which inherently involve very limited risks for the citizen.

Fourth, the cooperation from the managers working in the Public Administration was positive. As a result, most of the information collected came directly from people who have been involved in the design, management or application of these systems.

Finally, **it must be remembered that artificial intelligence is merely a tool, and it is only the most optimal, efficient or effective tool for some specific problems**. It is not an end in itself. The Public Administration must decide whether there is an alternative system to AI that is better for each problem. This idea was clearly understood in the conversations with the specialists.

Taula d'entitats
del Tercer Sector Social
de Catalunya

m4Social

with support of:

Generalitat de Catalunya
Departament
de Drets Socials

AMB Àrea Metropolitana de Barcelona