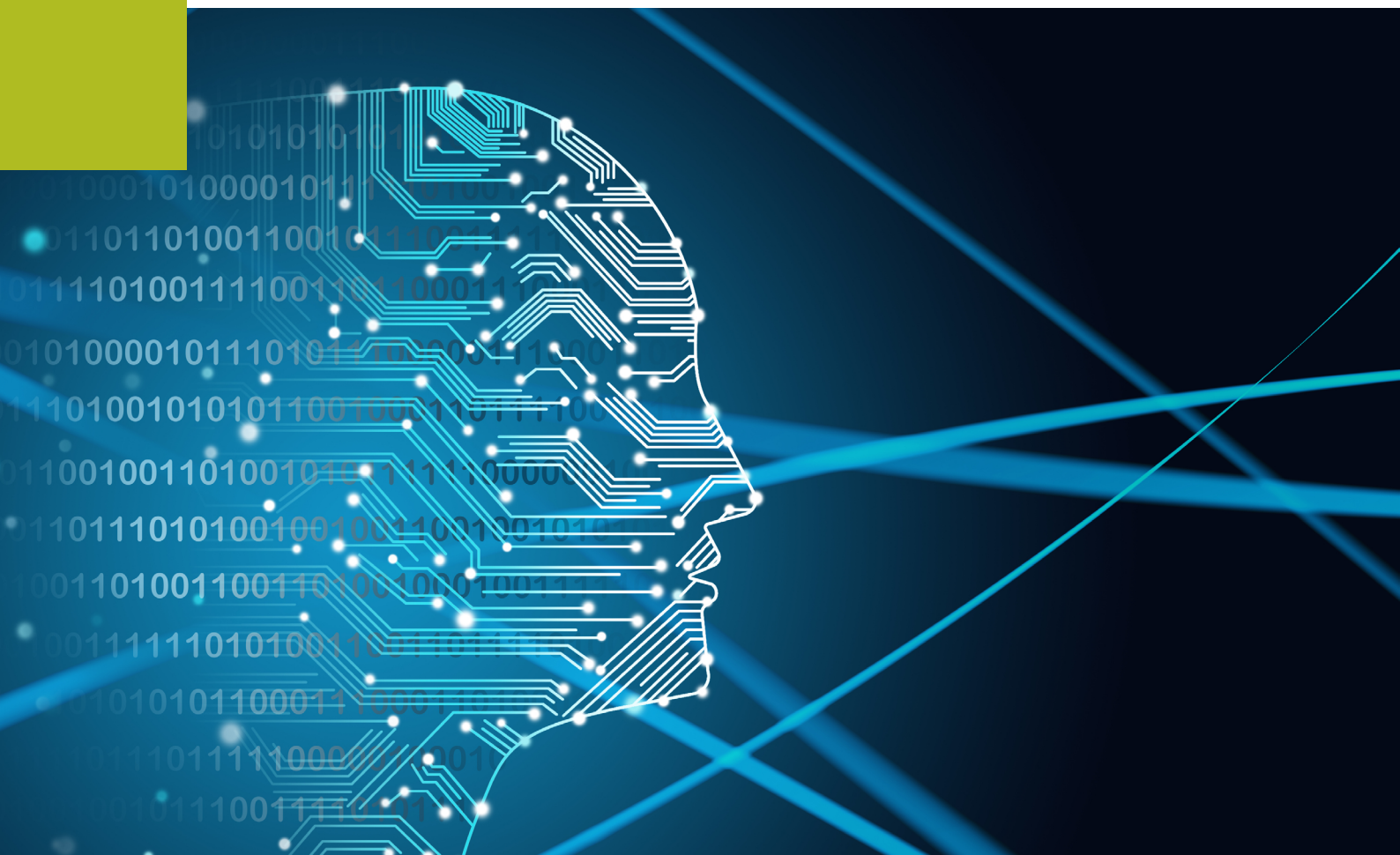
issa | EXCELLENCE IN SOCIAL SECURITY

UNU EGOV

ARTIFICIAL INTELLIGENCE IN SOCIAL SECURITY ORGANIZATIONS

INTERNATIONAL SOCIAL SECURITY ASSOCIATION



This report was jointly prepared by Moinul Zaber and Oxana Casu of the United Nations University Operating Unit on Policy-Driven Electronic Governance (UNU-EGOV) and Ernesto Brodersohn of the International Social Security Association (ISSA).

Available in electronic format at: www.issa.int.

Artificial Intelligence in Social Security Organizations

International Social Security Association
Geneva, 2024

Contents

Abstract

Social security institutions worldwide encounter formidable obstacles in delivering quality services in an increasingly challenging environment. Challenges include limited resources and infrastructure, with escalating demands, which hinder their ability to provide comprehensive support to their members and overall target population. Overcoming these hurdles necessitates innovative strategies and international cooperation to ensure comprehensive service delivery as well as equitable and sustainable social security provision. This is where Artificial Intelligence (AI) becomes a critical and enabling technology in social security. It can help significantly depressurize resources to focus on specific segments of the population, help gain insights into patterns previously undetected, and in general improve service delivery.

The rise of AI capable of leveraging diverse data types to construct efficient tools, or glean insights, has demonstrated the potential to revolutionize service delivery and decision-making processes within institutions, notably at social security organizations. Integrating AI and data facilitates proactive and automated delivery of services. Yet, owing to AI's developmental stage as a science, and deploying diverse AI tools within institutional frameworks proves challenging, the primary hurdle stems from the nature of the data itself – the fundamental ingredient of AI.

At social security institutions, the imperative for high-quality data, contextually relevant models, and stringent AI safety measures becomes paramount. To avert the peril of underutilizing AI, institutions are striving to implement AI tools in different ways. From leveraging intelligent chatbots for improved service delivery, to data-centric decision-making, or by leveraging machine learning, institutions must adapt to the use of AI as part of the new paradigms of digital transformation in order to fulfil their mission objectives. This article outlines the diverse facets of AI-empowered digital transformation, which is essential for social security organizations, emphasizing the importance of robust data quality, context-sensitive models and prioritizing safety standards. By focusing on these elements, it paves the way for AI-focused automation, ensuring that social security institutions harness AI effectively while safeguarding against potential risks.

This report examines different factors that can help social security institutions harness AI. It will look at advantages and challenges in onboarding AI projects to aid social security institutions to embrace it as a key component of their technology portfolio.

1. Introduction

In social security institutions worldwide, the increasing demands for comprehensive support and delivery of essential services faces multiple challenges as well as resource constraints. This report delves into the critical issues surrounding service delivery in these organizations, highlighting the transformative potential of AI-based solutions.

As social security organizations strive to overcome challenges in providing crucial aid, AI emerges as a promising tool. AI stands poised to revolutionize service delivery, offering potential solutions to streamline processes, enhance the use of resources, decision-making and access to vital services for individuals in need. However, the successful application of AI in this context requires a meticulous approach, ensuring the safety and contextualization of algorithms to ensure the alignment of the technology application to the business objectives and to prevent any inadvertent harm. The successful application of AI also requires data quality and management capacity, which are key components of AI systems implementation.

Social security institutions are increasingly applying AI combined with other so-called data-driven technologies, notably analytics and big data [1]. Briefly, a data-driven strategy in social security aims to leverage growing data resources to improve services and decision making processes [2]. It is an ongoing effort that involves the integration of digital technologies and data to improve business processes, create new business models, and deliver better services to customers [3]. As part of a data-driven strategy, the emerging adoption of AI-based technologies by social security institutions, which have various forms of data as their raw material, enables more proactive and automated service delivery, improving efficiency, effectiveness and responsiveness. This way, institutions can leverage data analysis to identify areas where services and benefits are lacking or need improvement. The advent of AI fields – such as machine learning, pattern recognition, natural language processing, computer vision and data visualization is shaping the way data in various forms can be used to make social service more effective and user-centric.

This report will explore the profound implications of responsible AI deployment within social security organizations. It will highlight the pivotal role AI has, and can bring, underscoring how its conscientious use, integrated with other technologies, can address the challenges of improving service delivery and backend processes.

Furthermore, this report will analyze the intricate web of challenges – both technical and non-technical – that impede the seamless integration of AI in data-centric social security institutions. From inadequate infrastructure and limited expertise, to ethical and cultural considerations, these hurdles pose substantial barriers requiring thoughtful navigation for effective implementation [4].

By exploring these issues, the report aims to outline the potential benefits of AI adoption in social security institutions, underlining the necessity of responsible and contextually-aware AI frameworks to foster better service delivery. It also provides insight into the required institutional capacity for a responsible AI application, as well as scenarios in which other technologies are better suited.

The subsequent parts of the report is structured as follows. Section 2 introduces AI in the context of social security. Section 3 presents AI application experiences in social security. Section 4 describes the global uptake of AI and ongoing regulations. Section 5 introduces approaches to assess AI readiness. Section 6 discusses institutional challenges in implementing data-centric AI approaches. Section 7 concludes the report and describes ongoing work.

2. Artificial intelligence and the challenges of social security organizations

AI is a groundbreaking field within computer science that empowers machines and devices to undertake tasks that traditionally demand human intelligence and cognitive prowess. These tasks encompass learning, creating, reasoning, translating, problem-solving and decision-making. AI encompasses various types, each possessing distinct levels of sophistication, capabilities, and applications. The primary classifications of AI include:

- **Artificial Narrow Intelligence (ANI)**, also referred to as "weak AI", is tailored to execute specific tasks within its programmed scope. ANI lacks the capacity to extend its learning beyond predefined capabilities. Examples of ANI include voice-controlled virtual assistants like Amazon's Alexa and Apple's Siri, employing speech recognition technology. Self-driving vehicles, such as Tesla's, utilizing vision recognition and image processing AI. Streaming platforms like Netflix, leveraging user data for personalized recommendations. Computer vision systems, enabling machines to identify and comprehend objects and individuals in images and videos, mimicking human visual cognition. ANI finds applications across diverse sectors, including health care, finance, manufacturing, customer service, security, and data science and analysis.
- **Artificial General Intelligence (AGI)**, also known as "strong AI", aims to mimic human-like intellectual capabilities. AGI endeavors to learn and adapt to new situations akin to human cognition, without confinement to singular tasks or domains. AGI has potential applications in various fields, from robotics to health care and transportation, promising enhanced efficiency and productivity
- **Artificial Super Intelligence (ASI)** transcends human-level intelligence, potentially surpassing humans across all realms of knowledge and activities. While ASI remains a theoretical concept with no practical realization to date, it fuels significant discussions and debates within the AI community.

Most of the implementation of AI in the social security sector can be categorized within ANI. In subsequent sections AI will be interchangeably used to signify ANI.

2.1. Brief overview of AI and its branches

The algorithmic processes of AI are unlike traditional processes and their application may be better suited than those of the latter for many tasks. In the case of AI, instead of writing a programme for each specific task, many examples are collected that specify the correct (or incorrect) output of a given input. AI algorithms then take this data as examples to produce a programme that performs the tasks and that is applicable to new cases. Programmes adapt to the changes in data as the essence of AI is to re-train themselves using the new data. As computational power continues to become more readily available for these tasks, its use becomes cheaper and a more viable option where other task-specific programme would have been the only option. This capability of scalability and harnessing insights from data has made AI an essential and complementary tool for policymakers and service providers aiming at the social good. Various AI tools are being used for: responding to a crisis; promoting economic empowerment; alleviating educational challenges; mitigating environmental challenges; ensuring equality and inclusion;

promoting health; information verification and validation; infrastructure management; public and social sector management; and even for security and fairness.

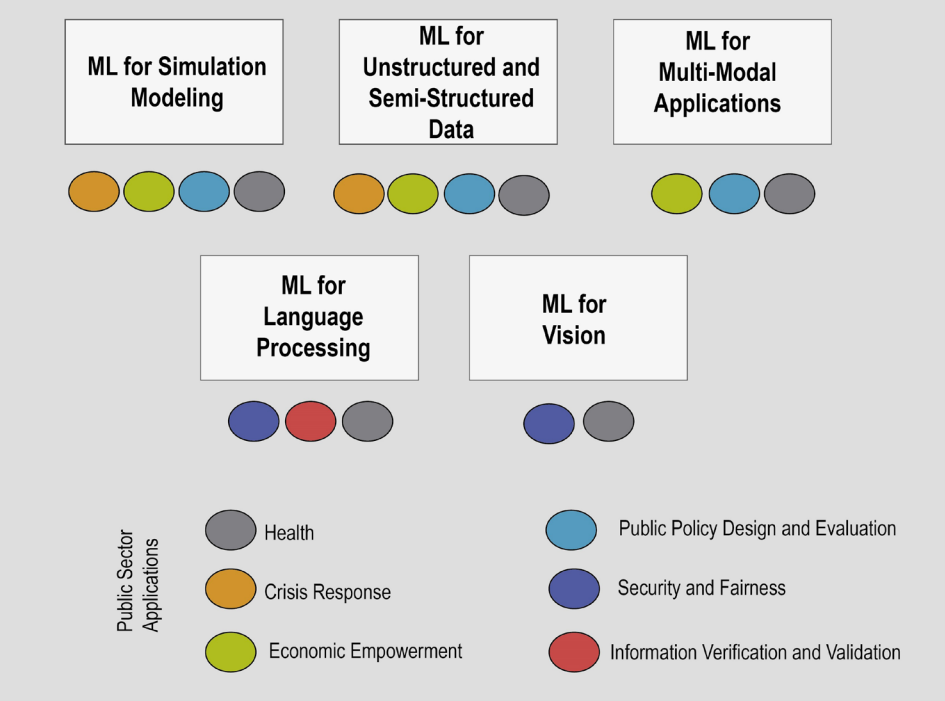
AI is a wide-reaching field characterized by its diverse techniques and methodologies which aim to mimic intelligent behaviour in machines and isn't a monolithic entity, but rather a diverse ecosystem with distinct branches working in tandem. Each branch focuses on different aspects of intelligence and cognitive processing. A few of the main branches of AI include:

- **Natural Language Processing (NLP)**, which empowers machines to understand and generate human language, enabling applications like chatbots and sentiment analysis.
- **Computer Vision**, which equips machines with the ability to "see" and interpret the visual world, facilitating tasks like image recognition and object detection in self-driving cars.
- **Robotics**, which enhances the capabilities of robots, enabling them to perform complex tasks autonomously or with minimal human intervention. AI integrates various disciplines to empower robots with the ability to perceive, understand, and interact with their environment effectively.
- **Machine Learning (ML)** that provides algorithms with the ability to learn from data without explicit programming, enabling them to improve their performance over time. ML allows programmes to learn from patterns and make inferences. A further subset of ML is Deep Learning, which utilizes complex neural networks with multiple layers to process data, facilitating sophisticated capabilities such as image and speech recognition. This underpins many AI applications, from spam filtering to personalized recommendations.
- Finally, **Data Science (DS)** encompasses the entire process of extracting knowledge and insights from data, playing a crucial role in preparing and analyzing data for use in various AI applications. DS overlaps with these AI domains as it applies analytical methods from AI and ML to interpret and analyze complex datasets, thereby aiding informed decision-making and prediction. Beyond the realms of AI, DS also spans a wide array of techniques from statistics and data analysis, among others, to derive insights and understanding from data.

Figure 1 represents how different categories of machine learning can be applied in different social security applications.

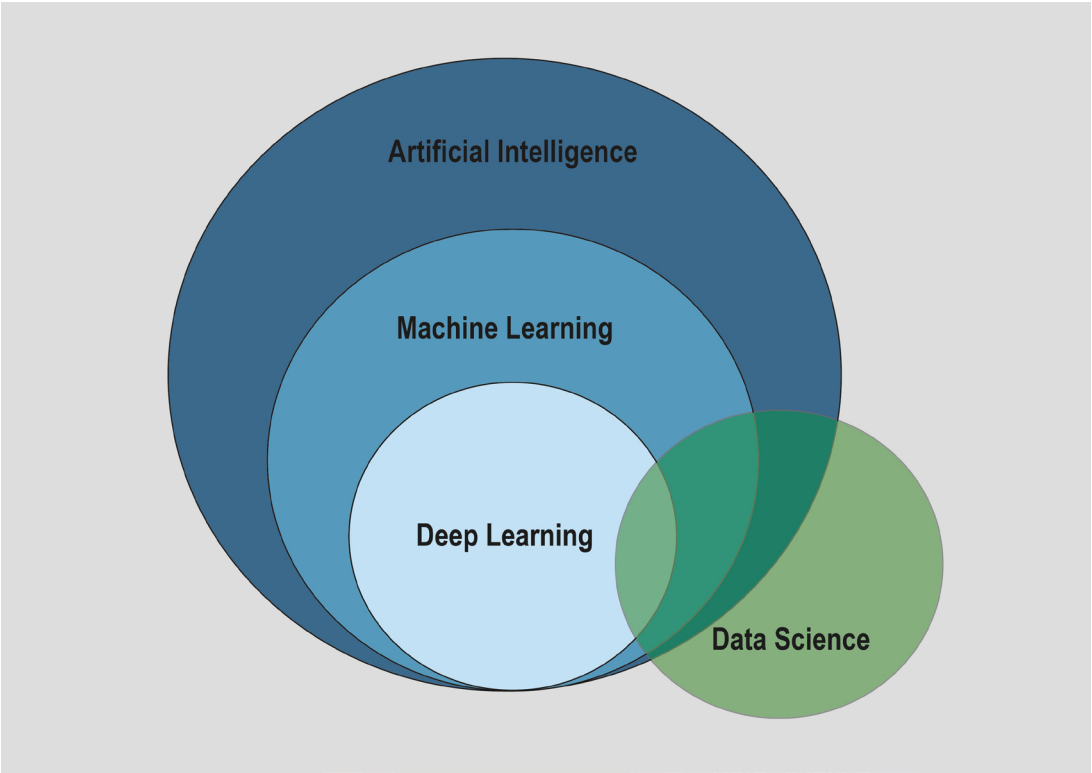
These interconnected branches work together to push the boundaries of what AI can achieve. This relationship is visually depicted in Figure 2, which shows the branches of AI.

Figure 1. *Machine learning categories as they are applied in different public sector applications*



Source: Authors' elaboration, with specific examples found in Annex, Table A.1.

Figure 2. *The relationship between AI and its sub-genres*



Source: Authors' elaboration.

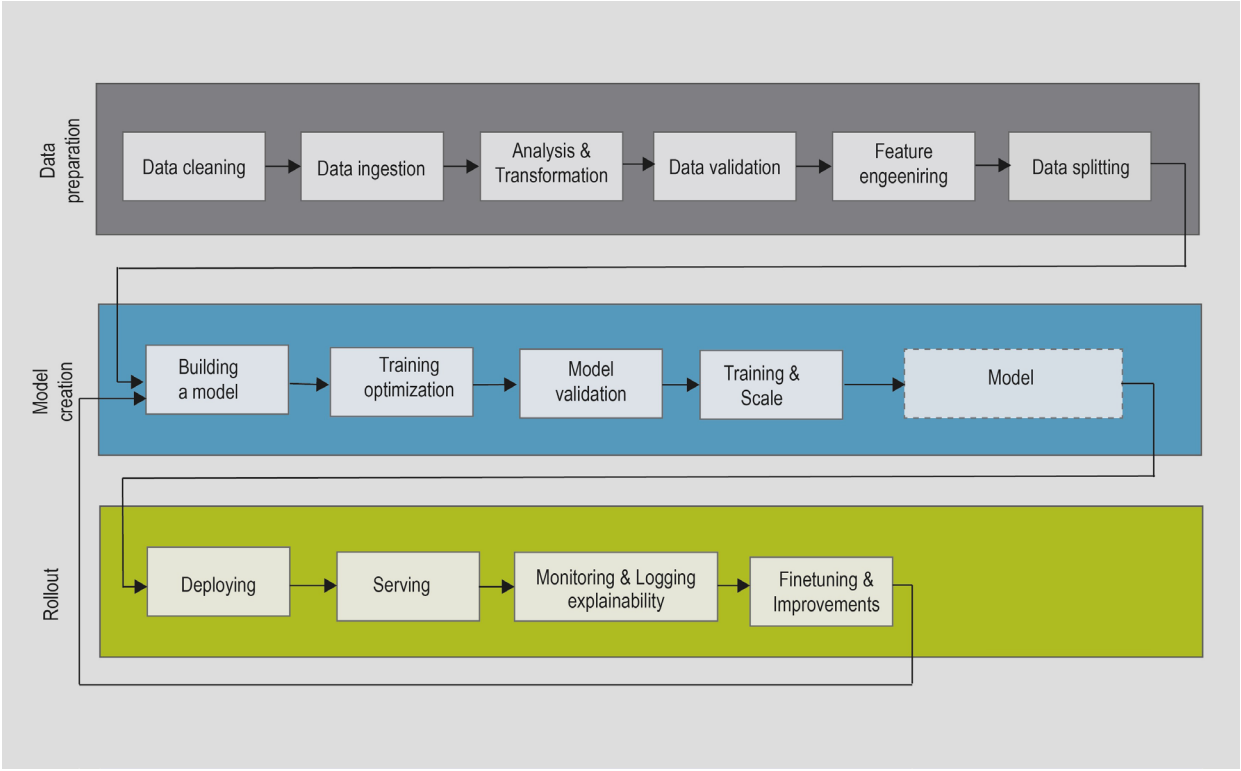
2.2. Machine learning and the need for better-quality data

Among the AI sub-fields, machine learning (ML) deals with the process of learning, reasoning, pattern finding and decision-making. It is an umbrella of methods that help build practical tools for other branches of AI. In short, ML is about finding ways to learn intelligently.

A learning problem can be defined as the problem of improving a measure of performance when executing tasks, using some type of training experience. For example, in learning to determine benefit eligibility in a case-management scenario, the task is to determine "eligible" or "not eligible" for any given resident's application. The performance metric may be to measure the accuracy of this eligibility classifier. The algorithm may be trained from a dataset containing historical eligibility information of applications, each of which is labeled in retrospect as being eligible or not. There may be many other alternate accuracy measures and training sets mixed with labeled and unlabeled data. Machine learning can be broadly categorized into three main branches: supervised learning, unsupervised learning, and reinforcement learning.

In the machine learning pipeline illustrated in Figure 3, the process is organized into three primary stages: data preparation, model creation and rollout. The data preparation stage encompasses data cleaning, ingestion, analysis and transformation, validation, feature engineering and data splitting. Subsequently, the model creation stage involves building a model, training optimization, model validation and training at scale to finalize the model. The rollout stage is focused on deploying, serving, monitoring and logging with explainability, followed by fine-tuning and improvements to enhance the model's performance post-deployment.

Figure 3. *A machine learning process pipeline showing how raw data is transformed into output*

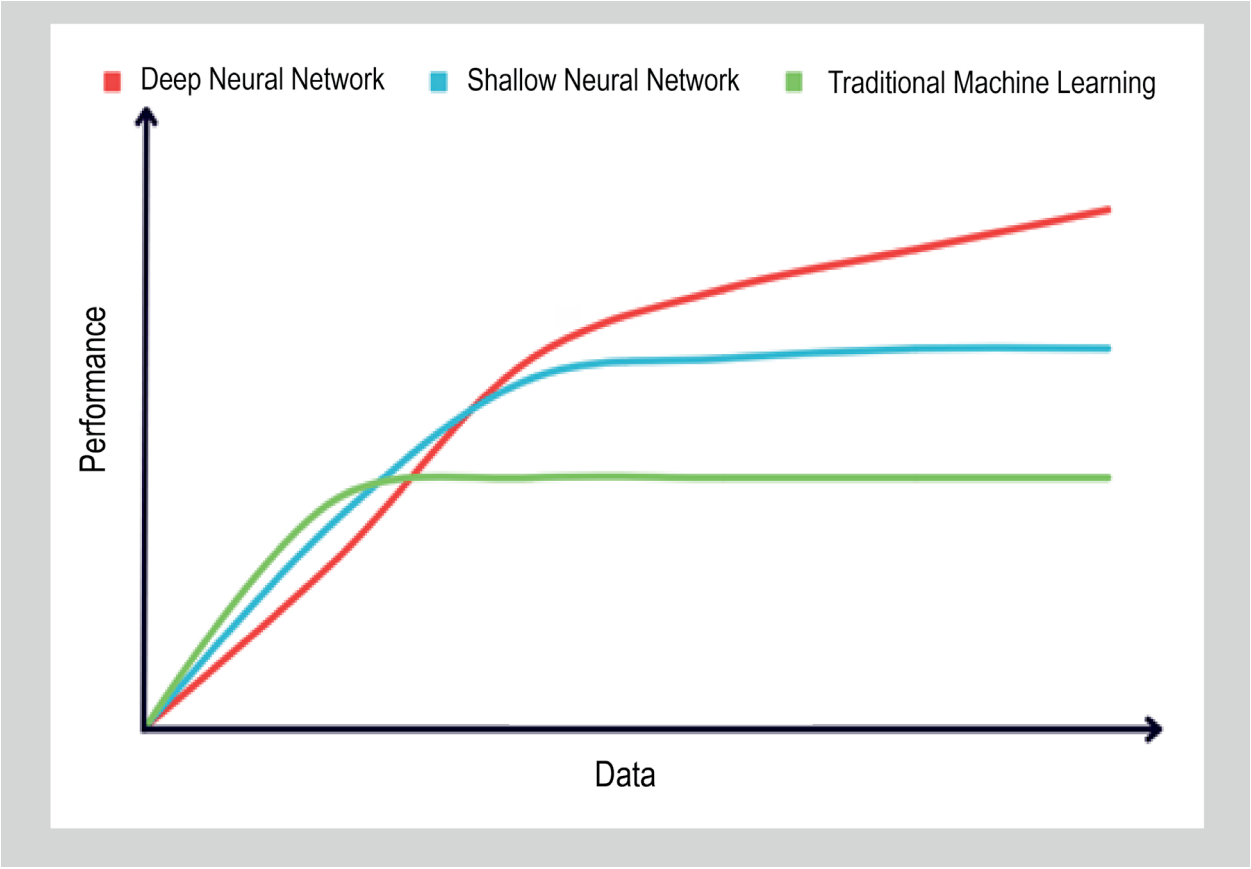


Source: Authors' elaboration.

Figure 4 depicts a comparative analysis of machine learning model performance as a function of increasing data volume. It is evident from the graph that with a smaller dataset, the performance of all three model types is somewhat similar. However, as the volume of data increases, deep neural networks exhibit a significant improvement in performance compared to shallow neural networks and traditional machine learning models, which level off, indicating a plateau in their ability to leverage additional data for performance gains.

By gathering and analyzing data on the supply side, such as response rates, customer satisfaction, waiting times and even responses themselves, it is possible to improve understanding of the customers, as well as the services and level of engagement social security institutions can provide to their members. AI-powered chatbots have served as excellent examples of how virtual assistants can provide round-the-clock assistance to citizens seeking information, or help with government services. On the demand side, with AI, government agencies such as social security institutions can identify areas that require new implementation strategies, allocate resources in the implementation of policies and even recommend new policies. For instance, by analyzing demographic and socioeconomic data, AI can help public officials identify disparities in the provision of services and their delivery and take action to address them. Machine learning algorithms are used to identify patterns in data. They can help government agencies detect fraud, waste and abuse, thus saving resources and improving the overall efficiency of the services on offer. Predictive analytics can be used to anticipate future needs and trends and enable government agencies to plan and allocate resources more effectively.

Figure 4. *Illustration of how machine learning algorithms increase their performance as the model gets exposed to more data*



Source: Authors' elaboration based on [5].

In planning for managing risks against unforeseen events, predictive analytics can help governments anticipate spikes in demand for emergency services during certain times of the year. Data visualization tools can present complex data in such a way that is easy to understand and interpret. This can help government agencies communicate with citizens more effectively and make data-driven decisions. Data and AI can also be used to identify potential risks and threats in real time, thereby helping organizations plan for, and mitigate crisis situations.

It is important, however, to ensure that these technologies are implemented responsibly, with appropriate safeguards to protect privacy and prevent bias. Since AI relies heavily on data to train models and make predictions, organizations need to ensure that their data is of high quality, reliable and accessible. This includes ensuring the legal use of personal data and respecting fairness, transparency and privacy, aligned with national data protection regulation. AI and data-based interventions also need to be integrated with systems, based on traditional technologies and practices of the agencies to ensure effectiveness. AI should be used when it is the best fit, considering the pros and cons of the different technologies.

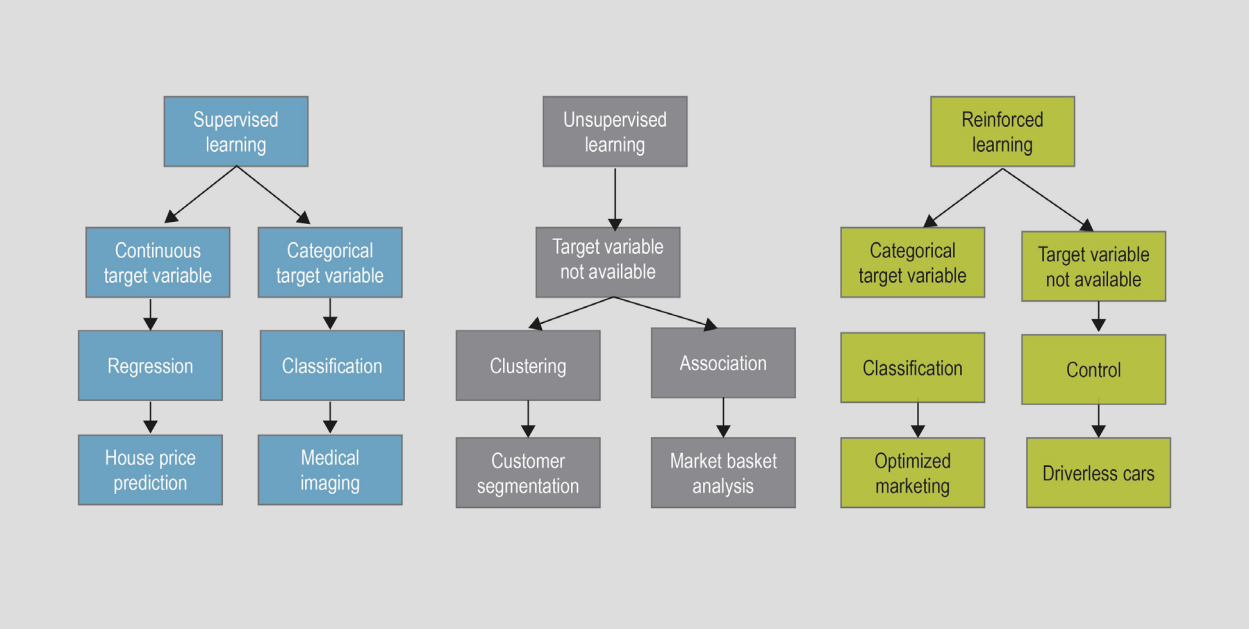
In order to leverage many of these advances and address these challenges, organizations will need to invest in human capital to reduce the skills gap. As new AI-based data-centric automation is implemented, skills in data science, machine learning and AI development are required. Computing traditionally focuses on code, while AI focuses on data. The performance of an AI-based system depends on a continuous supply of good-quality data that reinforces the AI solution. The accuracy and reliability of the results generated by the algorithms are heavily related to the quality of the data used to train them, just as much as the strengthening of the solution through continuous supervision. For example, machine learning algorithms rely on patterns and relationships in the data to make predictions or classify data. If the data is inaccurate, incomplete, or contains errors, the algorithm may learn incorrect patterns or make inaccurate predictions, leading to poor performance. In addition, the absence of quality data generates biased patterns that can result in discriminatory or unfair predictions. It is important to note that these systems are designed to be continuous learners. Hence there should be a permanent supply of good-quality data, with interventions on the solution to underscore the desired outcome, otherwise the output of the algorithms will not be commensurate with the context.

2.3. Types of machine learning algorithms

For many applications, it can be far easier to train a system by showing it examples of desired input-output behaviour than to programme it manually by anticipating the desired response for all possible inputs. Supervised learning is a type of machine learning in which the algorithm is trained on labeled data. Input here is paired with the desired output. The algorithm learns to map the input data to the output data, allowing it to make predictions on new, unseen data instances. Predicting whether an email is spam or not is an example of classification, while predicting the price of a house based on its features is an example of regression – two types of supervised learning algorithms. Figure 5 presents an overview of machine learning types, categorizing them into supervised learning, unsupervised learning, and reinforcement learning.

In unsupervised learning the algorithm is trained on unlabeled data, meaning that input data is not paired with any output variable. The algorithm learns to identify patterns and structure in the data without any prior knowledge of what the output should be. Clustering tasks, such as grouping customers into segments based on their purchase behavior, and dimensionality reduction tasks – such as reducing the number of variables in a dataset, are examples of unsupervised learning. Table 1 outlines five machine learning paradigms, each characterized by its use of data, data type and illustrative examples.

Figure 5. *Types of machine learning*



Source: Authors' elaboration.

Table 1. *Comparison of machine learning types by data usage and examples*

ML Type	Use of Data	Data Type	Example
Supervised	Large volumes, often thousands to millions of examples	Labeled data (input-output pairs)	Image recognition where photos are labeled with the object they contain
Unsupervised	Can vary, less dependent on volume than on data diversity	Unlabeled data	Market segmentation based on customer data without predefined categories
Semi-Supervised	Smaller amounts of labeled data, larger amounts of unlabeled data	Combination of labeled and unlabeled data	Enhancing the accuracy of a speech recognition system with limited labeled audio samples
Reinforcement	Data generated through extensive interactions over time	Feedback from the environment (rewards/punishments)	A robotic arm learning to pick up objects through trial and error
Generative AI	Large datasets for training, but specific amounts can vary	Can be either labeled or unlabeled, depending on the specific generative model	Generating new images that resemble the training set, such as creating new artworks in the style of a given artist

Source: Authors' elaboration.

Reinforcement learning is a type of machine learning in which an agent learns to interact with an environment, learns by trial and error and performs actions that maximize a reward signal. To illustrate how this works: in trying to teach a pet a certain task, for example, we may give them a treat (reward) if it performs the task correctly. If it does not, then we may say "NO" indicating a penalty. Over time, the pet gets to associate the correct behavior with the reward, and it gets better at performing the task. Reinforcement learning can be used in chatbots or conversational agents to improve their performance in understanding or responding to user queries leading to better user experience and increased satisfaction.

One of the high-impact areas of progress in supervised learning involves deep networks. Deep learning systems make use of gradient-based optimization algorithms to adjust parameters throughout multilayered networks based on errors in their output. Deep networks are based on an artificial neural network algorithm that is modeled after the structure and function of the human brain. Deep learning allows machines to learn from vast amounts of data and to recognize patterns that might be difficult or impossible for humans to identify.

Figure 6 demonstrates a neural network architecture for image classification. It learns from a dataset of labeled images, using layers of neurons to progressively detect features from simple patterns like edges to more complex shapes and animal parts. The network's final layer makes the prediction, and its accuracy improves through backpropagation, where it adjusts neuron connection weights to minimize prediction errors by learning from its mistakes.

Newer versions of deep learning algorithms give way to generative AI and foundation models. Generative AI represents a potent subset of artificial intelligence dedicated to producing entirely new content across various mediums, encompassing text, images and even music. Prominent tools like DALL-E 2 and Midjourney empower users to generate lifelike images based on textual cues, while platforms like Bard and GPT-3 excel in crafting diverse creative outputs such as poems, scripts and musical compositions.

Figure 6. *Inside a multi-layered neural network*

Source: Authors' elaboration.

In the public sector, the application of generative AI holds great promise:

- **Creative content creation:** Utilizing generative AI to craft educational resources, devise public awareness initiatives, or curate engaging content for social media campaigns.
- **Data augmentation:** Employing synthetic data generation to enhance the performance of AI models used in critical areas such as health care and environmental monitoring.
- **Personalized user interactions:** Tailoring governmental services and information dissemination to cater to individual preferences and requirements.

Nonetheless, challenges accompany the adoption of generative AI:

- **Bias and misinformation:** Given that generative AI models are trained on existing datasets that are prone to biases, cautionary measures such as meticulous data curation and rigorous model evaluation are imperative to mitigate biased outputs.
- **Deepfakes and malicious intent:** The realistic content generated by these systems can be exploited for nefarious purposes like creating deepfakes or disseminating misinformation, underscoring the necessity for robust regulatory frameworks and safeguards.
- **Ethical implications:** Discussions surrounding issues such as content ownership, potential workforce displacement and broader societal ramifications necessitate ongoing deliberation and the establishment of ethical guidelines.

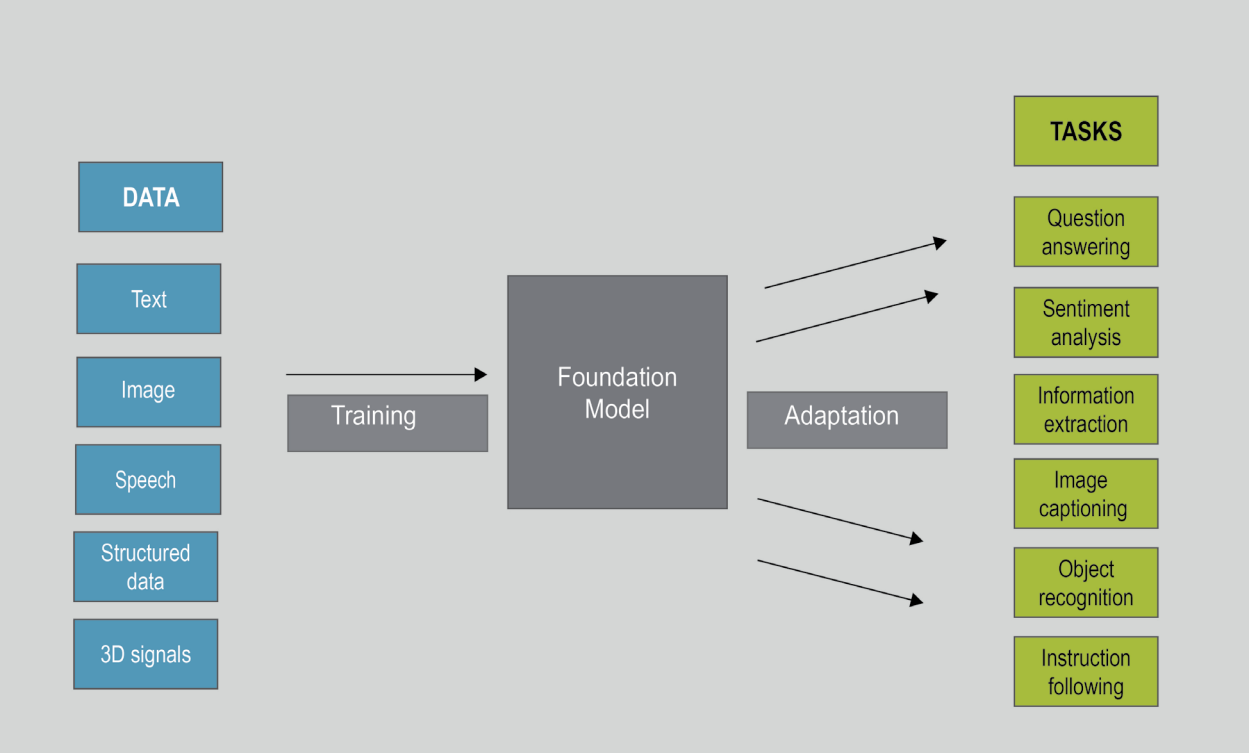
Effectively addressing these challenges is essential to unlock the full potential of generative AI while upholding principles of responsibility and ethics in its utilization within the public sector.

Deep learning models are often regarded as the fundamental building blocks of AI. They represent massive AI systems trained on extensive datasets, enabling them to discern general patterns across various domains such as language, code and images. Comparable to a well-read scholar who has perused an entire library, these models have acquired foundational knowledge and can undergo fine-tuning processes to cater to specific tasks.

Generative AI often relies on foundation models as underlying frameworks, leveraging their comprehensive understanding of data patterns to generate novel content across various modalities, from text to images and music. Foundation models provide the groundwork and knowledge base necessary for generative AI systems to produce original and contextually relevant outputs.

Foundation models are trained on vast datasets to learn general patterns across different domains, providing a broad understanding of language, code, images and more. Through fine-tuning processes, these models are adapted to specific tasks by focusing on relevant datasets, allowing them to specialize and excel in various applications. Figure 7 shows the diagram of a foundation model's workflow, illustrating the training from diverse data types to perform various AI tasks after adaptation.

Figure 7. *From training to task execution: A visual overview of a foundation model's learning and application process*



Source: Authors' elaboration based on [6].

By processing millions of data points, such as news article texts, these models decipher the structures of sentences, semantic relationships between words and even the synthesis of overarching themes. This process is similar to teaching the model a comprehensive encyclopedia of knowledge, from which it can draw insights and make informed decisions. Subsequently, through the process of fine-tuning, these models can be tailored to suit diverse applications, where specific datasets pertinent to the task at hand are employed to refine the model's understanding and specialization. For instance, feeding the model legal cases and contracts facilitates its adaptation to comprehend other legal documents. Despite their versatility, it is important to acknowledge that foundation models, whether closed-source or open-source, possess inherent limitations and potential biases. Thus, meticulous evaluation and responsible utilization are paramount, particularly in government settings where their deployment could significantly impact decision-making processes.

2.4. Data as the raw material of machine learning

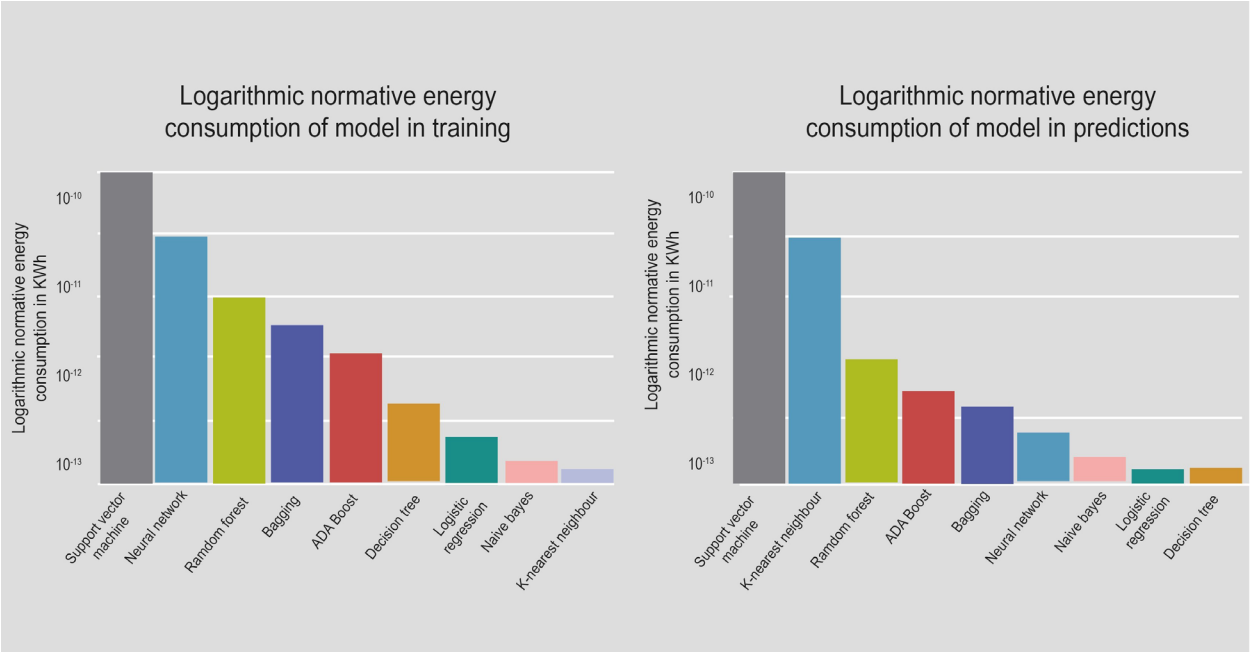
Machine learning and all other branches of AI are heavily reliant on the availability of training data. The amount of labeled data required to train a machine learning model varies depending on the dataset and model adopted. The requirement increases with the complexity of the datasets and the "depth" of the models. Deep networks allow far more generalization than shallow neural networks and traditional machine learning approaches, therefore they achieve significantly better accuracy. When applying deep learning to a problem the key challenge is the large amount of data required to train the models.

The algorithmic processes of AI are unlike traditional processes and may be better suited for many tasks. In the case of AI, instead of writing a programme for each specific task, many examples are collected that specify the correct (or incorrect) output of a given input. AI algorithms then take these examples to produce a programme that performs the tasks and that are applicable for new cases. Programmes adapt to the changes in data as the essence of AI programmes is to re-train themselves using the new data. As computational power becomes more readily available for these tasks, the use of these tools become more accessible and in some cases preferable to writing a task-specific programme. This capability of scalability and harnessing insights from data has made AI an essential and complementary tool for policymakers and service providers aiming at the social good.

2.5. Sustainability and AI

The use of enormous amounts of data and computing power puts heavy pressure on energy consumption. Green AI aims to address the growing concern of energy consumption in the AI domain. As AI models become more powerful, their carbon footprint increases. The pursuit of sustainability demands that we develop innovative strategies to reduce the energy demands of these models without sacrificing performance. This involves optimizing hardware for lower power usage, developing energy-efficient algorithms and exploring ways to integrate renewable energy sources within AI infrastructure. Additionally, researchers are emphasizing resource saving algorithms as mode of greener AI approach. In deep learning, the model's parameter count can provide an initial assessment of potential energy efficiency advantages. Smaller models often consume less energy. Many researchers are also focusing on the importance of comparing energy usage across different types of machine learning algorithms, which is complex due to their varied working principles. However, as of the writing, scientists have not yet come to a consensus on what can be a practical benchmarking for such purpose. Figure 8 shows the comparative bar graphs of logarithmic normative energy consumption for various models during training and prediction phases.

Figure 8. *Logarithmic normative energy consumption of models during training and prediction*



Source: Authors' elaboration based on [7].

2.6. Learning with limited data

When the amount of data is limited, models are often provided with biased data. Sometimes the designed algorithms adjust so well to the training dataset that they fail to give the right solution in the real world. In many cases it is difficult to accumulate large amounts of data, or perhaps it does not even exist (Defence Science and Technnology Laboratory [8]. For example, an institution wants to create a model to predict certain traits of the users, but they only have the historical data of 50 users. Traditional machine learning approaches to this model may have a biased output that goes against gender, age and ethnicity due to a lack of variation. Machine learning problems associated with small datasets require a different set of techniques and approaches compared to those involving large data, such as: engineering that helps create new features; regularization to avoid over-combining outputs from multiple models; and learning transfer using a model that has been pretrained on a larger dataset. Table 2 shows different types of machine learning methods based on the amount of data available.

Table 2. *Selection of different types of machine learning methods based on the amount of data*

Use of Data	Labeled or Unlabeled	Learning Method Used	Comment
Small amount	Mostly unlabeled	Zero-shot learning	Uses description of concept to train the model, concept ontology, semantic word embedding
Small amount	Mostly unlabeled	Manual labeling	Manually label the data
Small amount	Mostly labeled	Shallow machine learning, meta-learning, knowledge reasoning	Train a meta model to be applied for unseen tasks, support vector machines, decision trees, multi-layer perceptron, ontological approach leveraging description of objects
Large amount	Mostly labeled	Deep learning	Convolutional neural network
Large amount	Mostly unlabeled	Active learning, semi-supervised learning, self-supervised learning, Unsupervised learning	Utilizing both labeled and unlabeled data, query to select examples for a human operator to label, clustering, anomaly detection, latent variable, autonomous labelling

Source: Authors' elaboration.

2.7. The evolving landscape of machine learning

In the evolving landscape of machine learning and its intersection with human rights, social biases and ethical considerations, several seminal studies have been conducted, shedding light on both the promise and the hurdles of these technologies. We have explored some academic works that are relevant for the readers to get a brief understanding of the state-of-the-art of research that cross cuts between machine learning and social security.

2.7.1. Research focusing on human rights and social biases

In the realm of human rights, the application of machine learning algorithms has been explored by Greene, Park and Colaresi [9]. Their study focuses on the evolution of human rights standards over time. By mapping textual features in reports to human-coded scores, they were able to analyze these changes. Their findings reveal a dynamic landscape of human rights standards, highlighting the evolving interpretations of abuses and the consistent application of standards. This work underscores the potential of machine learning in providing valuable insights into the complex and ever-changing field of human rights.

On the other hand, the issue of social biases in machine learning algorithms is addressed by Sirotkin, Carballeira and Escudero-Viñolo [10]. In their study, they delve into the distribution of social biases in self-supervised learning (SSL) visual models, particularly those trained on datasets like ImageNet. Their research emphasizes the inherent biases present in deep neural networks and stresses the importance of careful model selection to balance bias reduction and performance. This study brings to light the critical need for addressing social biases in machine learning, highlighting the potential pitfalls and challenges in the development and application of these technologies.

2.7.2. Research focusing on methodology and ethical considerations

In the field of methodology and ethical considerations that need to be considered when using machine learning, Navarro et al [11] delve into the methodological quality of studies that employ supervized machine learning techniques. In their systematic review Risk of bias in studies on prediction models developed using supervised machine learning techniques, facilitated by the PROBAST tool, they highlight prevalent biases in these studies. They emphasize the importance of a thorough assessment across multiple domains to ensure the reliability and validity of the findings. This work underscores the need for rigorous methodological scrutiny in the application of machine learning in medical research to mitigate bias and enhance the quality of outcomes.

On the other hand, Watson [12] provides a philosophical lens to view unsupervised learning algorithms. Recognizing the potential of these algorithms in identifying natural kinds, the paper also highlights the ethical and epistemic risks that come with the use of unsupervised methods. Watson advocates for a pragmatic approach and a comprehensive analysis of generative models to navigate these risks. This philosophical insight into unsupervised learning underscores the need for a balanced approach that acknowledges both the potential benefits and the inherent risks of these algorithms. It serves as a reminder that progress in AI and machine learning must be pragmatic and include a comprehensive analysis of generative models.

2.7.3. Research focusing on practical applications and responsible development

The following studies represent a diverse range of applications and considerations in the field of machine learning and artificial intelligence. They explore practical applications in social security, health care, data collection, dataset documentation, automated decision systems and cybersecurity. Each study underscores the potential benefits of machine learning, while also highlighting the challenges of responsible development. From enhancing predictive accuracy in social security systems to integrating data cleaning with explainable AI in cybersecurity, these studies collectively illuminate the transformative potential and complexities of machine learning across various domains. They emphasize the importance of responsible AI practices, including fairness, accountability, transparency and human-centered design, in realizing the benefits of AI while mitigating potential risks.

The study by Sansone and Zhu shows the potential of machine learning in improving social security systems [13]. Their research demonstrates the efficacy of machine learning algorithms in enhancing predictive accuracy for beneficiaries in the Australian social security system. This highlights the potential benefits of machine learning in social security, while also posing challenges in ensuring the accuracy and fairness of such predictive systems.

In looking at AI applications in health care, Oprescu et al. [14] provide valuable insights into the expectations and preferences of pregnant women regarding AI applications in pregnancy care. Their work emphasizes the importance of responsible, trustworthy and safe AI solutions in health care. It underscores the potential of AI in improving patient care, while also highlighting the need for responsabiilty and accountability in its application.

Inel, Draws and Aroyo [15] propose a methodology focusing on data quality and reliability to address fairness and accountability in AI data collection. Their work underscores the importance of responsible data collection in the development and propose a methodology focusing on data quality and reliability to address fairness and accountability in AI data collection.

The importance of enhancing transparency in dataset documentation was reviewed by the study by Pushkarna, Zaldivar and Kjartansson [16] which introduces a structured approach for documenting machine learning datasets. Their work aims to enhance transparency and reflect a human-centered design approach in AI. This highlights the importance of clear and comprehensive dataset documentation in promoting responsible AI practices

Stoyanovich, Howe and Jagadish [17] address the societal impact of automated decision systems (ADS) advocating for responsible ADS design and oversight within the data management community. Their work emphasizes fairness, equity, accountability and transparency in ADS, highlighting the potential benefits and challenges of these systems in various domains, including human rights and public safety

Finally, the importance of integrating data cleaning with explainable AI in cybersecurity was reviewed by Hong et al. [18]. Their work presents a framework integrating data cleaning with explainable AI for intrusion detection, aiming to enhance decision-making in threat identification and mitigation in cybersecurity.

These studies collectively illuminate the diverse applications of machine learning, emphasizing its potential benefits and challenges across various domains, including human rights, health care, cybersecurity and public safety.

2.8. The challenges

2.8.1. Inefficient use of data for AI

AI tools replicate human thought processes and behaviours, implying that algorithms can often be inaccurate. Such inaccuracies pose risks to various aspects such as personal privacy, national security, fairness, transparency and accountability. Moreover, inaccuracy can affect data quality, algorithms and human interaction with the design process.

When AI systems are trained on biased datasets, they may perpetuate and even amplify existing biases in the data. Unrepresentative or homogeneous training data can lead to biased predictions by the AI system. The vast amount of data fed into machines to identify patterns, including unstructured data from sources like the web, social media, mobile devices, sensors and the Internet of Things (IoT), presents challenges in data absorption, linking, sorting and manipulation. Without meticulous data curation, datasets may contain incomplete, missing, inaccurate or biased data.

In terms of bias, there are four main types: sample bias, measurement bias, algorithmic bias and bias against groups or classes of objects and people. However, algorithmic bias is often underdiscussed. Correcting bias is challenging due to its subtle introduction during model construction, difficulty in retroactively identifying its origin, persistent biases in machine learning due to random data division for training, testing and validation, lack of contextual understanding, and variability of fairness across communities and institutions.

In addition, the risk of inadvertently revealing sensitive data exists if personal data is inconsistently removed across datasets. The responsibility for bias largely lies with developers curating data and designing algorithms, determining deployment strategies and ultimately regulating usage. The framing of problems by scientists and engineering team compositions can introduce bias, affecting the ethical deployment and application of AI.

Failure to address these issues has led to algorithms dictating political advertisements, job seeker filtering by recruiters and deployment of security agents, among other scenarios. Additionally, the human-machine interaction requires careful evaluation to prevent accidents and injuries caused by over-reliance on AI systems.

Human judgment remains crucial. Human overriding of AI systems can mitigate risks stemming from data management lapses, scripting errors, misjudgments in model training and unintentional bias induced by data collectors favoring certain demographics over others.

2.8.2. Importance of data governance

Traditionally public institutions work as silos that create barriers to data access and availability for other institutions. Lack of data access results in poor data analytics and AI tools. Implementing data-centric AI may require significant changes to existing IT systems and processes. This means that organizations need to ensure that AI solutions are integrated seamlessly with their systems and workflows without causing disruption.

The accuracy and reliability of AI models depends on the quality of the data used to train them. To get the most out of the data, it is imperative to employ a data governance model that manages and ensures the quality, accuracy, completeness and security of the data used to train and develop AI algorithms. In aiming to convert data into information, data must go through a pipeline that consists of a series of steps, and the results of one step may influence the next. There is a specific order that may not be linear, as data processing may be an iterative process. Figure 3 above shows a flowchart depicting the machine learning pipeline: data preparation, model creation, and rollout stages.

The steps start with data collection, in which raw data is gathered, it is then preprocessed, where data is cleaned and transformed to ensure quality. Data is then stored in various forms in data warehouses, from where it moves to the analysis phase, in which various patterns are identified. It is then modeled, using various mathematical models to detect anomalies or predict outcomes and finally moves on to visualization, during which the insights are visually summarized.

To ensure the proper management of these steps, an appropriate data governance model is required, which involves defining policies and procedures for each of the above-mentioned steps. This includes identifying the sources of data, establishing data quality standards, defining data ownership and stewardship, as well as ensuring compliance with relevant regulations and industry standards. It is important to ensure that the data used to train AI models is consistent, accurate and relevant.

Data governance for AI also involves establishing processes for data preparation and pre-processing, including data cleaning, normalization and feature engineering. At the policy level data governance also addresses ethical and privacy concerns.

2.8.3. Model contextualization

Model contextualization, particularly in advanced AI systems, presents various challenges that intersect with safety, ethics and the responsibility of AI developers and users.

From a safety perspective, contextualizing a model's responses appropriately is crucial to avoid misinformation or harmful advice. AI models, no matter how sophisticated, lack real-world understanding and experience. This limitation can lead to incorrect, inappropriate, or unsafe responses, especially in complex or nuanced situations. For instance, in medical or legal advice scenarios, even a slight mis-contextualization can lead to dangerous outcomes. Ensuring the safety of AI responses is a continuous challenge, requiring constant monitoring and improvement of the model's algorithms and training data.

Ethically, model contextualization raises questions about bias and representation. AI models are trained on large datasets that can include biased or unrepresentative information. If not carefully curated, these biases can be reflected in the model's responses, perpetuating stereotypes or unfair representations of certain groups or topics. Ensuring that AI systems are fair, unbiased and representative is an ongoing challenge that involves scrutinizing training data and algorithms for potential biases and implementing corrective measures.

Finally, the responsibility of AI touches on the accountability of both creators and users of AI technology. Creators of AI systems have a responsibility to design and train models that are safe, ethical and as unbiased as possible. This involves making conscious choices about the data used for training, the algorithms employed and the contexts in which the model is intended to operate. In addition, users of AI technology must also be responsible in how they employ AI systems, understanding their limitations

and avoiding reliance on them for critical decisions without human oversight. This dual responsibility emphasizes the need for clear guidelines and standards in the development and use of AI technologies.

Overall, the challenges of model contextualization in AI are multifaceted and continuously evolving. Addressing these challenges requires a concerted effort from AI developers, users and policymakers to ensure that AI systems are safe, ethical and responsibly used.

2.8.4. Ensuring AI safety and managing risk

Ensuring AI safety and managing risk in the development and deployment of artificial intelligence systems is a multifaceted task, requiring attention to various aspects such as the design of the AI, the data it is trained on and the environment in which it operates.

At the core of AI safety is the design and development process. This involves creating algorithms that are robust and reliable, capable of handling unexpected inputs or situations without malfunctioning or producing harmful outputs. It's crucial that AI systems are designed with fail-safes and mechanisms to detect and correct errors. This could include implementing checks that identify when an AI is operating outside of its reliable parameters and programming the system to seek human intervention in such cases. Additionally, AI models should be continuously updated and improved, incorporating new data and feedback to refine their decision-making capabilities.

The data used to train AI models significantly impacts their safety and reliability. Biased or incomplete datasets can lead to skewed or unfair AI decisions. Therefore, it's essential to use diverse, comprehensive and well-curated datasets that represent a wide range of scenarios and demographics. Regular audits and updates of the training data can help mitigate the risk of bias and ensure that the AI remains accurate and fair over time.

Another crucial aspect of ensuring AI safety is setting up appropriate testing environments. Before deployment, AI systems should undergo rigorous testing in controlled environments that simulate real-world scenarios as closely as possible. This testing should cover a wide range of conditions and edge cases to ensure that the AI can safely handle unexpected situations. Continuous monitoring post-deployment is also essential to promptly identify and address any issues that arise in real-world applications.

The ethical considerations in AI development cannot be overstated. AI developers and companies must adhere to ethical guidelines that prioritize human welfare, privacy and rights. This includes ensuring that AI systems do not infringe on privacy, propagate harmful biases, or cause unintended harm. Developing transparent AI systems, where decisions can be understood and scrutinized, also plays a key role in managing ethical risks.

Lastly, the responsibility for AI safety and risk management is not solely on the developers. Users of AI systems, including businesses and individuals, must be educated about the capabilities and limitations of AI. They should understand how to interact safely and effectively with AI systems and be aware of the potential risks involved. Regulatory bodies and policymakers also have a critical role in setting standards and guidelines for AI safety and ethical use, ensuring a balanced approach that fosters innovation while protecting the public interest.

In summary, ensuring AI safety and managing risks is a complex endeavor that requires careful consideration of the design, data, testing, ethical implications and regulation of AI systems. A collaborative approach

involving developers, users, and regulators is essential to harness the benefits of AI while minimizing potential harms.

2.9. Responsible and explainable AI

Machine learning algorithms learn patterns and relationships from vast amounts of data, often without explicit programming of the rules or decision-making criteria. As a consequence, the algorithms can produce results that are accurate but may not be intuitive or easily understood by humans. Much of AI (particularly deep learning) is plagued by the "black box problem." These models can be highly complex, with many layers and interconnected nodes. We often know the inputs and outputs of the model, but we do not know what happens in between. To ensure trust and accountability it is imperative to ascertain how an intelligent machine suggests certain decisions. Additionally, if AI systems become explainable, they may be able to significantly increase the benefits to organizations, increase the accuracy of models by 15 per cent to 30 per cent and reduce monitoring efforts by up to 50 per cent.

There are several reasons for the lack of explainability of machine learning algorithms. The foremost is related to data and their use. Machine learning algorithms can perpetuate biases present in the data resulting in outputs that reinforce existing societal biases. As a consequence, these models are also prone to adversarial attacks and biases. More significantly, and due to the black box nature of the algorithms, it is hard to pinpoint the features that cause such biases. Machine learning algorithms often operate in high-dimensional spaces. This results in non-linear relationships between features and output predictions making it difficult to explain.

Responsible Artificial intelligence constitutes a framework or guiding principles aimed at ensuring that AI systems are developed and utilized in a manner that upholds ethics, transparency, accountability and fairness. It encompasses a comprehensive array of considerations geared towards addressing the societal, ethical and legal ramifications of AI technologies. The overarching objective of responsible AI is to engender systems that not only efficiently fulfill their designated functions, but do so in a manner congruent with human values and ethical norms, fostering societal benefit while mitigating potential negative consequences. Key facets of responsible AI encompass:

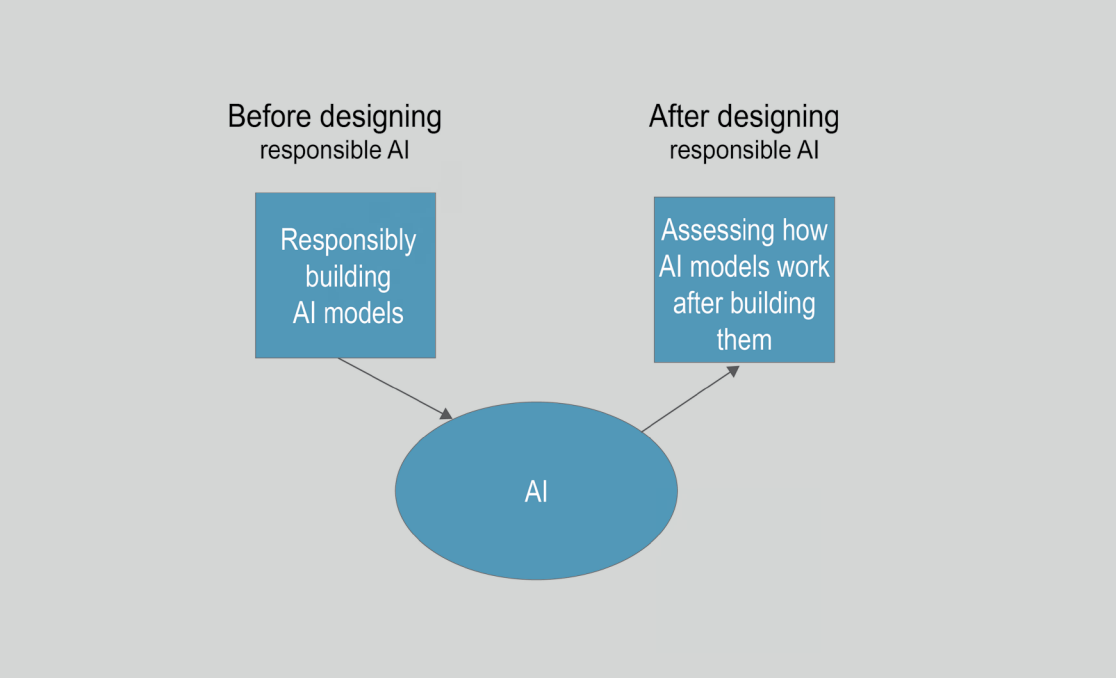
- **Ethical AI**, which embeds adherence to moral principles throughout the lifecycle of AI systems;
- **Transparent AI**, which underscores the importance of elucidating AI processes and decisions to facilitate user trust and validation;
- **Accountable AI**, which emphasizes the establishment of mechanisms for attributing responsibility for AI outcomes;
- **Fair AI**, which aims to eliminate bias and ensure impartial outcomes for all individuals;
- **Privacy-preserving AI**, which safeguards individuals' privacy rights and sensitive data;
- **Safe and secure AI**, which entail fortifying AI systems against both unintentional harm and malicious exploits;
- **Sustainable AI**, which promotes the development of environmentally sustainable AI solutions contributing to long-term planetary well-being.

Figure 9 shows a conceptual diagram highlighting that responsible AI paradigm is about setting the experimental environment before conception of AI design while explainable AI focuses on methods that will explain why machine behaves the way it did.

Making machine learning models explainable is an active research field. Some of the notable work done as mentioned by Došilović, Brčić and Hlupić [19] to understand the feature-output relationship are Shapley Additive exPlanations (SHAP); Local Interpretable Model-Agnostic Explanations (LIME); and Gradient-weighted Class Activation Mapping (Grad-CAM).

One of the popular counters to the black box problem is Explainable AI (XAI) – a set of machine learning processes that allows human users to comprehend, trust and manage AI. The goal of XAI is to enable interactions between people and AI systems by providing information about how decisions and events come about as mentioned by Tjoa and Guan in their survey on explainable artificial intelligence [20]. This has been so widely embraced that it is mentioned by the different publications. Hence, there is no expected result with many AI algorithms. Systems learn what the best prediction is, which makes it difficult to validate. Such unpredictability is a challenge. This means that auditing datasets and the output is not sufficient for evaluating AI tools.

Figure 9. *Reponsible AI focuses on pre-design environment setting and explainable AI focuses on post-design assessment*



Source: Authors' elaboration.

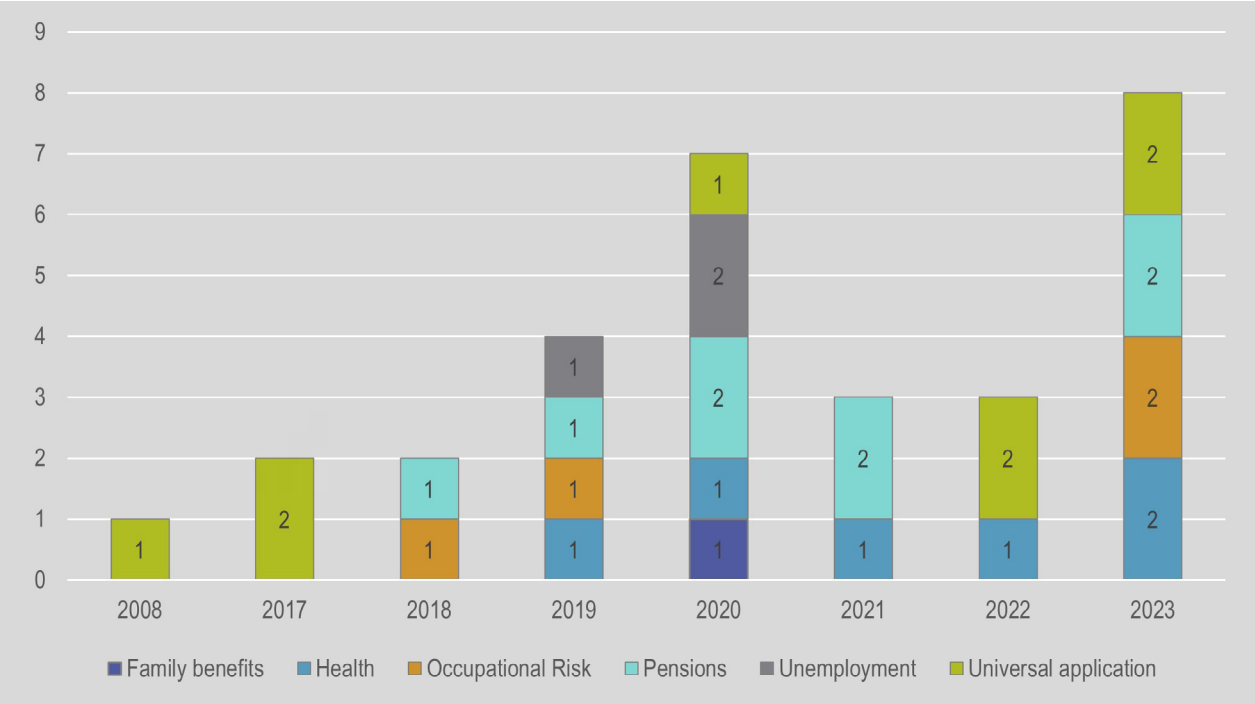
3. AI-based applications in social security

AI has been leveraged by social security institutions to improve their services in different branches with initiatives dating from 2008. However, as seen in Figure 10, the use of AI has seen a proliferation since 2018 with more than half of the observed experiences significantly increasing from 2020 onwards.

While most applications of artificial intelligence have been applied to universal solutions for different branches in social security; specific implementations of AI tools can already be seen in specific branches of social security. The predominance of implementations in AI have focused on pensions or health branches, however, some institutions have established applications that can be universally used in and across different branches.

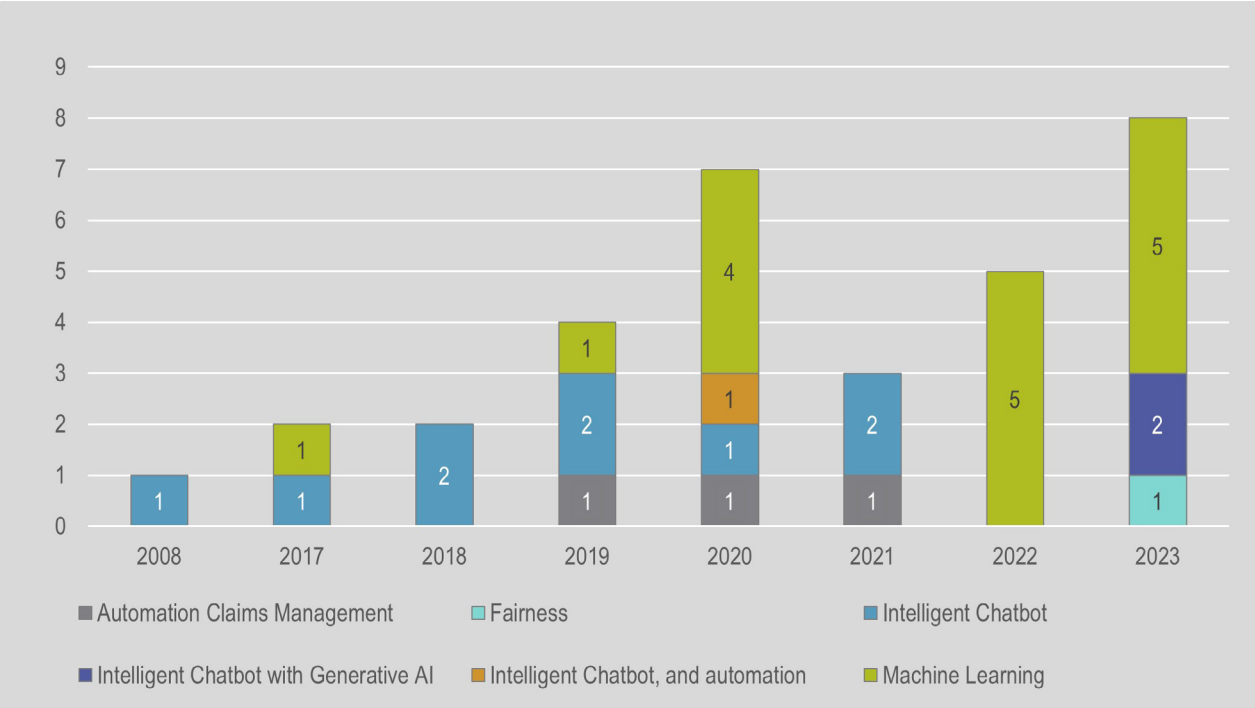
While artificial intelligence solutions dating before 2020 were predominately intelligent chatbots [21] [22], a clear tendency to leverage a broader range of AI tools have been proliferating from 2020 onward. Figure 11 shows how different social security institutions have significantly accelerated the use of machine learning techniques to improve automation and case management and even started to deploy generative artificial intelligence.

Figure 10. *Distribution of AI applications indifferent branches*



Source: Authors' elaboration based on ISSA Good Practices and other ISSA events.

Figure 11. *Type of AI applications in social security*



Source: Authors' elaboration.

Overall, AI has helped social security institutions to strengthen their capacity in different ways.

3.1. Strengthening administrative capacity through AI applications in social security

This section provides an overview of AI applications and implementations implemented in institutions and what capacities and functions this new technology has supported.

Analyzing the different applications of artificial intelligence in social security shows how different AI solutions have been supporting social security institutions. Figure 12 shows a topology of the type of capacity that has, or can be, significantly improved by the use of artificial intelligence. These capacities have been grouped from the documented AI solutions identified so far through different ISSA webinars, ISSA ICT conferences, good practices and other sources providing insights from the application of AI techniques and tools leveraged by institutions. It highlights the different ways they strengthen the organization and administrative support to provide better social outcomes. These have been identified as:

- **Service delivery:** Allows institutions to improve services by providing better, more opportune information leveraging channels with greater accessibility to different types of customers.
- **Automation and case management:** Refers to how social security institutions have used AI to automate the way they process cases as well as provide better grievance management support for individuals to follow up on a service.
- **Prospective and proactive social security:** Provides social security institutions with tools to gain insight, vision and prospective analysis to identify potential outcomes. Institutions can

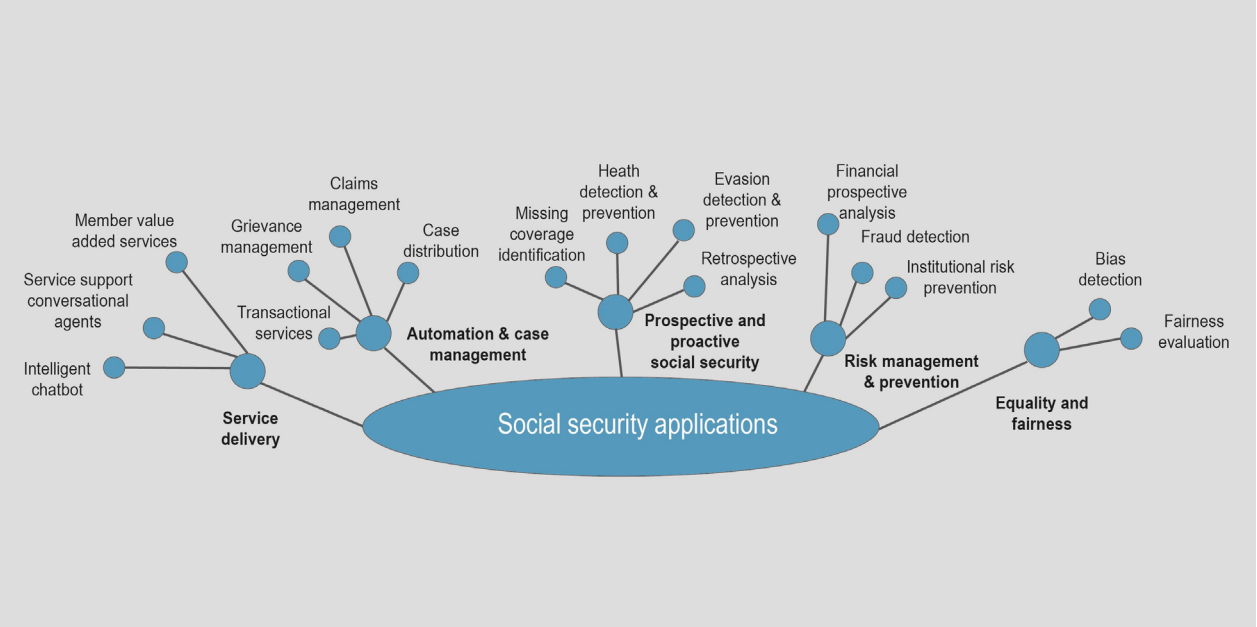
leverage these tools to establish approaches to proactively and pre-emptively improve cohorts and individuals' lives.

- **Risk management and prevention:** Supports institutions to identify risks, strengthening their capabilities to mitigate and/or take action on risks, as well as analyze member information that will allow them to provide better services.
- **Equality and fairness:** Evaluate the different AI solutions as well as programmes and responses in terms of the institution's mandate that would uphold the principles of fairness and equity.

Intelligent chatbots was among the most frequent solutions implemented, largely driven by the significant needs to unburden call centres and leverage digital channels. The surge of these was also seen during 2020 and coincided with the need for solutions that could help social security institutions cope with the surge in demand for services during the COVID-19 crisis. Within the pensions branch, the solutions predominantly look at leveraging intelligent chatbots, some of which have included the use of generative AI as part of their chatbot solution.

From 2020 onwards intelligent chatbots have continued to become more sophisticated, with constant review of the output provided by intelligent chatbots to ensure the adequacy of the ICT system response, as well as part of the continuous review and strengthening of communication with the customers. Some institutions have already introduced generative AI tools to further enhance their services and create a more natural flow as part of their communication with their members. The following sections describes in more detail different experiences of institutions leveraging AI tools that have enhanced capacity outlined in Figure 12.

Figure 12. *AI applications in social security*



Source: Authors' elaboration.

3.2. Service delivery and intelligent chatbots in social security

The Belgian initiative Chatbot Ori emerged to address the growing demand for customer service during the COVID-19 pandemic, which led to an increase in phone calls. Marc, the first version released in 2020, focused on tax form questions related to work interruptions. Ori 1.0 seamlessly integrated answers to frequently asked questions, improving visibility for the customer on the website. Ori 2.0, released in December 2021, improved query context retention, incorporated more topics, and made it easier to collaborate with administrators to align responses with the call centre. This example highlights the importance of constantly monitoring and optimizing algorithms helps the AI solution quickly adapt to changing user needs, even when starting with a narrow focus [23].

The Savings-Pensions Branch of the Deposit and Management Fund, a pension plan and solidarity fund management organization in Morocco, undertook a customer-centric digital transformation using AI. This initiative broke with traditional models by adopting a user-centered approach. The implementation included conversational chatbots, facial recognition for life monitoring services, customer service through WhatsApp and voice message-based call-bots. These digital assets, designed in collaboration with customers, improved the user experience, resulting in a 300 per cent increase in customers connecting to mobile applications, 10,000 users interacting with the WhatsApp service and 11,000 customers interacting with the chatbot. The adoption of AI-powered solutions had a significant impact on customer engagement [24].

The Superintendency of Occupational Risks (SRT) of Argentina successfully implemented an AI chatbot called Julieta. Designed to revolutionize customer service, it is available 24 hours a day, offering immediate responses to common queries. This has freed up phone lines for more complex issues. Julieta has an expanding knowledge base, and its results have shown a decrease in margins of error and a substantial number of user interactions [21].

Facing the challenges of the COVID-19 pandemic, Brazil's National Social Security Institute (INSS) implemented Helô, a virtual assistant chatbot powered by AI. Initially designed to respond to queries about the Meu INSS (My INSS) application, Helô expanded its functions to incorporate remote services, improving interaction with users and allowing more complex responses. The initial evaluation revealed significant success, with Helô handling more than one million calls in the first month and 57 per cent satisfaction. In this first phase, Helô operates using predefined rules and a database of frequent questions [21].

The Social Security Fund (CSS) of Panama during the pandemic developed a significant project focused on the use of AI and successfully launched ROVI, a virtual chatbot. The chatbot automated customer services and answers frequently asked questions. The strategy included digital consultations, efficient bed management and online payment services, with specific objectives such as achieving 100 per cent proficiency in the use of digital consultations and increased use of online payments [21].

The Social Security Bank (BPS) of Uruguay implemented an initiative for the inclusion of domestic workers in the social security system. The strategy focused on persuading both employers and workers to recognize that they had a formal employment relationship. Procedures for employers were centralized and automated, promoting self-management through online services, a mobile application and multi-channel assistance. A chatbot played a crucial role, answering 97 per cent of queries with a 100 per cent satisfaction rate. The results included a significant 24.4 per cent reduction in the evasion rate, with 57 per cent of employers registering online and 42 per cent making payments online [21, 23].

The Employees Provident Fund (EPF) in Malaysia successfully implemented the ELYA smart chatbot. The goal was to ease the burden on the contact centre by allowing customers to access information about EPF products and services themselves. ELYA evolved in three phases, from a basic chatbot to an advisory one for complex queries. Operating on the EPF website, powered by AI and natural language processing, ELYA offered real-time interaction, multilingual support and 24/7 availability. The chatbot has seen notable success, with millions of sessions, daily engagement and it contributed to financial education.

The German Federal Pension Insurance (DRV Bund) implemented the ZfA-Chatbot-PoC powered by artificial intelligence. The chatbot aimed to improve accessibility and communication with users. Achievements include improving the organization's image, introducing a new communication channel, increasing customer satisfaction and reducing employee workload [25].

The Social Insurance Institution (Kela) in Finland has positively implemented a multilingual chatbot, called Kela-Kelpo, to make it easier for customers to access benefit information. This chatbot works with Finnish, Swedish, and English. The solution uses conversational AI with natural language processing, offering a more human interaction. Initially, Kela had eight individual chatbots, but they were successfully consolidated into one, simplifying the user experience. Now, customers can easily switch between languages during a conversation. Additionally, the chatbot provides personalized advice in the self-service portal, based on frequently asked questions and contextual variables specific to each page [26].

3.3. Machine learning in social security

The past few years have seen a significant increase in machine learning solutions that strengthen the social security institution's capacity. This section analyses different machine learning solutions that have been built by social security institutions.

3.3.1. Improving health risk prevention and return to work in social security

Social security institutions focused on health and specialized occupational, safety and health (OSH) have already leveraged machine learning solutions to identify risks related to their members. These have created solutions that vary from enhancing an institution's capacity to detect illnesses to pattern detection to mitigate fraud risks when members file health claims.

As an example, the Mutual for Safety CChC in Chile successfully implemented an AI system to address the challenge of timely detection of pneumoconiosis, specifically silicosis, among workers. Silicosis is an incurable lung disease caused by inhaling crystalline silica dust, and the high volume of patients coupled with a limited number of certified doctors has led to delays and errors in the detection process. To tackle this issue, the organization developed an AI model capable of detecting respiratory pathologies through the analysis of radiological images. The system achieved an impressive accuracy rate of 99 per cent, allowing efficient differentiation between healthy individuals and those with pneumoconiosis. The project's overall objective facilitated the detection and treatment of occupational lung diseases, ensuring workers are promptly removed from exposure [27].

The German Social Accident Insurance Institution for the Construction Industry (BG BAU) has undertaken a pioneering role in leveraging AI to enhance accident prevention in the construction sector. In response to the challenge of limited resources and the high incidence of serious and fatal accidents, BG BAU initiated the lighthouse project "AI-based support for targeted accident prevention". The project involved training an AI application with around 10 million data points to predict workplaces with a high probability of

health and safety deficiencies leading to accidents. The AI system supports supervisors suggesting which companies require preventive measures, prioritized by an AI score indicating the urgency of advice [28].

3.3.2. Machine learning for improve health services and well-being

Artificial intelligence applications provide an overview of the significant opportunities to improve health services. The different applications, leveraging significant data sets, are able to categorize and identify individuals with potential illnesses and become preemptive in providing health services to patients.

The National Health Insurance Service (NHIS) in the Republic of Korea effectively utilized AI in its response to the COVID-19 crisis. The implementation focused on four main areas: information, finance, facilities, and human resources. NHIS leveraged big data to analyze medical information, categorize infected patients based on underlying diseases, and facilitate efficient patient treatment and resource management. This allowed for the classification of patients with mild symptoms to community treatment centres and those with severe symptoms to medical institutions, ensuring effective hospital bed capacity management. The AI-driven Enrolment Verification System helped health care institutions identify at-risk individuals in real-time, preventing infection spread within hospitals [29].

Similarly in Australia, the eHealth NSW developed and implemented a machine learning prototype product aimed at earlier sepsis detection in the emergency departments of New South Wales public health system. The AI-driven prototype leverages historical data extracted from four hospitals, spanning the years 2017–2019, and employs logistic regression and XGBoost (an open-source siftware library) algorithms for modeling. This helped the institution identify patients at risk of developing sepsis in the emergency department waiting room, the clinical decision support tool aims to enable early detection, leading to reduced sepsis-related deaths, ICU admissions, and readmissions [27].

A case with a slightly different objective is the Family Allowances Fund (COMPENSAR) in Colombia. They implemented the "well-being on demand" platform, integrating analytics, big data, and artificial intelligence to enhance user well-being based on Amartya Sen's theory. The platform implemented analysis techniques, big data, and AI to personalize well-being experiences for individual users, families and businesses. The platform utilizes a recommendation engine that analyses user behaviour and preferences, offering services aligned with their well-being needs. Key objectives include adapting to user needs, enhancing personalized experiences and supporting the Colombian business community in addressing the comprehensive well-being of their employees. The strategic plan involves analytical models, technology integration and human resource development, and aimed to reach 200,000 users by 2023, with 79 per cent actively engaged as of February 2023 [30].

Brazil's INSS has implemented an innovative artificial intelligence solution to improve decision making. Using a decision support system (DSS), the solution includes optical character recognition, document detection, text extraction and facial recognition. The system intelligently classifies requests, identifying cases of immediate denial or need for additional documentation, applying rules and diagnostic engines to automate decisions. It includes a citizen diagnostic dashboard, workflow management and a scenario engine. The objective is to accelerate decision-making, minimize delays and optimize the efficiency of the social security system [31].

3.3.3. Leveraging AI for ensuring proper use of resources and mitigating fraud

The Social Security Administering Body for the Health Sector (BPJS Kesehatan) in Indonesia implemented a machine-learning-based fraud detection system to address the increasing challenges of health care fraud. The AI machine learning-based solution allowed BPJS Kesehatan to shift from retrospective to concurrent fraud detection, where hospital claims are reviewed before payments are made, and the system was able to detect potential fraud more efficiently. The implementation provided a faster and more efficient fraud detection process, reduced review time and more accurate predictions with large datasets. Among the lessons learned was the importance of accurate historical claims data, having a data scientist team for ongoing support and the understanding that machine learning requires continuous improvement. Risks included the need for model modification when applying the system to different organizations and variations in metrics based on training data [32].

The National Social Security Fund (NSSF) of Uganda leveraged a machine learning model as part of their big data analytics and automation in business processes. Their implementation of predictive analytics models allowed them to forecast expected call volumes and determine resource requirements to better identify customers and data matching, transaction authentication, internal user patterns mapping that ultimately helped them in addressing fraud on pension contributions [33].

In 2022, Azerbaijan improved a benefit first implemented in 2006 for improving economic opportunities and reducing the financial burden on low-income households with the adoption of the Targeted State Social Assistance Strategy (TSSA). Initially, the TSSA application process was a manual process through a "one-stop shop," where the Centralized Electronic Information System (CEIS) allowed social workers to assess household conditions. However, aware of challenges such as corruption and operational deficiencies, the government of Azerbaijan implemented an AI-based application that transformed the TSSA allocation system. This innovation automated processes, reduced complaints and significantly improved transparency, efficiency in social security management [34].

In France, the URSSAF National Fund (URSSAF Caisse Nationale) has successfully implemented a machine learning project to improve the accuracy and efficiency of social security contribution collection forecasts. The organization has focused on optimizing financing and reducing costs through machine learning models. The project initially focused on the automated generation of daily forecasts for social security contributions in the private sector. The evaluation of results demonstrated that the machine learning algorithm outperforms the manual method in metrics such as mean absolute error (MAE), forecast quality indicator (PI) and absolute error standard deviation (AESD), thus achieving a substantial improvement in accuracy and efficiency [35].

In Estonia, the AI tool OTT was specifically developed to enhance decision making. The OTT decision support tool uses an advanced machine learning model that evaluates job seeker profiles, optimizing decisions on support channels, contact frequency, intervention strategies and risk levels. The implementation, backed by extensive research and collaboration with multiple organizations, included pilot testing in five offices before its final rollout in October 2020. The AI model, based on gradient boosting and backed by five years of unemployment data, provides valuable information about the likelihood of employment for job seekers. Ongoing support and user feedback were essential to perfecting the tool and ensuring constant optimization [31].

Employment and Social Development Canada (ESDC) generated a response to a policy that affected the Guaranteed Income Supplement (GIS), a benefit for low-income Old Age Security (OAS) beneficiaries. Some people, especially those with partners in long-term care facilities, were left vulnerable. The organization implemented an AI solution to quickly identify and support the analysis of the individuals with this supplement. The solution led to a machine learning model to process call centre notes and has identified more than 2,000 affected people. With an effectiveness rate of 92–98 per cent, the solution successfully identifies affected persons. The process highlighted the importance of data quality, collaboration and the need to address real problems for the successful implementation of AI [36, 37].

3.3.4. Leveraging AI to streamline claims management processes

Claims adjudication and management can significantly help social security institutions to streamline the case management and focus specialized (and many times scarce) resources to more complicated cases. This is especially important in managing health claims where the institutions rely on doctors from very different specialties to analyze health claims and make adjudication decisions based on their knowledge.

Austria's Federation of Social Insurances implemented an AI-based system that supports automatic processing of claims and matches them with doctors. An example is the case of medical bill reimbursement management using artificial intelligence. Prior to the implementation of the AI solution, long manual processing times resulted in months-long waits for refunds. With the platform, processing time was significantly reduced to just a few days, without the need for additional staff. The AI-based approach involved the collection and processing of documents using technologies such as optical character (OCR) and entity recognition, as well as open-source tools and languages, showing a commitment to efficiency and transparency [26].

openIMIS, an open source AI tool implemented through the German Society for International Cooperation (GIZ) collaborated with the Swiss Tropical and Public Health Institute (Swiss TPH) to integrate an AI-solution into the claims adjudication process. The solution has an innovative approach as it implements the use of three different machine learning algorithms including Extreme Gradient Boosting, Random Forest and Extra Trees to obtain more reliable results. This solution aims to reduce processing times, facilitate access to health care and positively contribute towards the goal of achieving universal health coverage [27, 38].

In addition to the different applications documented above, further examples on the potential use of artificial intelligence that can be applied to social security in tackling different challenges. Table 3 summarizes some cases of the use AI that can be further applied in of social security.

Table 3. *AI solutions for different social security challenges*

Social Security Challenge	AI Solution	Description
Old-age pension management	AI for pension automation	AI systems automate the management and disbursement of pensions, ensuring timely and accurate payments to retirees, reducing administrative costs, and improving efficiency [39]
Child care	AI monitoring and educational tools	Smart monitoring systems use AI to ensure child safety, while AI-based educational platforms provide personalized learning experiences and developmental tracking for children [40]
Fraud detection – credit	AI and machine learning in fraud detection	AI algorithms analyze transaction patterns in real-time to detect and prevent credit fraud, enhancing security for financial institutions and their customers [41]
Know Your Customer (KYC)	Automated KYC verification	AI-driven systems streamline the KYC process by verifying identities, documents, and data quickly and accurately, reducing the risk of fraud and improving customer onboarding [42]
Old-age support	Robots for assistance	Robotics powered by AI offer physical and cognitive assistance to the elderly and invalid, helping with daily tasks, medication reminders, and companionship, improving their quality of life [43]
Emotional support	AI chatbots for mental health	AI-powered chatbots provide emotional support and counseling services, offering an accessible and immediate resource for individuals seeking mental health assistance [44]

Source: Authors' elaboration.

4. Global uptake of AI and ongoing regulations

Social security institutions are part of the global trend of leveraging artificial intelligence to enhance their capacity and deliver better, improved and more efficient services.

4.1. Indices indicating AI capabilities of the countries

When examining the global landscape of national AI capabilities, there are three distinct indices that offer insights from various angles: the *Global AI Index 2023* by Tortoise Media [45]; the *Artificial Intelligence Index Report 2023* by the Stanford Human-Centered Artificial Intelligence (HAI) [46]; and the *Government AI Readiness Index 2023* by Oxford Insights [47]. Each index utilizes a unique array of indicators and sub-indicators to evaluate AI readiness and capabilities, reflecting the diverse impacts of AI on global competitiveness, innovation and governance.

The Global AI Index by Tortoise Media assesses countries across three primary dimensions: implementation, innovation, and investment. These dimensions encompass sub-indicators such as talent, infrastructure, and operating environment (under implementation); research and development (under innovation); and government strategy and commercial factors (under investment). Offering a comprehensive overview, this index identifies top-performing countries like China, Singapore and the United States (US), alongside bottom-ranking nations like Egypt and Uruguay, highlighting disparities in AI readiness across regions.

In contrast, the *Artificial Intelligence Index Report* by HAI Stanford takes a broader approach, considering indicators such as research and development, technical performance, and policy and governance. This report provides in-depth analysis without a specific ranking, offering insights into global AI trends and their ethical, economic and societal implications. Meanwhile, Oxford Insights' Government AI Readiness Index focuses on evaluating nations based on their governments' preparedness to implement AI technologies, emphasizing governance, infrastructure and policy aspects. Together, these indices offer a nuanced understanding of global AI capabilities, emphasizing the need for targeted policies and international collaboration to address disparities and ensure widespread benefits from AI advancements.

4.2. Institutional and regulatory challenges of embracing AI

The challenges faced by public institutions in implementing data-centric AI can be broadly categorized into three main areas: legal, regulatory and human resource availability.

Legal challenges stem from fundamental issues related to data and machine learning models, encompassing concerns such as ethics, data privacy, bias and discrimination, transparency and explainability, accountability and intellectual property. Many countries struggle to update their policies on these matters, making it increasingly difficult for legal institutions to keep pace with the rapidly advancing technologies.

Global regulatory initiatives for AI are complex, involving inclusive engagement, international collaboration for knowledge sharing and coordination among existing global bodies. The regulatory framework aims to identify best practices, manage risks and foster economic growth, while fostering consensus on AI standards across diverse stakeholders. This involves discussions on both non-binding and binding frameworks, as well as mechanisms for mutual reassurance, information exchange and liability.

In addition, regulatory efforts stress the importance of emergency response capabilities, mechanisms for oversight and adherence to ethical and governance principles throughout AI's lifecycle. Aligned with foundational United Nations (UN) principles and international commitments like the UN Charter and Human Declaration of Human Rights, AI governance prioritizes the promotion of human rights and sustainable development goals. Despite challenges posed by the global nature of AI governance and existing territorial frameworks, efforts are underway to address risks and ensure accountability. The envisioned regulatory framework involves a network of institutions tasked with assessing AI's future impacts, fostering innovation, managing risks, and promoting evidence-based policy assessment and accountability.

Despite the existence of prominent regulations like the European Union's (EU) General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) in the United States, along with various data protection acts in countries like Australia, Brazil, Canada and India, regulatory initiatives continue to evolve. These initiatives include ethical considerations surrounding AI use, compliance with

sector-specific regulations such as Health Insurance Portability and Accountability Act (HIPAA) for health care or Family Educational Rights and Privacy Act (FERPA) for educational institutions in the US, liability attribution for AI model decisions and procurement regulations like the Federal Acquisition Regulation (FAR) in the US. Public institutions must navigate these challenges to harness data for informed decision-making, service improvement and stakeholder satisfaction while ensuring compliance with regulations and safeguarding privacy and security.

The European Union has adopted the EU AI Act [48], a pioneering comprehensive legal framework for AI. This significant legislation is designed to oversee the progression, deployment and utilization of AI systems across the EU, with the dual objectives of addressing potential risks and encouraging innovation and ethical development. The act classifies AI systems according to their level of risk, imposing more stringent requirements on high-risk applications such as facial recognition and credit scoring. Through the establishment of clear guidelines and the promotion of responsible AI practices, the EU AI Act seeks to ensure that AI systems are developed and utilized in a manner that upholds fundamental rights, fosters fairness and cultivates societal trust.

In addition to the legal and regulatory challenges at the policy level, there are significant human resource challenges at the implementation level. Public agencies often operate with slower processes compared to private organizations, hindering their ability to adapt to technological advancements swiftly. To effectively leverage AI and data for public policy design, civil servants must possess a comprehensive understanding of AI's potential, risks and challenges. This necessitates specialized skills such as data science, machine learning, software development and policy engineering. Organizations must strategically hire, retain and upskill their staff to meet these demands. Furthermore, the implementation of AI may lead to job displacement, requiring organizations to develop reskilling and redeployment plans for affected employees. Additionally, organizations must plan for, and manage, the significant changes to processes, workflows and job roles that accompany AI implementation through effective change management strategies.

For successful digital transformation driven by AI and data, governments must adapt their functioning paradigms. This transformation begins with enhancing public employees' competencies to understand the implications of data-centric AI. Awareness of skill requirements across various organizational levels is crucial, necessitating an understanding of capacity-building needs at individual, team, departmental and governmental levels. Enhanced collaboration and communication between institutions facilitates the sharing of insights, while continuous monitoring of capacity-building initiatives ensures their ongoing effectiveness.

4.3. Open source implementation of AI

Unlike closed-source models, which are maintained by organizations and withhold their code from public access or audit, open-source models stand as free resources available for public use and modification, driving innovation forward. Examples of closed-source large language models (LLMs) include PaLM from Google, the family of GPT models from OpenAI and Claude from Anthropic. While third parties may utilize certain closed-source models through an Application Programming Interface (API), the core code remains inaccessible for manipulation.

In contrast, open-source models such as Falcon LLM by the Technology Innovation Institute and Llama2 from Meta are freely available for public use and modification, fostering innovation as mentioned by Stefano Maffulli's article "Open Source AI: Establishing a common ground" [49]. While concerns may

arise regarding potential security risks stemming from malicious alterations to open-source code, these vulnerabilities are akin to cybersecurity threats encountered by all software. The acceptance of both closed- and open-source development has been a longstanding practice, as evidenced by the widespread utilization of open-source software like Linux in cloud computing across national governments.

The adoption of open-source models can yield numerous benefits:

- **Enhanced creativity, innovation, and competition:** Open-source AI models have significantly reduced the time and resources required for application development, democratizing access to AI development and fostering competition among a broader array of developers beyond major tech companies.
- **Safer AI:** Publicly available models not only facilitate application development but also enable developers to enhance product safety. For instance, Seismograph, an open-source model, aids AI writing assistants in responsibly interacting with sensitive content.
- **Increased transparency:** The accessibility of datasets and codes in open-source models allows third-party auditing and verification, ensuring quality and reliability.

It's essential to underscore the significance of transparency in AI development. While President Biden's *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* addresses crucial aspects of responsible AI, such as safety, security and trustworthiness, it places less emphasis on transparency. Without transparency, it becomes challenging for individuals to assess the safety, security and trustworthiness of AI tools. While regulatory measures should explore various avenues for enhancing transparency, open-source models inherently embody transparency, serving as a vital beacon in the pursuit of responsible AI development.

The challenges of open source AI models stems from the fact that there is no single acceptable definition of AI. The latest draft of the Open Source AI Definition takes inspiration from the *GNU Manifesto* written by Richard Stallman in 1985 to promote the development of the GNU operating system. The manifesto's simple slogan, "If I like a programme, I must share it with other people who like it," is a cornerstone of the Free Software Movement, emphasizing the principles of software freedom. While the GNU Manifesto refers to "programmes" without needing to define them, defining AI systems poses a challenge due to differing interpretations in the policy world.

4.4. Need for human capital for proper AI implementation

While AI holds immense promise, its success hinges on a strong foundation: robust human capital. By investing in people and their knowledge, skills and abilities, countries can build workforces equipped to harness the power of AI effectively. This requires a redefined educational system, one that equips individuals with the necessary skills in data analysis, machine learning and AI development. Additionally, fostering a culture of continuous learning allows individuals to adapt to the rapidly changing AI landscape.

Countries that prioritize education and human capital development are best positioned to reap the transformative benefits of AI. This not only fosters innovation, economic growth and societal progress, but also mitigates potential disruptions and inequalities caused by rapid technological advancements. In the global arena, a well-educated and skilled workforce is the cornerstone of AI-driven prosperity and competitiveness. Remember, human ingenuity and expertise are the key to unlocking the full potential of AI and navigating its complexities.

5. Assessing AI readiness

Artificial intelligence holds the potential to transform various sectors by enhancing efficiency, accuracy and accessibility. However, it also presents ethical, privacy and security concerns. Recognizing these dual aspects, evaluating AI readiness becomes imperative for countries, particularly for social security institutions, to navigate the complexities of AI adoption responsibly and ethically.

The primary criteria for assessing AI readiness include evaluating the digital infrastructure, which serves as the foundation for AI implementation. This encompasses robust hardware, software and high-quality data flows free from biases. Additionally, the presence of ethical and legal frameworks governing AI use is crucial to address privacy, fairness and non-discrimination concerns.

Assessing a country's readiness for AI also involves evaluating its technical capabilities and the skill level of its workforce. This includes not only engineers and data scientists but also policymakers and public servants knowledgeable about AI and its implications. Furthermore, assessing the government's role in facilitating and using AI in the public sector, particularly in social security institutions, is essential. This entails evaluating how AI can enhance public service delivery and ensuring equitable benefits for all citizens.

Finally, ensuring inclusivity and accessibility of AI technologies to all segments of the population is paramount. This requires addressing digital divides and ensuring that vulnerable groups are not marginalized in the AI-driven transformation.

5.1. International organizations and AI readiness

The United Nations Development Programme (UNDP), in collaboration with partners like the International Telecommunication Union (ITU) and United Nations Educational, Scientific and Cultural Organization (UNESCO), emphasizes the importance of a comprehensive AI Readiness Assessment. Such assessments help countries gauge their preparedness across the aforementioned criteria, focusing on ethical considerations, digital infrastructure and capacity building. Through initiatives like the Joint Facility with ITU, UNDP supports countries in developing data governance frameworks and enhancing digital capacities, highlighting the need for a proactive governance approach to AI.

5.2. AI readiness and social security institutions

Social security institutions stand to benefit significantly from AI through improved service delivery, fraud detection and the formulation of support policies. However, they also face challenges related to data management, data privacy, ethical use of AI and ensuring equitable access to services. A readiness assessment helps these institutions to:

- Identify gaps in digital infrastructure, data management and capacity that could hinder the effective implementation, rollout and use of AI.
- Develop and implement ethical guidelines and legal frameworks that safeguard citizens' rights and data.
- Ensure that AI-driven solutions are designed with inclusivity and accessibility at their core, so that no group is disadvantaged.

UNPD takes a specific approach as part of their proposed AI Readiness Assessment [50] providing a structured approach for countries to prepare for the responsible and ethical adoption of AI technologies. By focusing on digital infrastructure, ethical and legal frameworks, capacity building, and inclusivity, countries can ensure that AI serves as a tool for societal advancement rather than a source of disparity. It is therefore important for social security institutions to develop guidelines to help assess and identify a roadmap to leverage AI for enhancing social security while safeguarding ethical standards and promoting equitable access to services.

6. Institutional challenges of implementing data-centric AI

The institutional challenges of implementing data-centric artificial intelligence in public institutions can be divided into three categories: legal, regulatory, and the availability of human resources. The legal challenges arise from the inherent issues of data and machine learning modeling such as ethics, data privacy, bias and discrimination, transparency and explainability, accountability and intellectual property. Many countries have not yet succeeded in reformulating their policies on these issues, and it is becoming extremely hard for legal institutions to keep up with the rapid pace of technological transformation.

Public institutions need to look for ways to address the regulatory ecosystem described in section 4.2 above so they can leverage data to make better decisions, improve services and better serve their stakeholders, ensuring compliance with regulations while protecting privacy and security. Apart from the legal and regulatory challenges at the policy level, there are several human resource challenges at the implementation level. Public agencies that are slow to respond and adopt technological change, must understand AI and data and be able to tap into their potential. Organizations need to know the opportunities offered by AI while being fully aware of its risks and challenges. Working in AI and data requires specialized skills, such as data scientists, machine learning engineers, software developers, and engineering policy specialists. Organizations need to plan on hiring and retaining these skilled professionals. Investing in training and upskilling the existing workforce can help overcome these challenges and help them in their implemention of AI while addressing concerns over the substitution of some jobs that can be automated. Since implementing AI can involve significant changes, a change management plan will help them and their employees navigate these changes effectively.

For digital transformation based on AI and data to succeed, governments and social security institutions need their staff at different levels to acquire the competencies required to understand the transformations that data-centric AI brings. Therefore, it is important to raise awareness of the required skills at different levels. This can be done by understanding the capacity-building needs at the individual, team, department and government level. Increased collaboration and communication between institutions would help these departments share insights. Most significantly, there should be continuous monitoring of the impact of capacity-building initiatives.

7. Conclusion

Social security institutions have already made significant strides in implementing artificial intelligence as part of their core business processes to improve internal operations as well as service delivery. AI has already shown how it can significantly increase capacity in different dimensions. The landscape of implementation of AI solutions continues to evolve and has accelerated over the last few years, and the use and business cases for leveraging AI show that its use can significantly improve social outcomes.

AI thrives on data which a key component and fundamental pillar in its adoption. AI's integration into public service operations offers a pathway to heightened administrative efficiency, leveraging diverse datasets to automate tasks and inform decision-making processes. Leveraging AI in the automation of services and processes can significantly improve efficiency as well as enhance and increase the capacity of organizations to provide better and more specialized attention to specific groups.

As AI becomes increasingly indispensable, challenges persist. Its reliance on data requires a continuous stream of high-quality information to avert unwanted biased or erroneous outcomes. Key obstacles include ensuring data availability and quality, vital for effectively training AI systems. Establishing robust data governance protocols becomes essential, encompassing internal data utilization and compliance with data protection regulations while considering external data sources [4].

Despite AI's capacity to accommodate datasets of varying sizes, its evolving nature warrants careful scrutiny prior to real-world deployment, particularly regarding inherent limitations and risks. Methodological disparities between AI and traditional software development pose significant challenges, notably in ensuring transparency and explainability, crucial for accountable decision-making. It is therefore important for institutions to identify where and how AI adds value to the services and benefits provided by social security institutions, and understand how these improvements cannot be achieved without implementing AI.

Supervision of AI solutions is crucial in social security. It ensures that the output, performance and accuracy of results are in line with the institution's objectives. Monitoring the behaviour and output of AI solutions is needed to mitigate risks related to data and algorithms, helping institutions make necessary adjustments and corrections to ensure it meets its needs. Additionally, continuous oversight allows for the iterative improvement of AI models, adapting them to new data or changing circumstances, thus maintaining their relevance and effectiveness over time. It also mitigates unintended consequences that could undermine trust and utility in the advancements of AI tools.

Institutions should integrate AI into their technology portfolio to be used together with other technologies better suited to specific situations. As the field of AI continues to develop it is important for social security institutions to stay updated on the solutions and innovations around the use of AI. As it continues to evolve, social security should continue experimenting with the use of AI solutions, to help better understand the implications of data, infrastructure and strategies to tackle challenges and limitations of the AI tools. This way, institutions can be better prepared and strategically better positioned to fully engage when new AI solutions that can be fully leveraged by institutions for their specific needs. This implies assessing the expertise of their personnel and prioritize capacity building to facilitate a seamless transition toward a data-centric digital landscape. Similarly, it also means adopting new guidelines that can help social security institutions ensure that capacity is developed and that it considers regulatory, privacy and data protection regulation.

A glance over the use of AI in social security shows that capacity needs to be built to take full advantage of the growing number of AI tools and techniques. The need for data quality in the responsible application of AI will lead to a future where today's value-added services and data scientist can become indispensable for the future of social security.

References

[1] **ISSA.** 2019. *Applying emerging technologies in social security – Summary report 2017–2019*. Geneva, International Social Security Association – Technical Commission on Information and Communication Technology.

[2] **ISSA.** 2022. *Data-driven innovation in social security: Good practices from Asia and the Pacific* (Analysis). Geneva, International Social Security Association.

[3] **ISSA.** 2020. *Artificial intelligence in social security: Background and experiences* (Analysis). Geneva, International Social Security Association.

[4] **ISSA.** 2022. *ISSA Guidelines on Information and Communication Technology*. Geneva, International Social Security Association.

[5] **Alzubaidi, L. et al.** 2021. "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions", in *Journal of Big Data*, Vol. 8.

[6] **Merritt, R.** 2023. "*What are foundation models?*", in *NVIDIA Blog*, 13 March.

[7] **Wiesalla, L.** 2023. "*Benchmark of sustainable machine learning algorithms*", in *NextLytics Blog,* 14 February.

[8] DSTL. 2020. *Annual report and accounts 2020 to 2021*. London, Defence Science and Technology Laboratory.

[9] **Greene, K. T.;Park, B.; Colaresi, M.** 2019. "Machine learning human rights and wrongs: How the successes and failures of supervised learning algorithms can inform the debate about information effects", in *Political Analysis*, Vol. 27, No. 2.

[10] **Sirotkin, K.; Carballeira, P.; Escudero-Viñolo, M.** 2022. "A study on the distribution of social biases in self-supervised learning visual models", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* New York, NY.

[11] **Navarro, C. L. A. et al.** 2021. "Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review", in *BMJ,* Vol. 375, No. 2281.

[12] **Watson, D. S.** 2023. "On the philosophy of unsupervised learning", in *Philosophy & Technology*, Vol. 36, No. 2.

[13] **Sansone, D.; Zhu, A.** 2023. "Using machine learning to create an early warning system for welfare recipients", in *Oxford Bulletin of Economics and Statistics*, Vol. 85, No. 5.

[14] **Oprescu, A. M. et al.** 2022. "Towards a data collection methodology for responsible artificial intelligence in health: A prospective and qualitative study in pregnancy", in *Information Fusion*, No. 83.

[15] **Inel, O.; Draws, T.; Aroyo, L.** 2023. "Collect, measure, repeat: Reliability factors for responsible AI data collection", in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 11, No. 1.

[16] **Pushkarna, M.; Zaldivar, A.; Kjartansson, O.** 2022. "Data cards: Purposeful and transparent dataset documentation for responsible ai", in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, Association for Computing Machinery.

[17] **Stoyanovich, J.; Howe, B.; Jagadish, H. V.** 2020. "Responsible data management", in *Proceedings of the VLDB Endowment*, Vol. 13, No. 12.

[18] **Liu, H. et al.** 2021. "FAIXID: a framework for enhancing ai explainability of intrusion detection results using data cleaning techniques", in *Journal of Network and Systems Management*, Vol. 29, No. 4.

[19] **Došilović, F. K.; Brčić, M.; Hlupić, N.** 2018. "Explainable artificial intelligence: A survey", in *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. New York, NY, IEEE.

[20] **Tjoa, E.; Guan, C.** 2020. "A survey on explainable artificial intelligence (xai): Toward medical xai", in *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 11.

[21] **ISSA.** 2021. *The application of chatbots in social security: Experiences from Latin America* (Analysis). Geneva, International Social Security Association.

[22] **ISSA.** 2022. *Artificial intelligence in social security institutions: The case of intelligent chatbots* (Analysis). Geneva, International Social Security Association.

[23] **ISSA.** 2023. *Artificial intelligence: Deconstructing chatbots* (ISSA Webinar, 20 April). Geneva, International Social Security Association.

[24] **ISSA.** 2022. *Customer-centric application development* (ISSA Webinar, 14 September). Geneva, International Social Security Association.

[25] **ISSA.** 2021. *Improving customer services with intelligent chatbots* (ISSA Webinar, 8 December). Geneva, International Social Security Association.

[26] **ISSA.** 2022. *Digital transformation for adaptable and people centric social security* (16th ISSA International Conference on Information and Communication Technology in Social Security, Tallinn, 4–6 May). Geneva, International Social Security Association.

[27] **ISSA.** 2023. *Artificial intelligence applications in health care* (ISSA Webinar, 31 January). Geneva, International Social Security Association.

[28] **German Social Accident Insurance.** 2023. *Improving occupational health and safety in the construction sector thanks to artificial intelligence* (Good Practice). Geneva, International Social Security Association.

[29] **National Health Insurance Service.** 2020. *National Health Insurance Service big data and ICT usage in COVID-19 crisis* (Good Practice). Geneva, International Social Security Association.

[30] **COMPENSAR.** 2023. *Well-being on demand: A comprehensive well-being solution* (Good Practice). Geneva, International Social Security Association.

[31] **ISSA.** 2023. *Artificial intelligence for enhanced decision-making* (ISSA Webinar, 29 June). Geneva, International Social Security Association.

[32] **BPJS Kesehatan.** 2021. *Machine learning and big data: Supporting decision making on fraud detection* (Good Practice). Geneva, International Social Security Association.

[33] **National Social Security Fund.** 2020. *Application of machine learning to improve data quality* (Good Practice). Geneva, International Social Security Association.

[34] **State Social Protection Fund.** 2022. *An artificial-intelligence-based solution to the examination of household conditions for the Targeted State Social Assistance assignment in Azerbaijan* (Good Practice). Geneva, International Social Security Association.

[35] **URSSAF**. 2024. *Cash flow: Machine learning forecast of daily private sector social security contributions – A good practice in compliance management and financial efficiency* (Good Practice). Geneva, International Social Security Association.

[36] **ISSA.** 2020. *Artificial intelligence for social security institutions* (ISSA European Network Webinar, 22 September). Geneva, International Social Security Association.

[37] **Employment and Social Development Canada.** 2020. *Using artificial intelligence (AI) to identify vulnerable Canadians* (Good Practice). Geneva, International Social Security Association.

[38] **openIMIS.** 2024. *openIMIS home*. Paris, UNESCO.

[39] **Moreeng-Mogotsi, M. B.** 2019. *Investigating factors affecting efficiency of pension administration system in North-West province* (Diss.). Potchefstroom, North-West University.

[40] **Devi, J. S., et al.** 2022. "A path towards child-centric artificial intelligence based education", in *International Journal of Early Childhood*, Vol. 14, No. 3.

[41] **Hassan, M.; Abdel-Rahman Aziz, L.; Andriansyah, Y.** 2023. "The role artificial intelligence in modern banking: An exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance", in *Reviews of Contemporary Business Analytics*, Vol. 6, No. 1.

[42] **Gupta, A.; Dwivedi, D. N.; Shah, J.** 2023. *Artificial intelligence applications in banking and financial services: Anti money laundering and compliance*. Berlin, Springer Nature.

[43] **Cortellessa, G. et al.** 2021."AI and robotics to help older adults: Revisiting projects in search of lessons learned", in *Paladyn, Journal of Behavioral Robotics*, Vol. 12, No. 1.

[44] **Khawaja, Z.; Bélisle-Pipon, J.-C.** 2023. "Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots", in *Frontiers in Digital Health*, Vol. 5.

[45] **Cesareo, S.; White, J.** 2023. *The Global AI Index*. London, Tortoise Media.

[46] **AI Index Steering Committee.** 2023. *The AI Index 2023 Annual Report*. Stanford, CA, Stanford University – Institute for Human-Centered AI.

[47] **Hankins, E. et al.** 2023. "Government AI Readiness Index 2023", in *Oxford Insights*.

[48] **European Commission.** 2024. *Artificial Intelligence Act*. Brussels.

[49] **Maffulli, S.** 2023. "*Open Source AI: Establishing a common ground*", in *OpenSource Opinions*, 28 November.

[50] **Hamdar, Y.; Ngodup Massally, K.; Peiris, H.** 2023. "*Are countries ready for AI? How they can ensure ethical and responsible adoption*", in *UNDP Blog*, 23 April.

[51] **Gillis, M. et al.** 2021. "A simulation–optimization framework for optimizing response strategies to epidemics", in *Operations Research Perspectives*, Vol. 8.

[52] **Nearing, G. et al.** 2023. "AI increases global access to reliable flood forecasts", in *arXiv preprint arXiv:2307*, 16104.

[53] **Kankanamge, N. et al.** 2020. "Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets", in *International Journal of Disaster Risk Reduction*, Vol. 42.

[54] **Nicholls, R. J. et al.** 2021. "Integrating new sea-level scenarios into coastal risk and adaptation assessments: An ongoing process", in *Wiley Interdisciplinary Reviews: Climate Change*, Vol. 12, No. 3.

[55] **Yaqoob, N. et al.** 2023. "The effects of agriculture productivity, land intensification, on sustainable economic growth: A panel analysis from Bangladesh, India, and Pakistan Economies", in *Environmental Science and Pollution Research,* Vol. 30, No. 55.

[56] **Castañeda, P. et al.** 2021. "Saving for the future: Evaluating the sustainability and design of Pension Reserve Funds", in *Pacific-Basin Finance Journal,* Vol. 68.

[57] **Wang, C.; Zhu, H.** 2020. "Representing fine-grained co-occurrences for behavior-based fraud detection in online payment services", in *IEEE Transactions on Dependable and Secure Computing*, Vol. 19, No. 1.

[58] **KM, A. K.; Abawajy, J.** 2022. "Detection of false income level claims using machine learning", in *International Journal of Modern Education & Computer Science*, Vol. 14, No. 1.

[59] **Bui, D. T.** 2021. *Applications of machine learning in eKYC's identity document recognition* (Diss). Kouvola, Sout-Eastern Finland University of Applied Sciences.

[60] **Franklin, T. W.; Henry, T. K. S.** 2020. "Racial disparities in federal sentencing outcomes: Clarifying the role of criminal history", in *Crime & Delinquency,* Vol. 66, No. 1.

[61] **Amrutha, C. V.; Jyotsna, C.; Amudha, J.** 2020. "Deep learning approach for suspicious activity detection from surveillance video", in *2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*. New York, NY, IEEE.

[62] **Al-Hashedi, K. G.; Magalingam, P.** 2021. "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019", in *Computer Science Review*, Vol. 40.

[63] **Tavoschi, L. et al.** 2020. "Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from September 2016 to August 2017 in Italy", in *Human Vaccines & Immunotherapeutics*, Vol. 16, No. 5.

[64] **Hasan, I. et al.** 2021. "The AI enabled chatbot framework for intelligent citizen-government interaction for delivery of services", in *8th International Conference on Computing for Sustainable Global Development (INDIACom)*. New York, NY, IEEE.

[65] **Weber, R. B.; Grobe, D.; Scott, E. K.** 2018. "Predictors of low-income parent child care selections",

in *Children and Youth Services Review*, Vol. 88, May.

[66] **Mertoğlu, U.; Genç, B.** 2020. "Automated fake news detection in the age of digital libraries", in *Information Technology and Libraries*, Vol. 39, No. 4.

[67] **Jiawei, L. et al.** 2020. "Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study", in *JMIR Public Health and Surveillance*, Vol. 6, No. 2.

[68] **Hughes, G.; Shaw, S. E.; Greenhalgh, T.** 2020. "Rethinking integrated care: a systematic hermeneutic review of the literature on integrated care strategies and concepts", in *The Milbank Quarterly*, Vol. 98, No. 2.

[69] **Norman, K.; Haß, U.; Pirlich, M.** 2021. "Malnutrition in older adults: Recent advances and remaining challenges", in *Nutrients*, Vol. 13, No. 8.

[70] **Wright, K.; Singh, S.** 2022. "Reducing falls in dementia inpatients using vision-based technology", in *Journal of Patient Safety*, Vol. 18, No. 3.

[71] **Sezgin, E. et al.** 2020. "A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors", in *Translational Behavioral Medicine*, Vol. 10, No. 3.

[72] **Zhang, T. et al.** 2022. "Natural language processing applied to mental illness detection: A narrative review", in *NPJ Digital Medicine*, Vol. 5, No. 1.

[73] **Jyväkorpi, S. K. et al.** 2021. "The sarcopenia and physical frailty in older people: multi-component treatment strategies (SPRINTT) project: Description and feasibility of a nutrition intervention in community-dwelling older Europeans", in *European Geriatric Medicine,* Vol. 12, No. 2.

Annex

Table A.1. *ML domain and category that could be applied in social protection*

Domain	ML Category	Specific Capabilities	Use case/Scenario
Crisis Response	ML for simulation modeling	Predicting outcomes of crises, optimizing response strategies	A simulation-optimization framework for optimizing response strategies to epidemics [51]
			Global access to reliable flood forecasts [52]
	ML for unstructured and semi-structured data	Analysing social media for real-time disaster monitoring	Determining disaster severity through social media analysis [53]
	ML for simulation modeling	Climate change impact predictions	Simulating sea level rise scenarios and their impacts on coastal areas [54]
Economic Empowerment	ML for simulation modeling	Modeling economic growth scenarios under different policies	The effects of agriculture productivity, land intensification, on sustainable economic growth [55]
		Economic modeling for policy impact analysis	Saving for the future: Evaluating the sustainability and design of Pension Reserve Funds [56]
		Simulating behavioral patterns to predict fraudulent activities	Representing fine-grained co-occurrences for behavior-based fraud detection in online payment services [57]
	ML for multi-modal applications	Integrating textual, audio, and visual data for comprehensive analysis	Detection of false income level claims using machine learning [58]
		Combining textual, visual, and possibly audio data for comprehensive identity verification	Applications of machine learning in eKYC's identity document recognition [59]
Security and Justice	ML for language processing	Analysing legal documents for fairness	Identifying bias in historical sentencing data [60]
	ML for vision	Enhancing surveillance for harm prevention	Real-time analysis of CCTV footage to detect suspicious activities [61]
		Analysing patterns in visual data for anomaly detection	Use image analysis to detect forged documents in financial applications [62]

Public and Social Sector	ML for unstructured and semi-structured data	analysing public feedback for policy making	Sentiment analysis of social media to gauge public opinion on policies [64]
	ML for multi-modal applications	Enhancing public service delivery	Interactive chatbots for citizen services [64]
	ML for simulation modeling	Forecasting demand for child care services	Model and predict future child care needs to aid in planning and resource allocation [65]
Information Verification and Validation	ML for language processing	Detecting and flagging false news	Automated systems to verify news authenticity before spreading [66]
Health and Hunger	ML for unstructured and semi-structured data	Predicting disease outbreaks	Analysing health reports and social media for early warning signs [67]
	ML for multi-modal applications	Personalized treatment and care plans	Integrating patient data across sources for holistic care plans [68]
	ML for vision	Detecting signs of malnutrition or deterioration	Detect malnutrition signs and activity changes in elderly [69]
		Monitoring health and safety, recognizing faces and objects to assist in daily tasks	Use vision to detect falls or accidents in the home and alert caregivers or emergency services [70]
	ML for language processing	- Analysing patient health records - Voice-based dietary tracking	Provide medication and dietary reminders via voice [71]
		Understanding spoken language to provide conversational support and detect emotional cues	Engage in natural language conversations to provide social interaction and detect signs of emotional distress [72]
	ML for simulation modeling	Predicting dietary impact on health	Project health impacts of dietary habits in elderly [73]

ISSA General Secretariat
Route des Morillons 4, Case postale 1
CH-1211 Geneva 22, Switzerland

E: **ISSA@ilo.org**
T: **+41 22 799 66 17**
F: **+41 22 799 85 09**

www.issa.int

International Social Security Association

The International Social Security Association (ISSA) is the world's leading international organization for social security institutions, government departments and agencies. The ISSA promotes excellence in social security through professional guidelines, expert knowledge, services and support to enable its members to develop dynamic social security systems and policy throughout the world. Founded in 1927 under the auspices of the International Labour Organization, the ISSA is based in Geneva, Switzerland.