EUROPEAN UNION AGENCY
FOR FUNDAMENTAL RIGHTS

GETTING THE FUTURE RIGHT

ARTIFICIAL INTELLIGENCE AND FUNDAMENTAL RIGHTS

REPORT

Neither the European Union Agency for Fundamental Rights nor any person acting on behalf of the Agency is responsible for the use that might be made of the following information.

Luxembourg: Publications Office of the European Union, 2020

Print	ISBN 978-92-9474-861-4	doi:10.2811/58563	TK-03-20-119-EN-C
PDF	ISBN 978-92-9474-860-7	doi:10.2811/774118	TK-03-20-119-EN-N

© Photo credits:

Cover: HQUALITY/Adobe Stock
Page 5: Mimi Potter/Adobe Stock
Page 8: monsitj/Adobe Stock
Page 14: Monsitj/Adobe Stock
Page 16: Mykola Mazuryk/Adobe Stock
Page 20: metamorworks/Adobe Stock
Page 25: Gorodenkoff/Adobe Stock
Page 28: Dimco/Adobe Stock
Page 32: VideoFlow/Adobe Stock
Page 37: zapp2photo/Adobe Stock
Page 41: bestforbest/Adobe Stock
Page 44: zapp2photo/Adobe Stock
Page 47: European Communities
Page 52: blacksalmon/Adobe Stock
Page 61: Copyright © 2020 CODED BIAS - All Rights Reserved
Page 63: Siberian Art/Adobe Stock
Page 68: Good Studio/Adobe Stock
Page 75: Sikov/Adobe Stock
Page 79: robsonphoto/Adobe Stock
Page 82: thodonal/Adobe Stock
Page 86: blackboard/Adobe Stock
Page 88: blackboard/Adobe Stock
Page 92: Monopoly919/Adobe Stock
Page 95: Gorodenkoff/Adobe Stock
Page 96: Freedomz/Adobe Stock
Page 100: Copyright © 2020 CODED BIAS - All Rights Reserved
Page 103: Copyright © 2020 CODED BIAS - All Rights Reserved

Foreword

Did you know that artificial intelligence already plays a role in deciding what unemployment benefits someone gets, where a burglary is likely to take place, whether someone is at risk of cancer, or who sees that catchy advertisement for low mortgage rates?

We speak of artificial intelligence (AI) when machines do the kind of things that only people used to be able to do. Today, AI is more present in our lives than we realise – and its use keeps growing. The possibilities seem endless. But how can we fully uphold fundamental rights standards when using AI?

This report presents concrete examples of how companies and public administrations in the EU are using, or trying to use, AI. It discusses the potential implications for fundamental rights and shows whether and how those using AI are taking rights into account.

FRA interviewed just over a hundred public administration officials, private company staff, as well as diverse experts – including from supervisory and oversight authorities, non-governmental organisations and lawyers – who variously work in the AI field.

Based on these interviews, the report analyses how fundamental rights are taken into consideration when using or developing AI applications. It focuses on four core areas – social benefits, predictive policing, health services and targeted advertising. The AI uses differ in terms of how complex they are, how much automation is involved, their potential impact on people, and how widely they are being applied.

The findings underscore that a lot of work lies ahead – for everyone.

One way to foster rights protection is to ensure that people can seek remedies when something goes awry. To do so, they need to know that AI is being used. It also means that organisations using AI need to be able to explain their AI systems and how they deliver decisions based on them.

Yet the systems at issue can be truly complex. Both those using AI systems, and those responsible for regulating their use, acknowledge that they do not always fully understand them. Hiring staff with technical expertise is key.

Awareness of potential rights implications is also lacking. Most know that data protection can be a concern, and some refer to non-discrimination. They are less aware that other rights – such as human dignity, access to justice and consumer protection, among others – can also be at risk. Not surprisingly, when developers review the potential impact of AI systems, they tend to focus on technical aspects.

To tackle these challenges, let's encourage those working on human rights protection and those working on AI to cooperate and share much-needed knowledge – about tech and about rights.

Those who develop and use AI also need to have the right tools to assess comprehensively its fundamental rights implications, many of which may not be immediately obvious. Accessible fundamental rights impact assessments can encourage such reflection and help ensure that AI uses comply with legal standards.

The interviews suggest that AI use in the EU, while growing, is still in its infancy. But technology moves quicker than the law. We need to seize the chance now to ensure that the future EU regulatory framework for AI is firmly grounded in respect for human and fundamental rights.

We hope the empirical evidence and analysis presented in this report spurs policymakers to embrace that challenge.

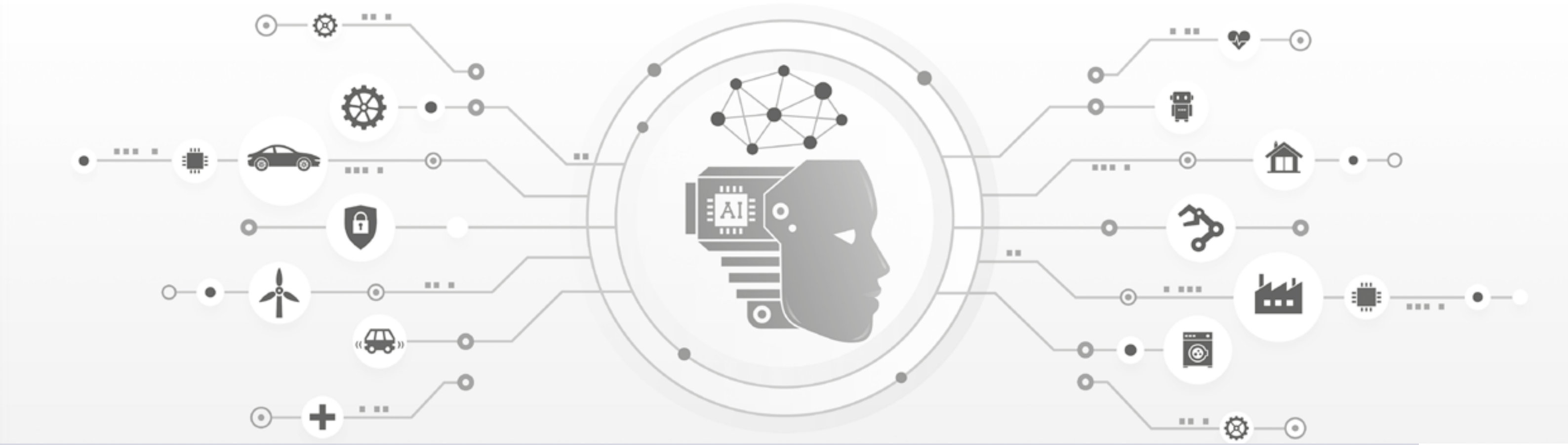
Michael O'Flaherty
Director

Contents

Figures

Key findings and FRA opinions

New technologies have profoundly changed how we organise and live our lives. In particular, new data-driven technologies have spurred the development of artificial intelligence (AI), including increased automation of tasks usually carried out by humans. The COVID-19 health crisis has boosted AI adoption and data sharing – creating new opportunities, but also challenges and threats to human and fundamental rights.



Developments in AI have received wide attention by the media, civil society, academia, human rights bodies and policymakers. Much of that attention focuses on its potential to support economic growth. How different technologies can affect fundamental rights has received less attention. To date, we do not yet have a large body of empirical evidence about the wide range of rights AI implicates, or about the safeguards needed to ensure that the use of AI complies with fundamental rights in practice.

On 19 February 2020, the European Commission published a White Paper on Artificial Intelligence – *A European approach to excellence and trust*. It outlines the main principles of a future EU regulatory framework for AI in Europe. The White Paper notes that it is vital that such a framework is grounded in the EU's fundamental values, including respect for human rights – Article 2 of the Treaty on European Union (TEU).

This report supports that goal by analysing fundamental rights implications when using artificial intelligence. Based on concrete 'use cases' of AI in selected areas, it focuses on the situation on the ground in terms of fundamental rights challenges and opportunities when using AI.

Legal framework

The overarching fundamental rights framework* that applies to the use of AI in the EU consists of the Charter of Fundamental Rights of the EU (the Charter) as well as the European Convention on Human Rights.

Multiple other Council of Europe and international human rights instruments are relevant. These include the 1948 Universal Declaration of Human Rights and the major UN human rights conventions.**

In addition, sector-specific secondary EU law, notably the EU data protection *acquis* and EU non-discrimination legislation, helps safeguard fundamental rights in the context of AI. Finally, the national laws of EU Member States also apply.

* *For more, see FRA (2012),* ***Bringing rights to life: The fundamental rights landscape of the European Union****, Luxembourg, Publications Office of the European Union.*

** *These major conventions include: the 1966 International Covenant on Civil and Political Rights; the 1966 International Covenant on Economic, Social and Cultural Rights; the 1965 International Convention on the Elimination of All Forms of Racial Discrimination; the 1979 Convention on the Elimination of All Forms of Discrimination against Women; the 1984 Convention against Torture; the 1989 Convention on the Rights of the Child; the 2006 Convention on the Rights of Persons with Disabilities; and the 2006 International Convention for the Protection of All Persons from Enforced Disappearance.*

For more on the universal international human rights law framework, including their enforcement mechanisms, see e.g. De Schutter, O. (2015), International Human Rights Law: Cases, Materials, Commentary*, Cambridge, Cambridge University Press, 2nd edition.*

The report is based on 91 interviews with officials in public administration and staff in private companies, in selected EU Member States. They were asked about their use of AI, their awareness of fundamental rights issues involved, and practices in terms of assessing and mitigating risks linked to the use of AI.

Moreover, 10 interviews were conducted with experts who deal, in various ways, with the potential fundamental rights challenges of AI. This group included public bodies (such as supervisory and oversight authorities), non-governmental organisations and lawyers.

SAFEGUARDING FUNDAMENTAL RIGHTS – SCOPE, IMPACT ASSESSMENTS AND ACCOUNTABILITY

Considering the full scope of fundamental rights with respect to AI

Using AI systems engages a wide range of fundamental rights, regardless of the field of application. These include – but also go beyond – privacy, data protection, non-discrimination and access to justice.

The EU Charter of Fundamental Rights (the Charter) became legally binding in December 2009 and has the same legal value as the EU treaties. It brings together civil, political, economic and social rights in a single text. Pursuant to Article 51 (1) of the Charter, the institutions, bodies, offices and agencies of the Union have to respect all the rights as embodied in the Charter. EU Member States have to do so when they are implementing Union law. This applies equally to AI as to any other field.

The fieldwork of this research shows that a large variety of systems are used under the heading of AI. The technologies analysed entail different levels of automation and complexity. They also vary in terms of the scale and potential impact on people.

FRA's findings show that using AI systems implicate a wide spectrum of fundamental rights, regardless of the field of application. These include, but also go beyond, privacy and data protection, non-discrimination and access to justice. Yet, when addressing the impact of AI with respect to fundamental rights, the interviews show, the scope is often delimited to specific rights.

A wider range of rights need to be considered when using AI, depending on the technology and area of use. In addition to rights concerning privacy and data protection, equality and non-discrimination, and access to justice, other rights could be considered. These include, for example, human dignity, the right to social security and social assistance, the right to good administration (mostly relevant for the public sector) and consumer protection (particularly important for businesses). Depending on the context of the AI use, any other right protected in the Charter needs consideration.

FRA OPINION 1

When introducing new policies and adopting new legislation on AI, the EU legislator and the Member States, acting within the scope of EU law, must ensure that respect for the full spectrum of fundamental rights, as enshrined in the Charter and the EU Treaties, is taken into account. Specific fundamental rights safeguards need to accompany relevant policies and laws.

In doing so, the EU and its Member States should rely on robust evidence concerning AI's impact on fundamental rights to ensure that any restrictions of certain fundamental rights respect the principles of necessity and proportionality.

Relevant safeguards need to be provided for by law to effectively protect against arbitrary interference with fundamental rights and to give legal certainty to both AI developers and users. Voluntary schemes for observing and safeguarding fundamental rights in the development and use of AI can further help mitigate rights violations. In line with the minimum requirements of legal clarity – as a basic principle of the rule of law and a prerequisite for securing fundamental rights – the legislator has to take due care when defining the scope of any such AI law.

Given the variety of technology subsumed under the term AI and the lack of knowledge about the full scope of its potential fundamental rights impact, the legal definition of AI-related terms might need to be assessed on a regular basis.

FRA OPINION 2

The EU legislator should consider making mandatory impact assessments that cover the full spectrum of fundamental rights. These should cover the private and public sectors, and be applied before any AI-system is used. The impact assessments should take into account the varying nature and scope of AI technologies, including the level of automation and complexity, as well as the potential harm. They should include basic screening requirements that can also serve to raise awareness of potential fundamental rights implications.

Impact assessments should draw on established good practice from other fields and be regularly repeated during deployment, where appropriate. These assessments should be conducted in a transparent manner. Their outcomes and recommendations should be in the public domain, to the extent possible. To aid the impact assessment process, companies and public administration should be required to collect the information needed for thoroughly assessing the potential fundamental rights impact.

The EU and Member States should consider targeted actions to support those developing, using or planning to use AI systems, to ensure effective compliance with their fundamental rights impact assessment obligations. Such actions could include funding, guidelines, training or awareness raising. They should particularly – but not exclusively – target the private sector.

The EU and Member States should consider using existing tools, such as checklists or self-evaluation tools, developed at European and international level. These include those developed by the EU High-Level Group on Artificial Intelligence.

Using effective impact assessments to prevent negative effects

Prior impact assessments mainly focus on technical issues. They rarely address potential effects on fundamental rights. This is because knowledge on how AI affects such rights is lacking.

Deploying AI systems engages a wide spectrum of fundamental rights, regardless of the field of application. Pursuant to Article 51 (1) of the Charter, EU Member States must respect all rights embodied in the Charter when they are implementing Union law. In line with existing international standards – notably the United National Guiding Principles on Business and Human Rights (UNGPs) – businesses should have in place “a human rights due diligence process to identify, prevent, mitigate and account for how they address their impacts on human rights” (Principles 15 and 17). This is irrespective of their size and sector, and encompasses businesses working with AI.

While pursuing its commitments to the UNGPs, the EU has adopted several legislative acts addressing sector-specific instruments, in particular in the context of due diligence-related obligations for human rights. Discussions are currently underway on proposing new EU secondary law. Such law would require businesses to carry out due diligence of the potential human rights and environmental impacts of their operations and supply chains. Such law would likely be cross-sectoral and provide for sanctions for non-compliance – which should encompass the use of AI. See FRA’s recent report on ***Business and Human rights – access to remedy***, which calls for improved horizontal human rights diligence rules for EU-based companies.

Impact assessments are an important tool for businesses and public administration alike to mitigate the potential negative impact of their activities on fundamental rights. EU law in specific sectors requires some forms of impact assessments, such as Data Protection Impact Assessments under the General Data Protection Regulation (GDPR). Many interviewees reported that a data protection impact assessment, as required by law, was conducted. However, these took different forms. Moreover, prior assessments, when conducted, focus mainly on technical aspects. They rarely address potential impacts on fundamental rights. According to some interviewees, fundamental rights impact assessments are not carried out when an AI system does not, or appears not to, affect fundamental rights negatively.

The research shows that the interviewees’ knowledge on fundamental rights – other than data protection and, to some extent, non-discrimination – is limited. The majority acknowledge, however, that the use of AI has an impact on fundamental rights. Some interviewees indicate that their systems do not affect fundamental rights, which is to some extent linked to the tasks the AI systems are used for.

All respondents are aware of data protection issues. Most respondents also realise that discrimination could – generally – be a problem when AI is used.

However, the exact meaning and applicability of rights related to data protection and non-discrimination remains unclear to many respondents.

The research findings show differences between the private and public sector. Interviewees from the private sector are often less aware of the wider range of fundamental rights that could be affected. Data protection issues are known to the private sector. However, other rights, such as non-discrimination or access to justice-related rights, are less well known among business representatives who work with AI. Some were fully aware of potential problems. But others said that the responsibility for checking fundamental rights issues lies with their clients.

Ensuring effective oversight and overall accountability

Businesses and public administrations that are developing and using AI are in contact with various bodies that are responsible for overseeing AI-related systems within their respective mandates and sectors. These bodies include data protection authorities. But those using AI are not always sure which bodies are responsible for overseeing AI systems.

In line with well-established international human rights standards – for example, Article 1 of the European Convention on Human Rights (ECHR) and Article 51 of the Charter – states are obliged to secure people's rights and freedoms. To effectively comply, states have to – among others – put in place effective monitoring and enforcement mechanisms. This applies equally with respect to AI.

At the level of monitoring, the findings point to the important role of specialised bodies established in specific sectors that are also responsible for AI oversight within their mandates. These include, for example, oversight in the area of banking, or data protection authorities. A variety of such bodies are potentially relevant to the oversight of AI from a fundamental rights perspective. However, the responsibilities of bodies concerning the oversight of AI remains unclear to many of those interviewed from the private and the public sector.

Public administrations' use of AI is sometimes audited, as part of their regular audits. Private companies in specific sectors also have specialised oversight bodies, for example in the area of health or financial services. These also check the use of AI and related technologies, for example as part of their certification schemes. Private sector interviewees expressed a wish for bodies that could provide expert advice on the possibilities and legality of potential AI uses.

In the EU, there is a well-developed set of independent bodies with a mandate to protect and promote fundamental rights. These include data protection authorities, equality bodies, national human rights institutions and ombuds institutions. The research shows that those using or planning to use AI often contacted different bodies about their use of AI, such as consumer protection bodies.

FRA OPINION 3

The EU and Member States should ensure that effective accountability systems are in place to monitor and, where needed, effectively address any negative impact of AI systems on fundamental rights. They should consider, in addition to fundamental rights impact assessments (see FRA opinion 2), introducing specific safeguards to ensure that the accountability regime is effective. This could include a legal requirement to make available enough information to allow for an assessment of the fundamental rights impact of AI systems. This would enable external monitoring and human rights oversight by competent bodies.

The EU and Member States should also make better use of existing oversight expert structures to protect fundamental rights when using AI. These include data protection authorities, equality bodies, national human rights institutions, ombuds institutions and consumer protection bodies.

Additional resources should be earmarked to establish effective accountability systems by 'upskilling' and diversifying staff working for oversight bodies. This would allow them to deal with complex issues linked to developing and using AI.

Similarly, the appropriate bodies should be equipped with sufficient resources, powers and – importantly – expertise to prevent and assess fundamental rights violations and effectively support those whose fundamental rights are affected by AI.

Facilitating cooperation between appropriate bodies at national and European level can help share expertise and experience. Engaging with other actors with relevant expertise – such as specialist civil society organisations – can also help. When implementing such actions at national level, Member States should consider using available EU funding mechanisms.

Most often, users of AI contacted data protection authorities to seek guidance, input or approval where personal data processing was involved. Interviewed experts highlight the relevance of data protection authorities for overseeing AI systems with respect to the use of personal data. However, they also note that data protection authorities are under-resourced for this task and lack specific expertise on AI issues.

Experts, including those working for oversight bodies such as equality bodies and data protection authorities, agree that the expertise of existing oversight bodies needs to be strengthened to allow them to provide effective oversight of AI related issues. According to the experts, this can be challenging given that these bodies' resources are already stretched. They also highlighted the important role of relevant civil society organisations specialised in the fields of technology, digital rights and algorithms. They can enhance accountability in the use of AI systems.

NON-DISCRIMINATION, DATA PROTECTION AND ACCESS TO JUSTICE: THREE HORIZONTAL THEMES

The research shows that the use of AI affects various fundamental rights. Apart from context-related specific aspects that affect different rights to a varying extent, the fundamental rights topics which emerged in the research to repeatedly apply to most AI cases include: the need to ensure non-discriminatory use of AI (right not to be discriminated); the requirement to process data legally (right to personal data protection); and the possibility to complain about AI-based decisions and seek redress (right to an effective remedy and to a fair trial).

The two main fundamental rights highlighted in the interviews are data protection and non-discrimination. In addition, effective ways to complain about the use of AI came up repeatedly, linked to the right to a fair trial and effective remedy. The following three FRA opinions, which reflect these findings, should be read alongside the other opinions, which call for a more comprehensive recognition of, and response to, the full range of fundamental rights affected by AI.

FRA OPINION 4

EU Member States should consider encouraging companies and public administration to assess any potentially discriminatory outcomes when using AI systems.

The European Commission and Member States should consider providing funding for targeted research on potentially discriminatory impacts of the use of AI and algorithms. Such research would benefit from the adaptation of established research methodologies, from the social sciences, that are employed to identify potential discrimination in different areas – ranging from recruitment to customer profiling.

Building on the results of such research, guidance and tools to support those using AI to detect possible discriminatory outcomes should be developed.

Specific safeguards to ensure non-discrimination when using AI

> Interviewees rarely mentioned carrying out detailed assessments of potential discrimination when using AI. This suggests a lack of in-depth assessments of such discrimination in automated decision making.

The obligation to respect the principle of non-discrimination is enshrined in Article 2 of the TEU, Article 10 of the TFEU (requiring the Union to combat discrimination on a number of grounds), and Articles 20 and 21 of the Charter (equality before the law and non-discrimination on a range of grounds). More specific and detailed provisions in several EU directives also enshrine this principle, with varying scopes of application.

Automation and the use of AI can greatly increase the efficiency of services and can scale up tasks that humans would not be able to undertake. However, it is necessary to ensure that services and decisions based on AI are not discriminatory. Recognising this, the European Commission recently highlighted the need for additional legislation to safeguard non-discrimination when using AI in the **EU anti-racism action plan 2020-2025**.

Most interviewees are in principle aware that discrimination might happen. Yet, they rarely raised this issue themselves. Only few believe their systems could actually discriminate.

Interviewees also rarely mentioned detailed assessments of potential discrimination, meaning that there is a lack of in-depth assessment of potential discrimination.

A common perception is that omitting information about protected attributes, such as gender, age or ethnic origin, can guarantee that an AI system does not discriminate. This is not necessarily true, however. Information potentially indicating protected characteristics (proxies), which can often be found in datasets, could lead to discrimination.

In certain cases, AI systems can also be used to test for and detect discriminatory behaviour, which can be encoded in datasets. However, very few interviewees mentioned the possibility of collecting such information about disadvantaged groups to detect potential discrimination. In the absence of in-depth analysis of potential discrimination in the actual use of AI systems, there is also almost no discussion and analysis of the potential positive effect of using algorithms to make decisions fairer. Moreover, none of the interviewees working on AI mentioned using AI to detect possible discrimination as a positive outcome, in the sense that discrimination can be better detected when data are analysed for potential bias.

Since detecting potential discrimination through the use of AI and algorithms remains challenging, and interviewees only briefly addressed the issue, different measures are needed to address this. These include the requirement to consider issues linked to discrimination when assessing the use of AI, and investment into further studies of potential discrimination that use a diverse range of methodologies.

This could involve, for example, discrimination testing. This could build on similar established methodologies for testing bias in everyday life, such as with respect to job applications, where the applicant's name is changed to (indirectly) identify ethnicity. In relation to AI applications, such tests could involve the possible creation of fake profiles for online tools, which only differ with respect to protected attributes. In this way, the outcomes can be checked with respect to potential discrimination. Research could also benefit from advanced statistical analysis to detect differences in datasets concerning protected groups, and therefore can be used as a basis for exploring potential discrimination.

Finally, some research interviews underscored that results from complex machine learning algorithms are often very difficult to understand and explain. Thus, further research to better understand and explain such results (so-called 'explainable AI') can also help to better detect discrimination when using AI.

FRA OPINION 5

The European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) should consider providing further guidance and support to effectively implement GDPR provisions that directly apply to the use of AI for safeguarding fundamental rights, in particular as regards the meaning of personal data and its use in AI, including in AI training datasets.

There is a high level of uncertainty concerning the meaning of automated decision making and the right to human review linked to the use of AI and automated decision making. Thus, the EDPB and the EDPS should also consider further clarifying the concepts of 'automated decision making' and 'human review', where they are mentioned in EU law.

In addition, national data protection bodies should provide practical guidance on how data protection provisions apply to the use of AI. Such guidance could include recommendations and checklists, based on concrete use cases of AI, to support compliance with data protection provisions.

More guidance on data protection

More clarity is needed on the scope and meaning of legal provisions regarding automated decision making.

Data protection is critical in the development and use of AI. Article 8 (1) of the Charter and Article 16 (1) of the TFEU provide that everyone has the right to the protection of their personal data. The GDPR and the Law Enforcement Directive (Directive (EU) 2018/680) further elaborate on this right, and include many provisions applicable to the use of AI.

The interviewees indicated that most of the AI systems they employ use personal data, meaning data protection is affected in many different ways. However, a few applications – according to the interviewees – do not use personal data, or only use anonymised data, and hence data protection law would not apply. If personal data are used, all data protection related principles and provisions apply.

This report highlights an important issue linked to data protection, which is also relevant for other fundamental rights with respect to automated decision making. According to a Eurobarometer survey, only 40 % of Europeans know that they can have a say when decisions are automated. Knowledge about this right is considerably higher among those working with AI – the majority of interviewees raised this issue. However, many of the interviewees, including experts, argued that more clarity is needed on the scope and meaning of legal provisions on automated decision making.

In the area of social benefits, interviewees mentioned only one example of fully automated, rule-based decisions. All other applications they mentioned are reviewed by humans. Interviewees in public administration stressed the importance of human review of any decisions. However, they rarely described what such human review actually involves and how other information was used when reviewing output from AI systems.

While interviewees disagree as to whether or not the existing legislation is sufficient, many called for more concrete interpretation of the existing data protection rules with respect to automated decision making, as enshrined in Article 22 of the GDPR.

Effective access to justice in cases involving AI-based decisions

To effectively contest decisions based on the use of AI, people need to know that AI is used, and how and where to complain. Organisations using AI need to be able to explain their AI system and decisions based on AI.

Access to justice is both a process and a goal, and is crucial for individuals seeking to benefit from other procedural and substantive rights. It encompasses a number of core human rights. These include the right to a fair trial and to an effective remedy under Article 6 and 13 of the ECHR and Article 47 of the EU Charter of Fundamental Rights. Accordingly, the notion of access to justice obliges states to guarantee each individual's right to go to court – or, in some circumstances, an alternative dispute resolution body – to obtain a remedy if it is found that the individual's rights have been violated.

In accordance with these standards, a victim of a human rights violation arising from the development or use of an AI system by a public or private entity has to be provided with access to remedy before a national authority. In line with relevant case law under Article 47 of the Charter and Article 13 of the ECHR, the remedy must be "effective in practice as well as in law".

The research findings identify the following preconditions for the remedy to be effective in practice in cases involving AI systems and their impact on fundamental rights: everyone needs to be aware when AI is used and informed of how and where to complain. Organisations using AI must ensure that the public is informed about their AI system and the decisions based on them.

The findings show that explaining AI systems and how they make decisions in layman terms can be challenging. Intellectual property rights can hamper the provision of detailed information about how an algorithm works. In addition, certain AI systems are complex. This makes it difficult to provide meaningful information about the way a system works, and on related decisions.

To tackle this problem, some companies interviewed avoid using complex methods for certain decision making altogether, because they would not be able to explain the decisions. Alternatively, they use simpler data analysis methods for the same problem to obtain some understanding about the main factors influencing certain outcomes. Some of the private sector interviewees pointed to efforts made to gradually improve their understanding of AI technology.

FRA OPINION 6

The EU legislator and Member States should ensure effective access to justice for individuals in cases involving AI-based decisions.

To ensure that available remedies are accessible in practice, the EU legislator and Member States could consider introducing a legal duty for public administration and private companies using AI systems to provide those seeking redress information about the operation of their AI systems. This includes information on how these AI systems arrive at automated decisions. This obligation would help achieve equality of arms in cases of individuals seeking justice. It would also support the effectiveness of external monitoring and human rights oversight of AI systems (see FRA opinion 3).

In view of the difficulty of explaining complex AI systems, the EU, jointly with the Member States, should consider developing guidelines to support transparency efforts in this area. In so doing, they should draw on the expertise of national human rights bodies and civil society organisations active in this field.

1

AI AND FUNDAMENTAL RIGHTS – WHY IT IS RELEVANT FOR POLICYMAKING

Artificial intelligence (AI) is increasingly used in the private and public sectors, affecting daily life. Some see AI as the end of human control over machines. Others view it as the technology that will help humanity address some of its most pressing challenges. While neither portrayal may be accurate, concerns about AI's fundamental rights impact are clearly mounting, meriting scrutiny of its use by human rights actors.

Examples of potential problems with using AI-related technologies in relation to fundamental rights have increasingly emerged. These include:

- an algorithm used to recruit human resources was found to generally prefer men over women;[1]
- an online chatbot[2] became 'racist' within a couple of hours;[3]
- machine translations showed gender bias;[4]
- facial recognition systems detect gender well for white men, but not for black women;[5]
- a public administration's use of algorithms to categorise unemployed people did not comply with the law;[6]
- and a court stopped an algorithmic system supporting social benefit decisions for breaching data protection laws.[7]

These examples raise profound questions about whether modern AI systems are fit for purpose and how fundamental rights standards can be upheld when using or considering using AI systems.

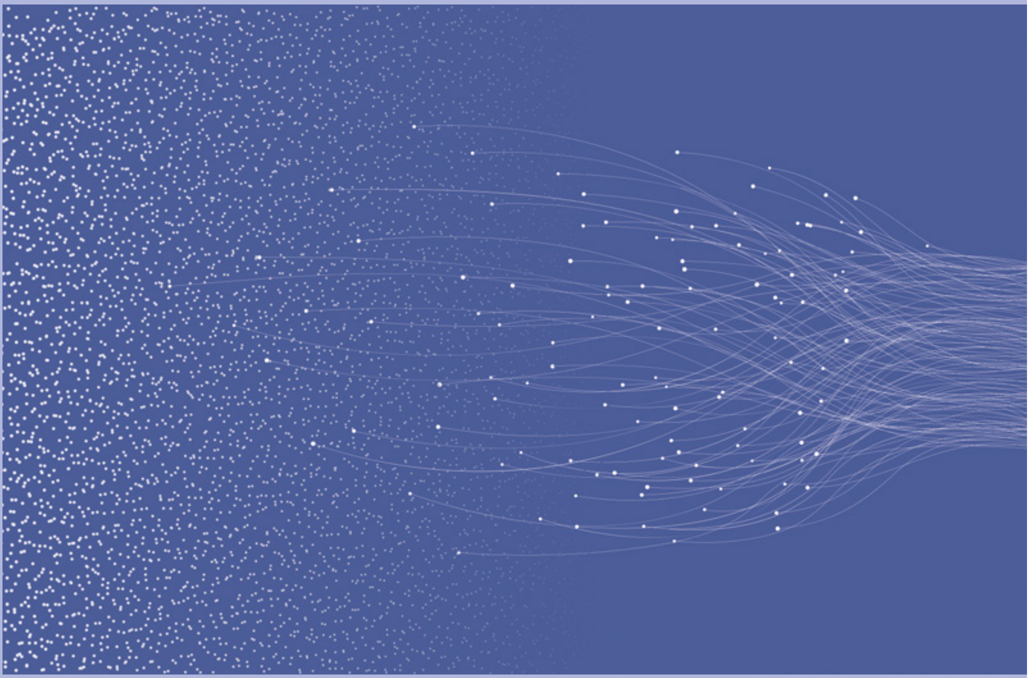
This report addresses these questions by providing a snapshot of the current use of AI-related technologies in the EU – based on selected use cases – and its implications on fundamental rights.

FRA's work on AI, big data and fundamental rights

This report is the main publication stemming from FRA's **project on Artificial intelligence, big data and fundamental rights**. The project aims to assess the positive and negative fundamental rights implications of new technologies, including AI and big data.

The current report builds on the findings of a number of earlier papers:

- **Facial recognition technology: fundamental rights considerations in the context of law enforcement** (2019): this paper outlines and analyses fundamental rights challenges triggered when public authorities deploy live FRT for law enforcement purposes. It also briefly presents steps to take to help avoid rights violations.
- **Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights** (2019): this paper highlights the importance of awareness and avoidance of poor data quality.
- **#BigData: Discrimination in data-supported decision making** (2018): this focus paper discusses how such discrimination can occur and suggests possible solutions.



As part of the project, FRA is also exploring the feasibility of studying concrete examples of fundamental rights challenges when using algorithms for decision making through either online experiments or simulation studies.

Several other FRA publications address relevant issues:

- The **Guide on Preventing unlawful profiling today and in the future** (2018) illustrates what profiling is, the legal frameworks that regulate it, and why conducting profiling lawfully is both necessary to comply with fundamental rights and crucial for effective policing and border management.
- The **Handbook on European data protection law** (2018 edition) is designed to familiarise legal practitioners not specialised in data protection with this area of law.
- Data from FRA's **Fundamental Rights Survey**. It surveyed a random sample of 35,000 people across the EU, including findings on people's opinions and experiences linked to **data protection and technology** (2020) and **security** (2020).
- FRA's report on **Business and human rights – access to remedy** analyses obstacles and promising practices in relation to access to remedies for victims of business-related human rights abuses. By analysing complaints mechanisms in EU Member States, the research maps what hinders and what facilitates access to remedies.

1.1. WHY THIS REPORT?

The growing attention to AI and its potential to drive economic growth has not been matched by a body of evidence about how different technologies can affect fundamental rights – positively or negatively. Only concrete examples allow for a thorough examination of whether, and to what extent, applying a technology interferes with various fundamental rights – and whether any such interference can be justified, in line with the principles of necessity and proportionality.

This report provides a fundamental rights-based analysis of concrete 'use cases' – or case studies. 'Use case' is a term in software engineering. This report loosely defines it as the specific application of a technology for a certain goal used by a specified actor.

The report illustrates some of the ways that companies and the public sector in the EU are looking to use AI to support their work, and whether – and how – they are taking fundamental rights considerations into account. In this way, it contributes empirical evidence, analysed from a fundamental rights perspective, that can inform EU and national policymaking efforts to regulate the use of AI tools.

What did the research cover?

FRA conducted fieldwork research in five EU Member States: Estonia, Finland, France, the Netherlands and Spain. It collected information from those involved in designing and using AI systems in key private and public sectors on how they address relevant fundamental rights issues.

The research – based on 91 personal interviews – gathered information on:

- the purpose and practical application of AI technologies;
- the assessments conducted when using AI and the applicable legal framework and oversight mechanisms;
- the awareness of fundamental rights issues and potential safeguards in place; and
- future plans.

In addition, 10 experts involved in monitoring or observing potential fundamental rights violations concerning the use of AI, including civil society, lawyers and oversight bodies, were interviewed.

Presenting the main findings

This report presents the main findings of the fieldwork. In particular, the report includes:

- An overview of the use of AI in the EU across a range of sectors, with a focus on: (1) social benefits, (2) predictive policing, (3) healthcare, and (4) targeted advertising.
- An analysis of the awareness of fundamental rights and further implications on selected rights, with a focus on the four use cases.
- A discussion of measures to assess and mitigate the impact of AI-related technologies on people's fundamental rights.

Two annexes, available on **FRA's website**, supplement the report:

- Annex 1 gives a detailed description of the research methodology and the questions asked in the interviews.
- Annex 2 provides examples of potential errors when using AI in selected areas.

In addition, country-specific information on each of the five Member States covered complements the fieldwork. This research, delivered by the contractor, is also available on **FRA's website**. It maps policy developments on AI and the legal framework governing its use in different sectors.

Supporting rights-compliant policymaking

This report provides evidence on the extent to which fundamental rights considerations are brought into discussions and activities to develop, test, employ and monitor AI systems in the EU. It also highlights how different technologies can affect some of the rights set out in the Charter, and reflects on how to protect these rights as AI becomes both more widespread and more sophisticated.

The analysis of selected fundamental rights challenges can help the EU and its Member States, as well as other stakeholders, assess the fundamental rights compatibility of AI systems in different contexts. The findings in the report about current views and practices among those using AI supports policymakers in identifying where further actions are needed.

The report does not aim to provide a comprehensive mapping of the use of different AI systems in the five EU Member States covered by the research, or to provide in-depth technical information about how the different systems mentioned by the interviewees work.

Conducting the interviews

Who?

This report is based on 91 semi-structured interviews with representatives from public administration and private companies who are involved in the use of AI for their services and businesses. FRA intentionally provided a very general definition of AI to those interviewed as part of the research, based on existing definitions.

The organisations interviewed were active in public administration in general, with some working in law enforcement.

The private companies include those working in health, retail, pricing and marketing, financial services, insurance, employment, transport and energy. Importantly, except for two interviewees, the research did not include companies that sell AI to other companies. Instead, the entities use AI to support their own operations.

In addition, ten interviews were conducted with experts dealing with potential challenges of AI in public administration (e.g. supervisory authorities), in non-governmental organisations or as lawyers working in this field.

Where?

Interviews were carried out in five EU Member States (Estonia, Finland, France, the Netherlands and Spain). These countries were selected based on their different levels of uptake of AI technology and of policy development in the area of AI, as well as to incorporate experience from across different parts of the EU.

How?

FRA outsourced the fieldwork to **Ecorys**. FRA staff supervised the work, and developed the research questions and methodology. Interviewers received dedicated training before conducting the fieldwork.

Interviews were carried out anonymously. As a consequence, no information identifying the organisation concerned is provided in the report. In addition, certain details of the applications described – most notably the country – are omitted to protect respondents' anonymity. This was communicated to interviewees, increasing their level of trust and allowing them to speak more freely about their work. It also proved useful for recruiting respondents.

1.2. WHAT DO WE MEAN BY ARTIFICIAL INTELLIGENCE?

There is no universally accepted definition of AI. Rather than referring to concrete applications, it reflects recent technological developments that encompass a variety of technologies. Although AI is usually defined very widely, a survey conducted in 2020 on behalf of the European Commission among companies in the EU showed that eight in ten people working at companies in the EU say they know what AI is. Slightly more than two in 10 respondents from companies in the EU-27 do not know (7 %) or are not sure about (14 %) what AI is.[8]

High-level expert group on artificial intelligence

"Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)."

*This **initial definition** of AI HLEG was subject to further discussion in the groups. See AI HLEG (2019), **A definition of AI: Main capabilities and disciplines**.*

FRA's research did not apply a strict definition of AI on the use cases it presents. For the interviews, AI was defined broadly, with reference to the definition provided by the High-Level Expert Group on Artificial Intelligence (AI HLEG).

The interviewees also expressed a variety of ways to think about AI. When identifying use cases to explore in the research, the project focused on applications that support decision making based on data and machine learning, and applications and systems that contribute to automating tasks that are usually undertaken by humans or which cannot be undertaken by humans due to their large scale. As such, the use cases in this report provide insight into the different technologies that are used and discussed in selected areas under the broad heading of AI. As there may be some contention concerning whether certain use cases constitute AI at the current level of use, the report often refers to 'AI and related technologies'.

The past years have seen an enormous increase in computing power, increased availability of data and the development of new technologies for analysing data. The increased amount and variety of data, sometimes available almost in real time over the internet, is often referred to as big data. Machine learning technologies and related algorithms, including deep learning, benefit enormously from this increased computing power and data availability, and their development and use is flourishing.

The use of these terms is, however, of limited use. It can even prove counterproductive, as it triggers ideas linked to science fiction rather than any real application of AI. A variety of myths exist about what AI is and can do,[9] often spread via (social) media. For example, some claim that AI can act on its own, being some form of entity. This distracts from the fact that all AI systems are made by humans and that computers only follow instructions made and given by humans. For a human-centric approach to AI, it is important to note that AI can never do anything on its own – it is human beings who use technology to achieve certain goals. However, the human work and decision making behind the AI systems is often not visible or the centre of attention.

"Currently, there is no lawyer who can tell the definition of AI and we've asked around pretty thoroughly. No one can tell."
(Public administration, Netherlands)

Entire studies and many discussions have explored possible AI definitions. The European Commission's Joint Research Centre analysed AI definitions. It highlights that they often refer to issues linked to the perception of the environment (i.e. the way a system receives input/data from its environment, e.g. through sensors), information processing, decision making and the achievement of specific goals. Definitions frequently refer to machines behaving like humans or taking over tasks associated with human intelligence. Given the difficulty of defining intelligence, many definitions remain vague. This makes the use of AI hard to measure in practice[10] and, equally, challenging to define in law.[11]

This report discusses the use of AI based on concrete applications. These differ in terms of their complexity, level of automation, potential impact on individuals, and the scale of application.

Most of the discussion around, and the actual use of AI, involves deploying machine learning technologies. These can be seen as a sub-domain of AI. There is also some confusion around the term "learning", which implies that machines learn like humans. In reality, much of current machine learning is based on statistical learning methodologies.[12] Machine learning uses statistical methods to find rules in the form of correlations that can help to predict certain outcomes.

This is different from traditional statistical analysis, because it does not involve detailed checks of how these predictions were produced (often referred to as 'black boxes'[13]). Traditional statistical analysis is based on specific theoretical assumptions about the data generation processes and the correlations used.[14] Machine learning is geared towards producing accurate outcomes, and can be used for automating workflows or decisions, if an acceptable level of accuracy can be obtained.

The usual example is an email spam filter, which uses statistical methods to predict if an email is spam. As it is not important to know why a certain email was blocked and because spam can be predicted with very high accuracy, we do not really need to understand how the algorithm works (i.e. based on what rules emails get blocked). However, depending on the complexity of the task, prediction is not always possible with high accuracy. Moreover, as this report highlights, not understanding why certain outcomes are predicted is not acceptable for certain tasks.

The area of machine learning incorporates several approaches. Most often, machine learning refers to finding rules that link data to a certain outcome based on a dataset that includes outcomes (supervised learning). For example, a dataset of emails, which are labelled as spam or not ('ham'), is used to find correlations and rules that are associated with spam emails in this dataset. These rules are then used to 'predict' with some degree of likelihood if any future email is spam or not.

Sometimes, machine learning is used to find hidden groups in datasets without defining a certain outcome (unsupervised learning) – for example, segmenting people into groups based on similarities in their demographics.

Finally, rules and correlations can be found through trial and error (reinforcement learning). These systems try to optimise a certain goal through experimentation, and update their rules automatically to have the best possible output. Such systems need enormous amounts of data and can hardly be used on humans, as it involves experimentation. They were mainly responsible for the success of winning board games against humans, which were often sensationalised by media.

1.3. AI AND FUNDAMENTAL RIGHTS IN THE EU POLICY FRAMEWORK: MOVING TOWARDS REGULATION

Policymakers have for some time highlighted the potential for AI and related technologies to improve efficiency and drive economic growth. Yet public authorities and international organisations have only recently reflected on the fundamental rights challenges associated with such technologies. Coupled with the growing use and accuracy of AI systems, this has turned attention to whether and how to regulate their use.

A 2017 European Parliament resolution marked a milestone in the EU's recognition of the fundamental rights implications of AI. The resolution stressed that "prospects and opportunities of big data can only be fully tapped into by citizens, the public and private sectors, academia and the scientific community when public trust in these technologies is ensured by a strong enforcement of fundamental rights".[15] It called on the European Commission, the Member States, and data protection authorities "to develop a strong and common ethical framework for the transparent processing of personal data and automated decision-making that may guide data usage and the ongoing enforcement of Union law". [16]

Later that year, the European Council called for a "sense of urgency to address emerging trends" including "issues such as artificial intelligence [...], while at the same time ensuring a high level of data protection, digital rights and ethical standards".[17] The European Council invited the European Commission to put forward a European approach to AI.

Responding to these calls, the European Commission published in 2018 its Communication on AI for Europe[18] and set up a High Level Expert Group on AI.[19] Both initiatives include a strong reference to fundamental rights.

The Commission-facilitated High Level Expert Group was made up of 52 independent experts from academia, civil society and industry (including a representative from FRA). It published 'Ethics Guidelines for Trustworthy AI' and 'Policy and investment recommendations for trustworthy AI' in 2019. These were developed further in 2020.[20] Its work triggered further discussion on the importance of framing AI in human rights terms, alongside ethical considerations. This led to the development of Ethics Guidelines that refer to the Charter and place fundamental rights consideration with respect to AI. The Ethics Guidelines include an assessment list for trustworthy AI, which has been translated into a checklist to guide those who develop and deploy AI.[21]

Indicating political support at the highest level, the European Council calls in its Strategic Guidelines for 2019-2024 to "ensure that Europe is digitally sovereign" and for policy to be "shaped in a way that embodies our societal values".[22] Similarly, Commission President Von der Leyen committed to "put forward legislation for a coordinated European approach on the human and ethical implications of [AI]".[23] This prompted significant moves towards setting out an EU legal framework to govern the development and use of AI and related technologies, including with respect to their impact on fundamental rights.

In February 2020, the European Commission published a White Paper on artificial intelligence. It sets out policy options for meeting the twin objectives of "promoting the uptake of AI and addressing the risks associated with certain uses of this new technology". The paper promotes a common European approach to AI. It deems this necessary "to reach sufficient scale and avoid the fragmentation of the single market". As it notes, "[t]he introduction of national initiatives risks to endanger legal certainty, to weaken citizens' trust and to prevent the emergence of a dynamic European industry".[24] Legal uncertainty is also a concern of companies planning to use AI.

The Commission White Paper on AI highlights risks to fundamental rights as one of the main concerns associated with AI. It acknowledges that "the use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights, be it as a result from flaws in the overall design of AI systems, or from the use of data without correcting possible bias". It also lists some of the wide range of rights that can be affected.[25]

The White Paper on AI indicates the Commission's preference for the possible new regulatory framework to follow a risk-based approach, in which mandatory requirements would, in principle, only apply to high-risk applications. These would be determined on the basis of two cumulative criteria: if it is employed in a sector, such as healthcare, transport or parts of the public sector, where significant risks can be expected to occur; and if it is used in a manner where significant risks are likely to arise. This latter risk could be assessed based on the impact on the affected parties, adding a harm-based element.

The White Paper also highlights some instances where AI use for certain purposes should be considered high-risk, irrespective of the sector. These include the use of AI applications in recruitment processes or for remote biometric identification, including facial recognition technologies.

Following a public consultation, which ran from February to June 2020,[26] the Commission is expected to propose legislation on AI in the first quarter of 2021.[27]

Ahead of the proposal, the EU's co-legislators have considered various aspects of the potential legal framework. In October 2020, the European Parliament adopted resolutions with recommendations to the European Commission on a framework of ethical aspects of AI, robotics and related technologies,[28] and a civil liability regime for AI.[29] It also adopted a resolution on intellectual property rights for the development of artificial intelligence technologies,[30] and continues to work on resolutions on AI in criminal law and its use by the police and judicial authorities in criminal matters,[31] and AI in education, culture and the audio-visual sector.[32] It also established a special committee on artificial intelligence in the digital age.[33]

Following their meeting on 1-2 October 2020, the heads of state and government of the EU Member States declared that the "EU needs to be a global leader in the development of secure, trustworthy and ethical Artificial Intelligence" and invited the Commission to "provide a clear, objective definition of high-risk Artificial Intelligence systems.[34] In addition, the Council of the EU adopted Conclusions on shaping Europe's digital future[35] and on seizing the opportunities of digitalisation for access to justice, which included a dedicated section on deploying AI systems in the justice sector.[36] The German Presidency of the Council of the EU published conclusions on the Charter of Fundamental Rights in the context of artificial intelligence and digital change; the text was supported, or not objected to, by 26 Member States.[37]

The growing reference to fundamental rights in these discussions indicates that a fundamental rights framework alongside other legal frameworks[38] is necessary for an effective and human rights compliant evaluation of the many opportunities and challenges brought by new technologies. Many existing AI initiatives are guided by ethical frameworks, which are typically voluntary.

A fundamental rights-centred approach to AI is underpinned by legal regulation, where the responsibility for respecting, protecting and fulfilling rights rests with the State. This should guarantee a high level of legal protection against possible misuse of new technologies. It also provides a clear legal basis from which to develop AI, where reference to fundamental rights – and their application in practice – is fully embedded.[39]

In addition to steps towards legal regulation, the EU is taking significant policy and financial actions to support the development of AI and related technologies. Alongside the White Paper, the Commission published the European Data Strategy.[40] It aims to set up a single market for data, including nine common European data spaces, covering areas such as health data and financial data. The proposal for the 2021-2027 Multiannual Financial Framework would create a Digital Europe Programme worth € 6.8 billion to invest in the EU's "strategic digital capacities", including AI, in addition to funding through Horizon Europe and the Connecting Europe Facility.[41]

Other international actors are also considering steps to regulate AI. Most notably, the Council of Europe is an active player in the field of AI and related technologies. In September 2019, the Committee of Ministers of the Council of Europe set up the Ad Hoc Committee on Artificial Intelligence (CAHAI). It aims to examine "the feasibility and potential elements of a legal framework for the development, design and application of AI, based on the Council of Europe's standards on human rights, democracy and the rule of law".[42] In April 2020, the Committee of Ministers of the Council of Europe adopted recommendations on the human rights impact of algorithmic systems.[43]

In addition, the Organisation for Economic Cooperation and Development (OECD) adopted AI principles and created an AI policy observatory.[44] At global level, UNESCO is starting to develop a global standard setting instrument on AI.[45] These are selected examples of the wide range of legal and policy initiatives aiming to contribute to standard setting in the area of AI. This includes, amongst others, actual (draft) legislation, soft-law, guidelines and recommendations on the use of AI, or reports with recommendations for law and policy.

FRA put together a (non-exhaustive) list of initiatives linked to AI policymaking.[46] While these also include legislative initiatives in EU Member States, many organisations and businesses launched initiatives to tackle ethical concerns of AI. However, while useful to tackle potential problems with AI, ethical approaches often rely on voluntary action. This does not sufficiently address the obligation to respect fundamental rights.

As FRA pointed out in its *Fundamental Rights Report 2019*: "only a rights-based approach guarantees a high level of protection against possible misuse of new technologies and wrongdoings using them."[47] The European Commission's initiative on regulating AI helps to avoid disjointed responses to AI across Member States, which can undermine businesses across the EU and with entities outside the EU.

Endnotes

[1] Reuters (2018), **‘Amazon scraps secret AI recruiting tool that showed bias against women’**, 10 October 2018.

[2] Chatbot or chatterbot is a common AI feature embedded in messaging applications to simulate human conversation through voice or text.

[3] Independent (2017), **‘AI robots learning racism, sexism and other prejudices from humans, study finds’**, 17 April 2017.

[4] Prates, M., Avelar, P. and Lamb, L. (2019) ‘**Assessing Gender Bias in Machine Translation – A Case Study with Google Translate**’, 11 March 2019.

[5] The **Gender Shades project evaluating the accuracy of AI powered gender classification products**.

[6] See for example: Der Standard (2020), ***Datenschutzbehörde kippt umstrittenen AMS-Algorithmus***, or AlgorithmWatch (2019), **Poland: Government to scrap controversial unemployment scoring system.**

[7] Privacy First (2020), **Dutch risk profiling system SyRI banned following court decision**.

[8] European Commission (2020), **European enterprise survey on the use of technologies based on artificial intelligence**, Luxembourg, July 2020.

[9] See, for example, the website “**AI myths**”.

[10] Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B. (2020), **AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence**, Luxembourg.

[11] Schuett, J. (2019), A legal definition of AI, **arXiv: 1909.01095**

[12] Hastie, T., Tibshirani R., and Friedman, J. (2009), **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**, Springer.

[13] See, for example: Pasquale, F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambrigde and London; and Rai, A. (2020), ‘**Explainable AI: from black box to glass box**’, *Journal of the Academy of Marketing Science*, Vol. 48, pp. 137-141.

[14] A seminal paper describing this difference is: Breiman, L. (2001), ‘Statistical Modeling: The Two Cultures’, *Statistical Science*, 2001, Vol. 16, No. 3, pp. 199-231.

[15] **European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement** (2016/2225(INI)), para. 1.

[16] *Ibid.*, para. 20.

[17] European Council (2017), **European Council meeting (19 October 2017) – Conclusions**, EUCO 14/17, Brussels, 19 October 2017, p. 8.

[18] European Commission (2018), **Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe**, COM(2018) 237 final, 25 April 2018.

[19] More information is available on the **webpage of the High level expert group**.

[20] High-Level Expert Group on Artificial Intelligence (2019), **Ethics Guidelines for Trustworthy Artificial Intelligence**; **Policy and investment recommendations for trustworthy AI**.

[21] High-Level Expert Group on Artificial Intelligence (2020), **Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment**.

[22] European Council, **A New Strategic Agenda 2019-2014**, p. 4.

[23] Vonder Leyen, Ursula, ***A Union that strives for more: My agenda for Europe***, p. 13.

[24] European Commission, ***White Paper On Artificial Intelligence – A European approach to excellence and trust***, COM(2020) 65 final, Brussels, 19 February 2020, p. 2.

[25] *Ibid.*, p. 12.

[26] European Commission (2020), **White paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust**,17 July 2020.

[27] European Commission (2020), **Adjusted Commission Work Programme 2020, Annex I: New initiatives**, 27 May 2020.

[28] European Parliament, Legislative Observatory, **Framework of ethical aspects of artificial intelligence, robotics and related technologies**, 2020/2012 (INL).

[29] **European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence**, 2020/2014 (INL).

[30] **European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies**, 2020/2015 (INI).

[31] European Parliament, **Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters**, 2020/2016 (INI).

[32] European Parliament, Legislative Observatory, **Artificial intelligence in education, culture and the audiovisual sector**, 2020/2017 (INI).

[33] **European Parliament decision of 18 June 2020 on setting up a special committee on artificial intelligence in a digital age, and defining its responsibilities, numerical strength and term of office**, 2020/2684 (RSO).

[34] European Council (2020), **Special meeting of the European Council (1 and 2 October 2020) – Conclusions**, EUCO 13/20, 2 October 2020.

[35] Council of the European Union (2020), **Shaping Europe’s Digital Future – Council Conclusions**, 9 June 2020.

[36] Council of the European Union, **Council Conclusions “Access to Justice – Seizing the Opportunities of Digitalisation”**, 13 October 2020.

[37] Council of the European Union, **Presidency Conclusions – the Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change**, 21 October 2020.

[38] See e.g. Pagallo, U., Casanovas, P. & Madelin, R. (2019), ‘**The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data**’, *The Theory and Practice of Legislation* 7 (1), pp. 1-25.

[39] See FRA (2019), ***Fundamental Rights Report 2019***, Luxembourg, Publications Office, Chapter 7**.**

[40] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, **A European strategy for data**, COM/2020/66 final.

[41] European Council, ***Conclusions from Special meeting of the European Council (17, 18, 19, 20 and 21 July 2020)***, EUCO 10/20, 21 July 2020.

[42] Council of Europe, Ad Hoc Committee on Artificial Intelligence (CAHAI), **Factsheet: Governance for digital transformation**.

[43] Council of Europe, ***Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*** (adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers’ Deputies).

[44] See the dedicated **OECD website**.

[45] See the dedicated **UNESCO website**.

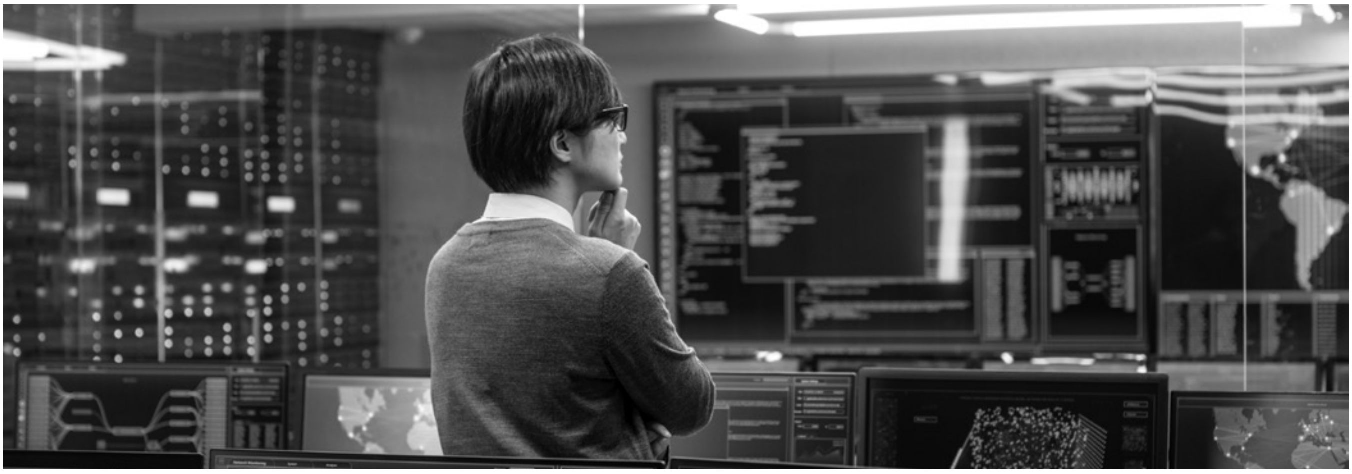
[46] See, for an overview by FRA, **AI Policy Initiatives**, or at the **Council of Europe website**.

[47] FRA (2019), ***Fundamental Rights Report***, Luxembourg, Publications Office, p. 166.

2

PUTTING FUNDAMENTAL RIGHTS IN CONTEXT – SELECTED USE CASES OF AI IN THE EU

In the EU, the use of AI-related technologies is relatively wide-spread. A recent survey shows that 42 % of companies use AI-related technologies – and that 18 % plan to do so.



Note on interviewees

The use cases presented in this chapter are based on information obtained in interviews with both public and private sector representatives.

The interviewed representatives from public administration work in the areas of health services, infrastructure and energy, the judiciary, law enforcement, migration and border management, social benefits, tax, as well as transportation and traffic control.

Interviewees from private companies mainly work in retail, marketing and pricing, the health sector, in financial services, energy, insurance, employment and transport, as well as in cross-cutting areas with a focus on AI development for different sectors.

This chapter presents selected cases of AI use – typically referred to as 'use cases' in the AI field. FRA collected information on such cases from five EU Member States: Estonia, France, Finland, the Netherlands and Spain. They involve different areas of application across public administration and private companies. Special focus is put on the use of AI in the areas of social benefits; predictive policing; health services; and targeted advertising.

The chapter provides information on the current use of AI, as well as basic information on EU competence, in these select areas. The use cases provide a good sense of what kind of AI and related technologies are currently being used.

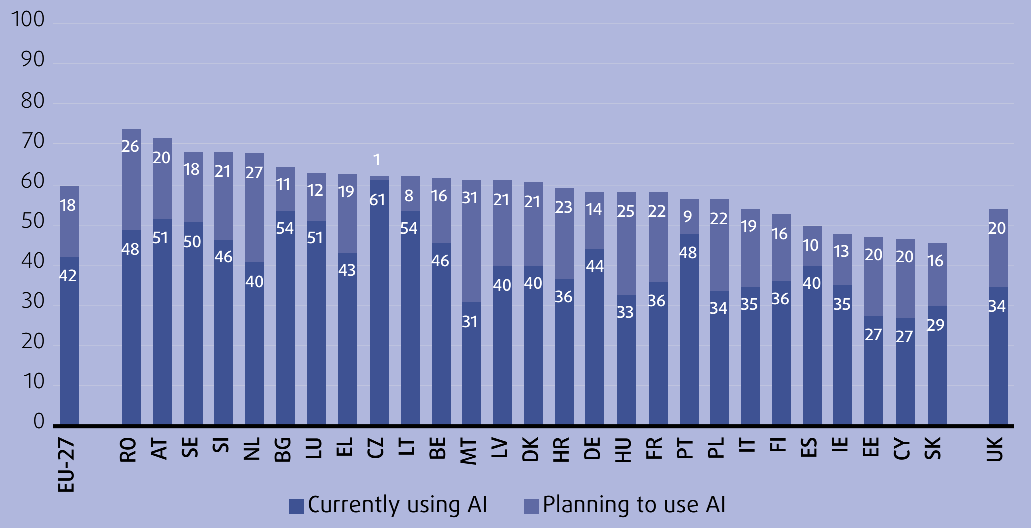
The examples also offer context for the fundamental rights analysis. Looking at a broad variety of use cases provides important insights on how the actual use of AI can affect people's fundamental rights. **Chapter 4** includes a discussion of fundamental rights implications, and makes reference to the cases described in this chapter.

Use of AI by companies in the EU in 2020

According to the European Enterprise Survey, at the beginning of 2020, 42 % of companies in the EU said they use technologies that depend on AI. This percentage ranges from 27 % in Estonia and Cyprus to 61 % in Czechia (see Figure 1). Another 18 % of companies are planning to use AI in the future.

The survey indicates that AI is used mostly in the IT sector (63 %). The technologies used comprise a variety of IT applications aiming at process or equipment optimisation, anomaly detection, process automation, and forecasting, price optimisation and decision making.

FIGURE 1: COMPANIES USING AI IN 2020, BY MEMBER STATE (%)



Notes: *The survey asked about the use or plans for use of ten different AI related technologies, such as speech recognition, visual diagnostics, fraud detection, analysis of emotions, forecasting based on machine learning and more. Includes the percentage of companies using at least one AI technologies. N = 9,640.*

Source: *FRA, 2020 [based on data extracted from European Commission,* **European enterprise survey on the use of technologies based on artificial intelligence***, Luxembourg, July 2020]*

As noted, this report focuses on four broad AI 'use cases':

- social benefits,
- predictive policing,
- health services, and
- targeted advertising.

These areas are particularly sensitive as regards fundamental rights. Two cover mainly the public administration's use of AI (social benefits allocation and predictive policing). The other two concern private companies (health services and targeted advertising). These use cases provide the basis for the report's fundamental rights analysis by offering the necessary context. Where appropriate, the report also highlights findings from interviews that cover areas other than these four areas.

"AI and machine learning are different concepts. AI is an umbrella term."
(Private company, Estonia)

Detailed studies on the taxonomy of AI are available,[1] providing further categorisations of the technology. As noted in the introduction, interviewees had different views about what AI is and some also stated that there is no clear definition of AI.

"What you see now is that everyone doing something with machine learning is labelling this as 'AI'."
(Public administration, Netherlands)

This report discusses specific use cases without further classifying the technology applied. Yet the use of AI in the cases examined differed: the use of technology described by the interviewees involved both varying levels of complexity and varying levels of automation.

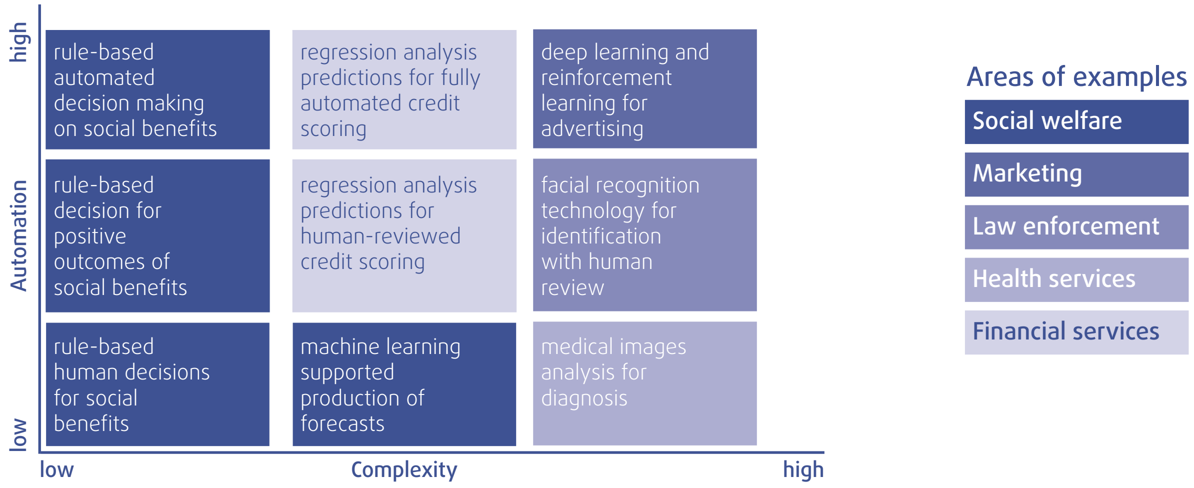
Figure 2 provides an overview of different examples of use that interviewees discussed under the heading of AI. Some applications are relatively straightforward to understand. In rule-based decision making, algorithms are defined based on 'if-then-rules' (for example, if a person has an income below a certain threshold, then they will be eligible for certain benefits). Such algorithms were used in the area of social benefits at different levels of automation, with examples of full, partial or no human review involved.

Other applications used more traditional statistical methods to inform decisions. These include, for example, *regression analysis*. This is a classical statistical method that analyses correlation between several pieces of information ('variables') and an outcome, which is a credit score in this example. Others used more complex machine learning methodologies to feed into the production of forecasts and statistics for government reports.

There are also algorithms with much higher levels of complexity, such as *deep learning* for diagnosis support in the area of health. Such tools still include a high level of human review, and hence do not include a high level of automation.

By contrast, targeted advertising is an example of potentially using highly complex algorithms without human review of each output and decision, also using highly complex algorithms including *deep learning* and *reinforcement learning*. (See **Chapter 1** for descriptions of these terms.) Human review would also not be possible in this area due to the scale at which such algorithms operate.

FIGURE 2: EXAMPLES OF DIFFERENT AUTOMATION AND COMPLEXITY LEVELS IN USE CASES COVERED



high	rule-based automated decision making on social benefits	regression analysis predictions for fully automated credit scoring	deep learning and reinforcement learning for advertising
Automation	rule-based decision for positive outcomes of social benefits	regression analysis predictions for human-reviewed credit scoring	facial recognition technology for identification with human review
low	rule-based human decisions for social benefits	machine learning supported production of forecasts	medical images analysis for diagnosis
	low	Complexity	high

Areas of examples

- Social welfare
- Marketing
- Law enforcement
- Health services
- Financial services

Source: FRA, 2020

Notes: The examples from financial services and the use of facial recognition technology are not covered in the detailed use case descriptions, but were mentioned in other interviews. The examples illustrate different levels of complexity and automation, as used in practice.

AI systems also vary according to the potential harm that could result from an erroneous decision based on the use of AI. Depending on the area of application, wrong decisions – based on erroneous outputs from the system – can have different impacts. When using AI for decision making, the consequences are different if a decision is affirmative but wrong (false positive) or negative but wrong (false negative).

These issues are particularly important when machine learning is used, as it is based on statistical calculations, which always come with some degree of error. While rule-based algorithms can also make mistakes (especially if they grow more complex), risks are lower because of the deterministic nature of the rules developed.

For example, when using AI to make decisions on social benefits, a false positive means that a person may erroneously receive benefits. This does not necessarily have a negative impact on the person concerned (unless the error is found out later and the money needs to be paid back). However, it negatively impacts on the public administration, as money is paid not in line with good administration practices. In contrast, a false negative would have a negative impact on the individual, because they would not receive benefits to which they are entitled. Annex 2, available on **FRA's website**, provides further hypothetical examples of effects of wrong decisions based on the use cases discussed.

Importantly, when automating tasks, the impact could also scale up, potentially exacerbating the negative effect on society as a whole. The severity and scale of potential harm is one aspect that needs to be taken into consideration when analysing potential limitations on fundamental rights with respect to the use of AI.

For example, small error rates when using facial recognition technology used by law enforcement might still lead to flagging many innocent people, if the technology is used at places where many people are analysed. This might apply to airports or train stations, where thousands of people could be scanned on a daily basis.[2] A potential bias in error rates could then lead to disproportionally targeting certain groups in society.

Technologies used across all cases identified in the research

Interviewees mostly mention **'machine learning'**, including the use of neural networks and its extensions (see Chapter 1 for a description of machine learning). Respondents either directly mentioned this, or mentioned subfields of machine learning, such as image recognition or facial recognition technology (FRT).

Most often, interviewees mentioned the use of **'supervised machine learning'** as mainly used to optimise for a specifically defined outcome. Yet sometimes **'unsupervised machine learning'** was also used to categorise or cluster data. Only one case referred to the use of **'reinforcement learning'**, without going much into detail.

Several respondents used **'natural language processing (NLP)'**. This is a technology to analyse text and speech, and is sometimes combined with machine learning algorithms.

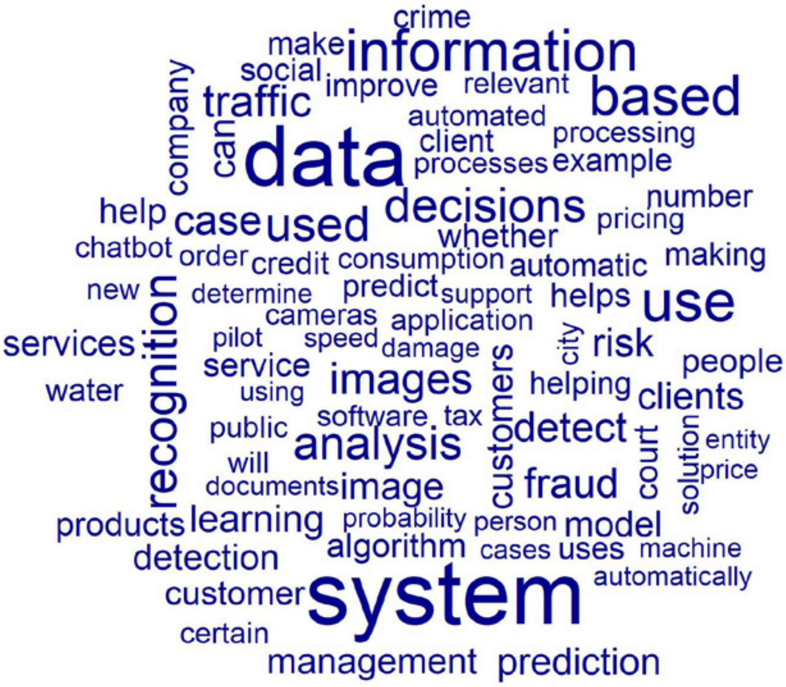
Few mention examples that involve rule-based algorithms, meaning that the rules for the algorithm to follow are directly encoded (i.e. based on 'if-then-rules').

In some cases, interviewees did not disclose or could not provide detailed information about the technology used.

Generally, the interviewees referred to more than one use case, but were asked to focus on one application during interviews.

- Importantly, the fieldwork shows that companies and public administrations are often still at the beginning of looking into the use of AI. Only about two thirds of the use cases are actually in use and deployed in practice. Many of the use cases described by interviewees are at pilot stage, under development, or still in the research phase.
- Two AI-driven applications were halted after tests.

FIGURE 3: WORDS INTERVIEWEES MOST OFTEN USED TO DESCRIBE THE AI 'USE CASES'



Notes: FRA visualisation of the words most frequently used in descriptions of use cases. The bigger the size of the word, the more often the interviewees mentioned the terms.

Source: FRA, 2020

Figure 3 shows the most frequently used words to describe the use cases covered in this report. It highlights the importance of data when using AI systems as well as its relevance to supporting decision making.

FRA has previously highlighted that a thorough description of the data used by AI applications is essential for identifying and mitigating potential fundamental rights challenges.[3] A variety of data were used for the AI systems covered in this report. However, it was difficult to obtain detailed information about the data used, because most respondents remained rather vague about their data sources.

Rather generically, many respondents mentioned using 'open data', 'historical data' or 'metadata'. More concretely, respondents mentioned using customer data, e.g. about purchases or browsing behaviour, or administrative records, such as data on social benefits and taxes. Interviewees also mentioned medical records, police records, court records, as well as social media and traffic data. Data included text data (e.g. e-mails), audio recordings, video, and geolocation data. Data come from internal databases of companies and public administration, but also from external sources.

The single most important reason for using AI is increased efficiency. The vast majority of respondents, across the public and private sector, mentioned using AI for greater speed, fewer errors and cost reduction, as fewer human resources are needed. Some interviewees from law enforcement also said they use AI for safety and security, as well as crime prevention.

Humans previously performed many of the use cases. Some respondents said they use AI because it entails fewer mistakes than having humans carry out

"It is mostly used to save time [...] when you have to go through a lot of material."
(Public administration, Netherlands)

"The most important is to deal with cases more efficiently. It's about making use of your workforce, the people who handle cases, as effectively as possible."
(Public administration, Netherlands)

certain tasks. Some respondents also use AI for tasks that humans did not previously carry out, as the quantity of information could not be processed by humans – for example, in the area of genome analysis or traffic predictions.

Importantly, for about half of the respondents interviewed, the use of AI is relevant for decision making. However, AI is mainly used to support decision making, and the final decisions remain largely in the hands of humans.

Interviewees pointed out that, while enthusiastic, public administration and companies are still cautious when deploying AI. Many of the use cases are still in the testing phase. And some, as described below, were stopped during this phase. Nevertheless, almost no interviewees were aware of any plans to reduce the level of technology used. In fact, most expressed intentions to invest in innovation or new ways to employ currently available AI systems.

2.1. EXAMPLES OF AI USE IN PUBLIC ADMINISTRATION

[Use case 1]

Automating social welfare systems – using algorithms in the area of social benefits

Background and EU legal framework

The United Nations Special Rapporteur on extreme poverty and human rights, Philip Alston, warned in his October 2019 report that introducing a 'digital welfare' state, including the use of AI, can lead to a "digital welfare dystopia". Digitalisation of welfare systems is often accompanied with reductions of overall welfare budgets, narrowing the beneficiary pool, and other measures that reduce the availability of welfare. Digitalisation also increases the power of states by offering opportunities to control people. This is particularly worrying in countries with significant rule of law deficits.[4]

The use of algorithms by public administration in welfare raises major concerns with respect to its potentially negative impact on poverty and inequality, if applied erroneously in the area of social benefits.[5] This includes areas such as child welfare services[6] and unemployment benefits.[7]

Yet public authorities are keen to use new technologies to make decision making on social security and other benefits more efficient and potentially fairer. Globally, new technologies are used in many ways to administer welfare systems. These include identity verification, eligibility assessments, benefit calculations, fraud prevention and detection, risk scoring and need classification, as well as communication between authorities and beneficiaries.

The OECD defines social benefits as transfers made to households in need after certain events or particular circumstances have arisen, including sickness, unemployment, retirement, housing, education or family circumstances.[8] However, there is no commonly agreed definition of social benefits. Social benefits, in particular social insurance, systems are different from private insurance schemes, as they involve compulsory contributions made by both employees and employers, sometimes in the form of taxation.[9]

Social policy, including social security and social protection, is an area of shared competence between the EU and the Member States (Article 4 (2) (b) of the TFEU). Pursuant to Article 151 of the TFEU, the EU pursues the objectives, among other things, to promote "improved living and working conditions" and "proper social protection". To this end, the EU supports and complements the activities of the Member States in a number of fields, including social security and social protection of workers and combating social exclusion (Article 153 (1) of the TFEU). EU actions can encourage cooperation between Member States and adopt directives with minimum requirements. Moreover, decisions on social security and social protection can only be adopted through special legislative procedure by a unanimous vote in the Council.[10]

Against this backdrop, EU Member States are mostly free to shape their social security and social protection policies. Since there is virtually no harmonisation, social security systems differ significantly across the EU in terms of what benefits are provided, conditions for eligibility, how benefits are calculated, what contributions need to be paid and by whom, etc.

Public administrations in EU Member States are working on implementing AI and related technologies in the area of public welfare. However, information about its applications is limited. FRA collected information about use cases linked to:

- using algorithms when it comes to compensating job seekers,
- processing social benefits applications, and
- machine learning-supported data analysis on the use of pensions.

Private insurance companies' use of AI

Several private insurance companies interviewed for this research use AI and related technologies. This includes handling requests of customers for complementary health insurance, insurance compensation decision support, evaluating the credit risk of individuals, insurance pricing, insurance claims management, and decision-making support related to management functions and credit decisions.

Private insurance companies generally embrace AI-related technologies, as these help make their business more profitable. An OECD report highlights the importance of technology for this sector. But it also argues that risk classification could lead to the exclusion of those belonging to certain vulnerable groups in ways that are undesirable from a societal and political perspective.*

** OECD (2020), The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector.*

Use in practice

The use cases outlined below exemplify some of the challenges when using or planning to use AI in the area of social benefits, linked to algorithmic decision making.

Experimenting with new technologies to support jobseekers

Over the course of a three-year project, a **public organisation** experimented with several AI-related technologies concerning all of their work related to processing benefits for job seekers and assisting them to return to work. The representative interviewed states that the tested technologies can improve and foster the relationship with job seekers and improve the advice given to both job seekers and companies. After testing is completed, the organisation will decide if and how it will apply these technologies in its day-to-day work.

Tests include machine learning-based detection of the attractiveness of job offers and a system for detecting whether job seekers are still actively looking for a job.

The tests are also looking into profiling job seekers to provide advice to them. This would include calculating the probability of someone being offered an available job within a given time, and identifying parameters that make job offers relevant. This may also be reflected in advice to companies on best practices for formulating job offers. The profiling would allow the organisation to determine appropriate services according to the profile and background of the job seeker, rather than having an analysis and advice drawn up by employees. Practically, this would be done by requiring job seekers to complete a monthly diary on their job search. However, it is still under consideration whether the programme should be limited to providing descriptive analyses, or whether it should go further and provide recommendations. The organisation is hesitant about the latter aspect.

Additionally, a natural language processing system is being tested for analysing the content of job seekers' e-mails. Here, e-mails are categorised, relevant data is extracted, and the urgency and relevance of e-mails is identified. Using a chatbot, and using automatic replies to emails, is being considered.

The data used for the systems come from several sources from within the organisation. The data on job seekers and their background, including personal tax data, as well as data on salaries and social security allowances, are used under very strict conditions. This is because they are derived from highly regulated data sources (e.g. salary statements cannot be accessed). Other data, such as job offers from companies, are also used to generate knowledge about the job market. The organisation currently does not use external data, such as from (professional) social media networks, because no legal provisions are in place for using such data.

Processing housing benefits – failure and success

A public body responsible for processing social benefits piloted an AI tool to process applications and subsequently support their staff in making decisions on housing benefits. The system selected cases from new benefit applications that were relatively straightforward to calculate. These include new applications for housing benefits submitted by an individual living alone or with children, and by an individual who does not have any other income than government benefits. Overall, these cases were deemed simple, with the result always being that the individual receives the benefits.

The technological solution was based on a decision-tree model following the rules for housing benefits. Calculating general housing benefits requires income estimates in advance. The data used during the testing stemmed from an internal database, which contains data on benefit application processes. The data was pseudonymised as there was no need to use personal information. A simple statistical model (linear regression) was used, where the input is the income and the cost limits, and the output is the amount of benefit.

However, even in such simplified cases, they found it too difficult to use AI in practice because of the frequent changes in the legislation. The test was terminated. According to the interviewee, the lack of a legal basis for using machine learning does not allow using it for administrative decisions. There are no further plans to use AI to support decision making on social benefits.

While the organisation is not pursuing this particular project due to the aforementioned legal challenges, the interviewee noted potential for further applications and solutions in this area in the future. It was noted that AI or related technologies that can support operations without having a legal impact were particularly good for the organisation.

The SyRI case

In the Netherlands, the so-called 'System Risk Indication' (SyRI)* was developed as a government tool to alert the Dutch public administration about fraud risk of citizens, by processing and linking large amounts of their personal data from public authorities.

A broad coalition of civil society organisations dealing with privacy issues initiated a lawsuit, prompting the District Court of The Hague to scrutinise the algorithm-based SyRI.**

The court ruled that SyRI impinges disproportionately on the private life of citizens. The court found that everyone whose data was analysed by SyRI was exposed to this risk. In addition, due to the opacity of the algorithm used, citizens could "neither anticipate the intrusion into their private life nor can they guard themselves against it." ***

A good description of SyRI can be found in Ilja Braun (2018), **High risk citizens, in: Algorithm Watch.*

*** The ruling of 5 February 2020 (in Dutch) is available **online**.*

**** Privacy First (2020), **Dutch risk profiling system SyRI banned following court decision**.*

At the same time, the organisation is using image processing for social benefits applications. Generally, benefit applicants have to complete several forms and attachments, which are often submitted in paper format. For more efficient and time-saving handling of those documents by the agency's staff, the hard copies received are scanned and then classified by an automated system.

A first step is to turn images the right way round. Algorithms re-align documents that were not aligned properly when they were scanned, remove spots and clean up and edit the colouring of the document, identify columns, paragraphs, tables, and other elements as distinctive blocks, recognise the script, etc. Then, the application checks if the received application form and attachment are marked correctly (e.g. if a document is marked as an invoice, the system determines whether this is correct).

The turning and the classification of the images are done by image recognition and Optical Character Recognition (OCR) technologies. They recognise text stemming from images, including from photographs and scans of documents or handwritten notes. OCR technology then converts the recognised text into text data that is machine-readable. Here, in a pattern recognition process, input from the scanned images is first isolated, then compared to 'glyphs' (i.e. variations of letters) stored by the system on a pixel to pixel basis.

The agency will continue processing images and further develop it, for example, by potentially making it possible to scan bar codes from attachments. This would help to speed up the confirmation of the correctness of documents and attachments. There will also be more solutions related to natural language processing.

Automating unemployment benefits

In one of the countries selected, most decisions on unemployment benefits are fully automated. The national institution responsible for unemployment insurance benefits updated its system in 2019 to fully automate most of the processing of benefit applications and decisions. This was done after the relevant legislation was adapted to allow automated decisions.

If a person registers as unemployed and lodges an application for benefits, the system draws on information about the applicant from various other databases. This includes, for example, the population register, and tax authorities' databases containing information about salaries and work experience, etc. If all conditions for receiving unemployment benefits are fulfilled, the system calculates the period of payments, based on the length the person has contributed to the insurance system, and the amount of benefits, based on the average daily salary.

The procedure is fully automated. However, an employee of the institution must intervene if necessary information cannot be extracted from the databases, if there is contradictory information in the databases or if the decision on a case involves a level of discretion (i.e. the decision cannot be definitively determined based on the data available and a human has some leeway in deciding on the case).

The main reason for using this system is improved efficiency. In addition, the system is believed to achieve consistency in the processes. This is because every application, not subject to discretion, is handled in the same way.

[Use case 2]

Predictive policing – trying to anticipate crime in advance

Background and EU legal framework

AI technologies are used in **law enforcement**, particularly in predictive policing. Existing research into how such tools can affect fundamental rights has highlighted particular issues concerning discrimination, among other rights. One recurrent concern is the potential for predictive policing to reproduce and entrench existing discriminatory practices, particularly through reliance on historical crime data that may be biased or incomplete. This is because many crimes – such as domestic violence or hate crime – remain largely unreported and therefore are under-counted in official police statistics.[11]

A focus on certain crimes, such as violence and drug-related crime in public places – rather than on business fraud and non-payment of taxes, for example – can also make law enforcement responses less equitable.[12] This is because the former are often associated with certain demographics and neighbourhoods. Ultimately, this can undermine police relations with particular communities.

Criminological research on crime 'hotspots' has been around for several decades – notably in the UK and USA.[13] It uses police data to map certain crimes and undertakes statistical tests to explore crime probabilities. Various police forces have used and developed them to address different types of crime concentrations or clusters ('hotspots').

More recently, adaptations of this area of applied research have used AI as a tool to enhance its effectiveness, with some suggesting that using algorithmic tools could reduce the police's reliance on subjective human judgments that may reflect biases or stereotypes.[14] Some studies have also indicated that predictive policing could potentially reduce unnecessary surveillance, questioning, and physical checks and searches,[15] reducing the humiliation and harassment of individuals that may occur during these activities.

Predictive policing aims to forecast the probability of crime and anticipate emerging trends and patterns to inform crime prevention and intervention strategies.[16] It may also be a part of an investigation into a crime that has already taken place. While there is no authoritative definition of predictive policing,[17] it is typically characterised by analysing data to identify common patterns and trends in crime by using algorithms to create models based on the analysis. This is used to forecast criminal activity that may occur in the future.

AI technologies in this area generally either aim to 'predict' crimes or to 'predict' which individuals will either commit or be victims of crimes. Tools aiming to predict crimes are generally fed with historical data – largely from official sources – on the time, place and type of crimes committed. This can be complemented by environmental variables, such as population density,

FRA ACTIVITY

Preventing unlawful profiling today and in the future: a guide

In developing and using algorithmic profiling, bias may be introduced at each step of the process. To avoid this and subsequent potential violations of fundamental rights, both IT experts and officers interpreting the data should have a clear understanding of fundamental rights.

This FRA guide explains what profiling is, the legal frameworks that regulate it, and why conducting profiling lawfully is both necessary to comply with fundamental rights and crucial for effective policing and border management.

For more information, see FRA (2018), ***Preventing unlawful profiling today and in the future: a guide****.*

presence of certain public places or services, and major events or holidays. They generally do not use personal data when applied.[18]

In contrast, AI systems focused on predicting potential perpetrators or victims of crime employ both historical and real-time personal data. This could include criminal records data, addresses, phone numbers, location data, data extracted from social media, information about known associates and health or income data. This is then combined with other criminal and environmental data.[19]

The EU and its Member States have shared competence in the area of freedom, security, and justice (Article 4 (2) (j) of the TFEU). This includes judicial cooperation in criminal matters and police cooperation (Articles 82-89 of the TFEU). Already when the Treaty of Lisbon was adopted, an annexed declaration on the protection of personal data in judicial cooperation in criminal matters and police cooperation observed that "specific rules on the protection of personal data and the free movement of such data in the fields of [...] police cooperation based on Article 16 of the [TFEU] [...] prove necessary because of the specific nature of these fields."[20]

Within the framework of predictive policing, the collection, storage, processing, analysis and exchange of information is particularly relevant. The processing of personal data in the context of law enforcement operations is regulated at EU level by the Law Enforcement Directive (Directive (EU) 2016/680). [21] It sets out comprehensive standards and safeguards for such processing, including the safeguarding against and the prevention of threats to public security.

Use in practice

The use cases collected by FRA signal the variety of ways in which law enforcement authorities already use, or plan to use, AI and related technologies to support their work.

Examples mentioned by interviewees range from data mining systems designed to map crime patterns, detecting online hate speech and making risk assessments on gender-based violence, to automating certain prison guard duties. Other use cases include detecting illicit objects from satellite images, and, more generally, recognising objects in images. In addition, a tool was mentioned in the research used in the private sector for fraud prevention and crime detection in money transfers.

Interviewees emphasised that AI or related technology systems are used to automate and speed up tasks previously done by humans, thus freeing up and/or better distributing resources.

Mapping crime to support the efficient allocation of investigation capacity

A national intelligence agency and public prosecutor's office employ a data-driven system to help their employees make choices on how, where and when to use the available investigation capacity. The aim is to improve the allocation of human resources, ensuring that officers can be present at the right time and place.

The interviewees suggest that this system could make more precise assessments compared to humans, who often rely on their gut feeling for decisions. Still, the system is always used in combination with human appraisal and other non-AI systems to make operational decisions.

Based on system-generated outcomes, analysts create a 'heat map'. This outlines the prevalence of certain crimes in certain areas. This replicates a long-standing manual version of this crime anticipation system, whereby

FRA ACTIVITY

Facial recognition technology on the rise: fundamental rights considerations in law enforcement

EU law recognises as 'sensitive data' people's facial images, which are a form of biometric data if processed by facial recognition software. But such images are also quite easy to capture in public places. Although the accuracy of matches is improving, the risk of errors remains real – particularly for certain minority groups. People whose images are captured and processed might not know this is happening – and so cannot challenge possible misuses.

The FRA paper outlines and analyses these and other fundamental rights challenges that are triggered when public authorities deploy live FRT for law enforcement purposes. It also briefly presents steps to take to help avoid rights violations.

For more information, see FRA (2019), **Facial recognition technology: fundamental rights considerations in the context of law enforcement**.

police officers put pins on a map to indicate specific risk areas. Using AI to increase the speed of this process also makes it more reliable, users believe, because it can analyse more data.

The system is based on data mining and machine learning processes. It is primarily built on unique police data contained in crime reports, witness statements, and suspect declarations. Gaps are, to the extent possible, addressed by using other data sources, such as criminology research, and social and demographic information obtained from the national office of statistics. The system also uses data from open sources.

The specific parameters for calculation depend on the type of crime, as predictive factors vary in relevance across crime areas. For example, in the case of burglaries, data on burglaries is collected and combined with data on the place of residence of known criminals and their distance to burgled houses. The relevant criteria are preselected to allow the system to produce the heat map.

Location-based predictions are made for the next six months, and indicate the time and location where a burglary may occur. The result is a map of small squares where the risk of crime occurring is indicated in different shades. The interviewees indicated that this visualisation helps officers to analyse neighbourhoods and observe correlations between different locations.

Assessing the risk of gender-based domestic violence

A national police force uses an internal system to track cases of gender-based domestic violence. The system helps police officers take decisions and distribute resources across domestic violence cases. The system categorises cases on the basis of the assessed risk of relapse and repetition, in order to focus on the 'riskiest' cases.

A specialist team could complete the risk analysis without using AI. However, the system is able to compute a large amount of data in a short amount of time and assist untrained or non-specialist police officers in risk analysis.

When a case of alleged gender-based domestic violence is reported, the police officer starts an initial investigation. This includes collecting evidence, taking witness statements and – potentially – making an arrest. Using information gathered from this process, the officer fills out two detailed questionnaires to assess the complaints, evaluate the probability of reoffending, examine the evolution of the case and assess the behaviour of the perpetrator and the victim. Police officers also indicate the level of gravity, the nature of threats faced and attitudes concerning the victim.

The system then produces a risk 'score' on a three point scale. The police officer can raise the level of risk manually, but cannot lower the risk level below that indicated by the system. Once the level is confirmed, specific measures are applied in line with established police protocols. The system also informs a judge about potentially 'severe cases' through an automated system.

FRA ACTIVITY

Detecting hate speech online

A public agency combatting hate crime uses an AI-based tool to detect online hate speech by analysing patterns of speech online. On the basis of the processing, the system determines which social groups are targeted. This helps law enforcement adopt measures to protect them before threats are realised.

Although the tool aims to identify potential victims, rather than perpetrators, law enforcement can use the information generated by the system to ask social media providers for information on users to pursue criminal investigations.

One particular challenge is understanding the context in which statements are made. For example, journalists or academics may use words associated with hate speech to report on or analyse its occurrence.

In 2021, FRA plans to initiate research on online hate present on social media. This will allow FRA to provide input to policy developments in the area of online content moderation, which uses AI.

2.2. EXAMPLES OF AI USE IN THE PRIVATE SECTOR

[Use case 3]

AI and health – analysing medical records to save lives

Background and EU legal framework

Healthcare is particularly prominent in discussions about the use of AI. Medical data and online applications have the potential to support improved health outcomes and – as a result – wider socio-economic benefits. The COVID-19 pandemic has further increased focus and interest in the area, particularly in terms of the potential for (online) data and applications to enhance the ability of governments and health services to track the spread of disease.

Health is also prominent in the general population's views on uses of AI. A 2019 Eurobarometer survey found that every second European thinks that AI can be best used to improve medical diagnostics, develop personalised medicine, or improve surgery.[22]

This use case covers applications of AI or related technologies by public and private sector stakeholders in the area of medical records and disease prediction. Feeding data from electronic medical records (EMR) and electronic health records (EHR) into AI systems and related technologies can support the development of preventative medicine that recognises early risks of disease and designs appropriate interventions. Researchers can predict clinical events such as mortality, hospitalisation, readmissions and length of stay in the hospital.

Beyond disease prediction, medical record data can be analysed to predict patients' adherence to treatment and their keeping of medical appointments. These technologies have the potential to support improved health outcomes, as well as increase the efficiency of the healthcare system.

Under Article 6 of the TFEU, the EU has supporting competence in protecting and improving human health. Member States retain full responsibility for defining their health policies, organising and managing their health systems, and for delivering health services (Article 168 (7) of the TFEU).

Within the EU competence, Union action, which has to complement national policies, is directed towards improving public health, preventing physical and mental illness and diseases, and obviating sources of danger to physical and mental health. Such action can cover health information and education, as well as monitoring, early warning of and combating serious cross-border threats to health (Article 168 (1) of the TFEU). In the latter areas, the EU can adopt incentive measures, excluding any harmonisation of the laws and regulations of the Member States.

Other rules and policies adopted at the EU level aim to ensure free movement of citizens, their equal treatment and non-discrimination abroad, as well as availability and safety of medical products and services in the single market. Considering the development of technologies and their application in health care, exchange of medical records, patients' rights in cross-border situations and disease prediction as a matter of public health are particularly relevant.

Under the GDPR, health and genetic data are considered as a special category of data (Article 9) called 'sensitive data'.[23] These require specific protection as their processing could create significant risks. Data subjects' health and genetic data can only be shared in specific circumstances under Article 9 (2) of the GDPR. The GDPR provides an exemption to the purpose limitation principle if data are used for research purposes, in line with its Article 89 (1). Researchers are required to ensure that technical and organisational safeguards – such as pseudonymisation and anonymity – are in place when using patient data.

The EU has also taken action regarding the exchange of medical records. European Commission Recommendation C(2019)800 on a European Electronic Health Record exchange format[24] "seeks to facilitate the cross-border interoperability of EHRs in the EU by supporting Members States in their efforts to ensure that citizens can securely access and exchange their health data wherever they are in the EU."[25] The recommendation lays out technical specifications for the exchange of such data between EU Member States.

The European Data Strategy (February 2020) also has a strong focus on health data.[26] A 'Common European health data space' is one of the nine common European data spaces whose establishment the European Commission will support.

The Early Warning and Response System (EWRS) is owned by the European Commission and operated by the European Centre for Disease Prevention and Control. It aims at "notifying at EU level on serious cross-border threats to health"[27] and enabling "the European Commission and EU countries to be in permanent communication for the purposes of alerting, assessing public health risks and determining the measures that may be required to protect public health."[28]

EMR, which is a computerised medical record created for patients of a healthcare organisation,[29] and EHR, which contains a patient's medical history beyond one organisation and involve sharing data across the healthcare system, can include a large amount of personal data. This can encompass, among others: the name and contact details of the individual and their next of kin; demographic information, diagnoses and test results; and medication and treatment.[30] They may also include patient-generated data from wearable devices.[31]

There is no uniform EMR/EHR system operating across all EU Member States.[32] Some, such as Germany, do not have a national EMR/EHR system. Others – including Belgium and Denmark – have different EMR/EHR systems at the regional level. The systems differ considerably depending on what data is recorded and by whom and who has access to what data.[33] The European Commission and other stakeholders have highlighted the diversity of country-level EMR/EHR systems and their lack of interoperability as a major barrier to the digital single market in health.[34]

Studies highlight the potential for AI or related technologies to enable earlier diagnosis, widen possibilities for disease prevention and improve patient safety,[35] strengthening the right to access preventive healthcare and benefit from medical treatment. EMR/EHR may also help to make healthcare more personalised,[36] while the possibility for rapid sharing of data can facilitate more coordinated and timely treatment.

However, use of EMR/EHR presents significant data protection risks. The healthcare sector leads in terms of personal data breaches.[37] The amount of the personal data stored, the highest among all industries, combined with the large data-sharing network and number of access points, makes the healthcare sector an attractive target for hackers.[38]

The quality of data in EMR/EHR also raises some concern. Studies where patients were shown their medical files and asked about their accuracy found that up to 50 % of information was incomplete or erroneous.[39] A lot of important data in EMR/EHR is unstructured in the form of free text, which further reduces data quality.[40] Low levels of accuracy, completeness and overall data quality increases the risk of medical error.[41]

Use in practice

The applications described in the interviews include both simple and more advanced models employed in the **public and private sectors**. The largest number of use cases refer to image-based diagnosis tools. However, interviewees also discussed tools to automate various working procedures, such as the mapping of text data, filing of medical records, and analyses and measurements of body tissues and nerve fibres.

A smaller number of examples touched on more advanced projects, such as systems to monitor remotely certain health indicators, such as heart rate. In each case, systems complement the expertise of health professionals. The next sections present examples of diagnostic and remote monitoring tools.

Image-based tools to help detect and diagnose disease

The tools used to support the detection and diagnosis of diseases described by interviewees work in similar ways. For example, a privately owned hospital uses an AI system to interpret images from CT-scans of stroke patients. After a stroke, imaging is used to detect where damage to the brain has occurred and where there may be blockages in the blood supply to the brain. It can also generate measures that can be compared to particular values by a medical specialist.

The interviewee feels that the application helps to determine such characteristics in images more quickly, potentially – depending on who uses the tool – improving the quality of the diagnosis. However, they highlight that it is not necessarily more efficient to rely on the AI application, since a medical professional must be present and they could examine the image. Rather, the tool can offer some support – for example, if the specialist finds it difficult to interpret a certain image or find abnormalities.

The system was built, trained and validated using a dataset partially based on a large scientific study to which the hospital contributed. This was supplemented by purchasing foreign datasets. The algorithm will not be further trained or adapted in the future based on new data. No new versions will be released.

The developers feel that allowing the system to continue to learn would make it difficult to validate its operation.

A **private company** developed an algorithm that supports the detection of breast cancer from mammography exams. The tool gives a probability and degree of certainty, which can help radiologists to speed up their analysis of the results and decide whether additional tests are warranted. The algorithm detects and characterises anomalies in a mammography as cancerous or not.

While the interviewee indicates that the system now has a very low rate of false negatives or false positives, they note that in many cases it does not deliver a clear outcome. The system was trained on radiography and mammography data from Europe and the EU, with written reports and past biopsies acting as control data.

Monitoring patients' vital statistics remotely

A **hospital** is piloting a system to support early detection of potential illness. Monitoring patients' health indicators – for example, blood pressure or heart rate – typically takes place manually and captures the situation at a specific moment in time. Constantly monitoring such indicators has the potential to identify trends that doctors may otherwise not recognise and detect health issues early to prevent illness. The system uses a biosensor – a kind of plaster – which gathers hemodynamic data from patients continuously by constantly monitoring heart pulsation and respiration.

The data used by the system come from the hospital and the patient. These data are anonymised before being shared with the third-party provider. No other information besides that gathered through the monitoring of the plaster is used to build and train the system. Data on environmental factors were not incorporated in the pilot because, the interviewee pointed out, they could contain biases.

In the future, the system will combine the information gathered by the biosensor with separate information from patients' EMR to draw conclusions from trends observed in the monitoring.

Using AI to target health inspections

A public authority responsible for inspecting food safety standards in restaurants uses machine learning to process customer review data from major online platforms. This helps to decide where and when to conduct inspections. Previously, this process was based on complaints the authority received and on previous reports. Since the introduction of the tool, the rate of non-compliant restaurants identified doubled from around 18 % to 36 %.

The first step involves text mining. The algorithm identifies reviews containing key words that may indicate health and safety issues, such as 'sick, 'nausea' or 'rodents'. For the second step, the authority compared results coming from customer reviews with previous inspection reports to improve the algorithm's accuracy and reliability.

[Use case 4]

Targeted advertising – profiling consumers to boost profit

Background and EU legal framework

The internet has transformed the way we live. Many people make use of internet services, often offered for free, on a daily basis. Companies offering their services for free mainly generate revenue through advertising, with adverts automatically targeted to individual consumers based on information about them.

The availability of data about online individual behaviour combined with machine learning technologies have considerably improved the ability of commercial enterprises to target individuals. This could even go as far as manipulating consumers by predicting their reactions based on irrational aspects of psychology and not reasoned choice.[42]

The Cambridge Analytica scandal underscored the particularly negative impact of such uses for political purposes. In that case, a company illegally obtained personal data on millions of social media users to target political adverts to different social groups based on certain psychological profiles.[43]

A recent declaration of the Committee of Ministers of the Council of Europe highlights the lack of knowledge about the manipulative power of algorithms. "The effects of the targeted use of constantly expanding volumes of aggregated data on the exercise of human rights in a broader sense, significantly beyond the current notions of personal data protection and privacy, remain understudied and require serious consideration."[44] Concerns have also been raised about how online advertising, powered by AI technologies, can affect data protection and privacy,[45] consumer protection,[46] the right to non-discrimination,[47] and even the way democracies work.[48]

The word 'advertising' is associated with messages designed to influence consumer behaviour. Advertising in one form or another has always targeted specific groups based on their characteristics and behaviour.[49]

The growth of social media, however, has taken targeted advertising to another level, using direct access to consumer data. Micro-targeting is directed towards very specific groups – and the more data that is gathered through online activities, the more targeted these activities can be. As social media providers and platforms like Google or Amazon gather comprehensive user data by monitoring the various activities of their users, advertisers can access more detailed and specific information.[50]

The area of targeted advertising and systems that recommend content (e.g. news or movies) is one of the few real life examples that also involves so-called reinforcement learning. This is a technology that is based on optimising a certain goal through experimenting and updating its rules automatically to have the best possible output. This means a systems tries out different ad placements through trial and error and so finds the best way to optimise revenue – including an element of self-learning.

While very little knowledge on the actual use of reinforcement learning is available for European countries, major companies working in the area are researching the issue.[51]

Issues related to targeted advertising fall under consumer protection. This falls under shared EU competence with Member States under Article 4 (2) (f) of the TFEU. EU consumer protection measures seek to protect the health, safety and economic interests of consumers; and promote their right to information, education and to organise themselves to safeguard their interests (Article 169 (1) of the TFEU). The EU can adopt minimum harmonisation measures to achieve a high level of consumer protection (Article 114 (3) of the TFEU), yet allowing EU Member States to introduce even more stringent measures nationally.

In secondary EU legislation, rules on advertising are covered by Directive 2006/114/EC concerning misleading and comparative advertising.[52] This directive provides a minimum level of protection from misleading advertising. It also harmonises rules on comparative advertising across the EU. The provisions of Directive 2006/114/EC apply to both consumer-to-business and business-to-business relations. However, they are practically only applied to the latter[53] after Directive 2005/29/EC on unfair business-to-consumer commercial practices in the internal market practices[54] took effect.

Further, Directive 2006/123/EC on services in the internal market[55] covers services that include advertising. Additionally, Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the internal market (E-Commerce Directive) also applies. This directive forms part of the legal framework for digital services in the EU. To meet significant developments in the area of new online services and practices, the E-Commerce Directive is currently being revised as part of the Digital Services Act package. That package aims to "strengthen the Single Market for digital services and foster innovation and competitiveness of the European online environment".[56]

FRA collected information about actual use cases from six European companies engaged in placing online ads, content recommendation, and personalised marketing.

Use in practice

The examples covered include:

- placing ads online based on click predictions (i.e. learning about the likelihood that online users click on certain links or adverts) and automated bidding at auctions for online advertisement space
- personalised and targeted marketing and communication via email.

Most tasks were fully automated. The examples concern analyses of user preferences and activity and calculations of probabilities of clicks and purchases, including a measurement of the effectiveness of previously made recommendations. This also includes methods of targeted communication on the basis of identified target groups to build (long term) trust between clients and service providers.

Targeted online ads based on click predictions

Business models working with click predictions and targeted advertisements often follow a 'click and buy' policy. **Companies** purchase advertising space on media platforms, and optimise the display of adverts by analysing the interests and preferences of website users and showing them advertisements that interest them. The purpose is to increase the relevance of advertisements shown by better matching them to the interests of those who see them.

In the present example, the company only gets paid if people click on an advertisement and buy something. Additionally, the company uses AI to detect inappropriate content in advertising, such as advertisements for alcohol, firearms or political content.

The company uses a range of machine learning techniques in the field of computational advertising. To estimate the probability of a user clicking on an advertisement displayed in a specific context (optimising the so-called click-through-rate), customers' interests and the relevance of products are measured via a mapping of individuals' browsing histories and transaction patterns. Further, information is derived from individuals' navigation on merchant websites worldwide, with whom the advertising company works. This is done via anonymised third party cookies and trackers.These are placed on these merchant websites and outline individuals' navigation across them, and also list the products seen and purchased.

The profiles of individuals are linked to devices used by them, although IP addresses are anonymised. Once a product has been purchased, a recommender system algorithm tries to determine other products that the customer could also buy. In this case, 'fresh' data is valued higher than older data. Browsing histories are stored for a maximum of one year, as interests change and purchases older than a year are no longer necessarily considered relevant.

Advertisements shown to the respective person are immediately adapted accordingly, and they vary across websites, to also match the content of the latter. Once an advertisement is posted, it is continuously analysed. The combination of elements taken into account on an individual's interest is confirmed when a purchase is made. Data is shared across platforms, which includes informing others once a purchase has been made, to stop advertisements of that particular item. If no purchase is made, the formula is reviewed and the algorithm is further adapted on individuals' continuous online behaviour.

In the future, the company covered in this example expects to work more on optimising its timing in terms of when it places advertisements, within their given budget for a certain time frame. It also expects to focus more on displayed ads that have an impact on consumers.

Another example is based on a European online market place, which links buyers and sellers on a range of specialised products. Here, AI is used to optimise advertising campaigns, to categorise products based on the advertisements that are shown on the website of the market place, to improve the search engine experience by predicting complementary and substitutable products, and to detect fraud attempts.

The company uses machine learning to predict the value of clicks of customers to buy advertisement space, which is offered in real-time auctions. With these examples, the company indicates that AI enables it to make decisions that otherwise would not be possible without AI, or which would have to be significantly scaled down.

Targeted communication with customers and clients

In the case of a retail **company** focusing on specialised supplies sold across physical stores and online, direct marketing or personalised advertising is used to increase appeal to customers, and at the same time measure the efficiency of a particular instance of marketing or advertising.

According to the company at issue in this example, marketing emails are opened at an average of 20-30 %, and particularly so when customers recognise relevant and favourite products being offered. Marketing emails are sent to around 250,000 registered individuals, and a system is used to establish what may be considered relevant by each of these individuals. This is done by analysing purchases made by the respective individuals in the previous six months. 80 % of the offers displayed are directly based on previous purchases. Meanwhile, 20 % are new suggestions, i.e. alternative products in the same category as the previous purchases.

A similar approach is used by a bank that sends emails to clients. Messages offering specific services or products are sent only to certain clients. Data analysts calculate the probability of clients being interested in a service or product. If this probability is above a certain threshold, the client will receive the message. The system used does not yet include machine learning models and is not fully automated. These points will be taken on when they further develop the system.

In a third example, a grocery retailer uses loyalty cards both to increase the customers' interaction and to personalise offers. Loyalty card systems can predict how many customers are likely to engage with a product offering. The system covered in this example also suggests new products to customers and tracks the results of these suggestions. It groups buyers with similar behavioural patterns into segments to make more personalised suggestions.

Every week, the company's loyalty card owners receive personalised offers by email, website or mobile application, and they can access offers through in-store terminals. The AI system selects the offerings based on the individual purchase history, and it recommends new items that might catch the buyer's interest and prompt a purchase.

Endnotes

[1] See, for instance, Samoili et al. (2020), **AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence**, Luxembourg; Karanasiou A. and Pinotsis D. (2017), '**A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning'**, *International Review of Law, Computers & Technology*, 2017, pp. 170-187.

[2] See FRA (2019), ***Facial recognition technology: fundamental rights considerations in the context of law enforcement***, Luxembourg, Publications Office, p. 9 and 22.

[3] FRA (2019), ***Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights***, Luxembourg, Publications Office, June 2019.

[4] UN, Human Rights Council (2019), *Report of the Special rapporteur on extreme poverty and human rights*, Philip Alston, A/74/48037.

[5] Eubanks, V. (2018), *Automating Inequality. How hightech tools profile, police, and punish the poor*, St. Martin's Press.

[6] Redden, Joanna, Dencik, Lina and Warne, Harry (2020), ***Datafied child welfare services: unpacking politics, economics and power***, Policy Studies.

[7] Panoptykon Foundation (2015), **Profiling the unemployed in poland: Social and political implications of algorithmic decision making**; see also Algorithm Watch (2019), **Poland: Government to scrap controversial unemployment scoring system**.

[8] OECD, Glossary of Statistical Terms - **Social Benefits Definition**, accessed 5 August 2020.

[9] J. Henry Richardson, ***CHAPTER IV, SOCIAL INSURANCE***, *Economic and Financial Aspects of Social Security* (University of Toronto Press, 1960). Pieters, *Social Security*.

[10] For an overview of the EU competence in this domain and Regulation (EC) No. 883/2004, see Paju, J. (2017), *The European Union and Social Security Law*, Oxford, Hart Publishing, Ch. 2.

[11] Erik Bakke (2018), 'Predictive policing: The argument for public transparency', *New York University Annual Survey of American Law*, Vol. 74, pp. 139-140; Andrew G. Ferguson (2017), '**Policing Predictive Policing**', *Washington University Law Review*, Vol. 94, pp. 1146-1150. For example, only one in five women who experienced violence brought the most serious incident to the attention of the police. See FRA (2014), ***Violence against women: an EU-wide survey. Main results report***, Luxembourg, Publications Office, p. 61.

[12] Elizabeth E. Joh (2015), **The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing**, UC Davis Legal Studies Research Paper No. 473, p.18.

[13] Braga A., et al (2019), **Hot spots policing of small geographic areas effects on crime**, *Campbell Systematic Reviews*, Vol. 15 (3).

[14] Elizabeth E. Joh (2015). The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing, UC Davis Legal Studies Research Paper No. 473, pp. 17-18. Available at: **SSRN**.

[15] Erik Bakke (2018), 'Predictive policing: The argument for public transparency', *New York University Annual Survey of American Law*, Vol. 74, pp. 137-138.

[16] Wim Hardyns and Anneleen Rummens (2017), 'Predictive Policing as a New Tool for Law Enforcement? Recent Developments and Challenges', *Eur J Crim Policy Res*, p. 3, DOI: 10.1007/s10610-017-9361-2.

[17] Albert Meijer & Martijn Wessels (2019), 'Predictive Policing: Review of Benefits and Drawbacks', I*nternational Journal of Public Administration* 42:12, p. 1032, DOI: 10.1080/01900692.2019.1575664.

[18] The Law Society Commission on the Use of Algorithm in the Justice System (2019), **Algorithms in the criminal justice system**, p. 36.

[19] Newbold, J. (N.D.), **'Predictive Policing', 'Preventative Policing' or 'Intelligence Led Policing'. What is the future?**

[20] Declaration No. 21 Annexed to the Final Act of the Intergovernmental Conference which adopted the Treaty of Lisbon, signed on 13 December 2007.

[21] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119, 4.5.2016, pp. 89-131.

[22] European Commission (2019), Standard Eurobarometer 92, Report, *Europeans and Artificial Intelligence*, p. 10.

[23] European Patients Forum (n.d.), **The new EU Regulation on the protection of personal data: what does it mean for patients? A guide for patients and patients' organisations**.

[24] Commission Recommendation (EU) 2019/243 of 6 February 2019 on a European Electronic Health Record exchange format, OJ L 39, 11.2.2019, pp. 18-27.

[25] Digital Health Society, ***Exchange of electronic health records across the EU***, 19 February 2020.

[26] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, **A European Strategy for data**, COM(2020)66 final, Brussels, 19 February 2020.

[27] Decision No. 1082/2013/EU of the European Parliament and of the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No. 2119/98/EC, OJ L 293, 5.11.2013, pp. 1-15.

[28] See Commission **webpage on Communicable diseases**.

[29] OECD and European Union (2018), **Healthcare at a glance: Europe 2018**, p. 192.

[30] Vera Ehrenstein, Hadi Kharrazi, Harold Lehmann and Casey Overby Taylor (2019), 'Obtaining Data From Electronic Health Records', in: Gliklich RE, Leavy MB, Dreyer NA (eds.), **Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide**, 3rd ed., Addendum 2.

[31] On the use of these data in the insurance industry currently and their potential see, for example, A. Spender, C. Bullen, L. Altmann-Richer, J. Cripps, R. Duffy, C. Falkous, M. Farrell, T. Horn, J. Wigzell and W. Yeap (2019), '**Wearables and the internet of things: considerations for the life and health insurance industry'**, *British Actuarial Journal* 24:22, pp. 1-31.

[32] See **WHO visualisation**.

[33] See a short overview of different EHR systems in Europe from nurses' perspective in HealthEurope (2019), **The world of cloud-based services: storing health data in the cloud**.

[34] College of Europe (2018), **Transformation Health and Care in the Digital Single Market. Synopsis report of the public consultation**.

[35] European Commission (2016), **Study on Big Data in public health, telemedine and healthcare**; Roberta Pastorino, Corrado De Vito, Giuseppe Migliara, Katrin Glocker, Ilona Binenbaum, Walter Ricciardi, Stefania Boccia (2019), **Benefits and challenges of Big Data in healthcare: an overview of the European initiatives**, *European Journal of Public Health*, Vol. 29, Issue Supplement 3, pp. 23–27.

[36] Ministry of Health, Welfare and Sport of The Netherlands (2016), **Digitalization in health care and benefits for patient safety: Literature and web reports (2015-2016)**.

[37] This is according to multiple reports by different cybersecurity companies and over time. See, for example, SC Magazine (2019), **Healthcare leads in cost of data breaches**; Shannon Williams (2020), **New report reveals 'wall of shame' in health care data breaches**; Tammy Lovell (2019), **Statistics reveal healthcare is the sector most affected by personal data breaches**.

[38] SC Magazine (2019), **Healthcare leads in cost of data breaches**.

[39] Annet Sollie (2016), **Reuse and Sharing of Electronic Health Record Data with a focus on Primary Care and Disease Coding**, Doctoral dissertation at the Vrije Univesiteit Amsterdam, pp. 28-30.

[40] Vera Ehrenstein, Hadi Kharrazi, Harold Lehmann and Casey Overby Taylor (2019), 'Obtaining Data From Electronic Health Records', in: Gliklich RE, Leavy MB, Dreyer NA (eds.), **Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide**, 3rd ed., Addendum 2.

[41] Mowafa Househ, Bakheet Aldosari, Abdullah Alanazi Show, Andre Kushniruk and Elizabeth M Borycki (2017), 'Big Data, Big Problems: A Healthcare Perspective', *Studies in health technology and informatics* 238, p. 38.

[42] Sartor, Giovanni (2020), **New aspects and challenges in consumer protection**, study for the committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg.

[43] Neudert, Lisa and Marchal Nahema (2019), **Polarisation and the use of technology in in political campaigns and communication**, study at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit, within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament; Information Commissioner's Office (ICO) (2018), **Investigation into the use of data analytics in political campaigns**.

[44] Council of Europe (2019), **Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes**, Decl(13/02/2019)1.

[45] For example, Costello, Róisín Áine (2020), **The Impacts of AdTech on Privacy Rights and the Rule of Law**, Technology and Regulation, 11–23; EDPS (2018), **Opinion 3/2018, EDPS Opinion on online manipulation and personal data**.

[46] Sartor, Giovanni (2020), **New aspects and challenges in consumer protection**; Jabłonowska, Agnieszka et al. (2018), '**Consumer law and artificial intelligence. Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence**' *EUI Working Papers, LAW 2018/11*.

[47] Wachter, Sandra (2020), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising', *Berkeley Technology Law Journal*, Vol. 35, No. 2, 2020, (forthcoming), available at **SSRN**.

[48] Zuboff, Shoshana (2018), *The Age of Surveillance Capitalism*, London; EDPS (2018), **Opinion 3/2018, EDPS Opinion on online manipulation and personal data**.

[49] Martin, Gillian (2011), The importance of marketing segmentation, *American Journal of Business Education*, Vol. 4, No. 6.

[50] Kaili Lambe and Becca Ricks (2020),**The basics on microtargeting and political ads on Facebook**.

[51] See for example, information on the **RecSys2020 Workshop on REVEAL 2020: Bandit and Reinforcement Learning from User Interactions** (accessed on 7 August 2020).

[52] Directive 2006/114/EC of the European Parliament and of the Council of 12 December 2006 concerning misleading and comparative advertising, OJ L 376, 27.12.2006, pp. 21-27.

[53] European Commission, **Misleading and comparative advertising directive: Objective of the directive**.

[54] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149, 11.6.2005, pp. 22-39.

[55] Directive 2006/114/EC of the European Parliament and of the Council of 12 December 2006 concerning misleading and comparative advertising, OJ L 376, 27.12.2006.

[56] See the European Commission's webpage on **The Digital Services Act package**.

3

FUNDAMENTAL RIGHTS FRAMEWORK APPLICABLE TO AI

The use of AI – as presented in the four use cases discussed in **Chapter 2** – can affect specific fundamental rights (as outlined in **Chapter 4**). Full compliance with fundamental rights is a prerequisite for using AI-driven technologies, irrespective of the area concerned.

This chapter introduces the general fundamental rights framework in the EU that governs the use of AI, including selected secondary EU legislation and national law (Section 3.1). This fundamental rights framework provides the normative basis and benchmarks for the design, development and deployment of AI tools.[1] It helps determine whether or not a specific use of AI is fundamental rights compliant. The requirements for justified interferences with fundamental rights are outlined in **Section 3.3.**

3.1. FUNDAMENTAL RIGHTS FRAMEWORK GOVERNING THE USE OF AI

The cornerstone instrument of the EU fundamental rights framework applicable to the use of AI is the Charter. Together with the unwritten general principles of EU law, it is the main source of fundamental rights in the EU. The Charter enshrines a wide array of fundamental rights and has the same legal value as the EU Treaties. All EU institutions and bodies are bound by the Charter, as are Member States when they act within the scope of EU law (Article 51 (1) of the Charter).[2]

Many Charter rights are the same as those set out in the European Convention on Human Rights (ECHR).[3] Their meaning and scope must be the same as the corresponding ECHR rights (Article 52 (3) of the Charter). However, this cannot prevent Union law from providing more extensive protection.

Fundamental rights can also be found in provisions of the Treaties (see e.g. Article 6 (2) of the TEU and Titles V and X of the TFEU), and in EU secondary law.[4] These rights are further safeguarded in different pieces of secondary EU law.

A central piece of EU secondary law in the context of AI is the General Data Protection Regulation (GDPR – Regulation (EU) 2016/679).[5] It governs automated processing of personal data in the European Economic Area and processing of personal data by any other means which form part of a filing system – within the scope of EU law. (As a result, the GDPR does not apply to national security-related data processing.)

The GDPR is coupled with the Law Enforcement Directive, which applies to police and judicial cooperation in criminal matters. Both EU instruments include numerous provisions on the protection of personal data, determining the key principles of data processing, such as lawfulness, fairness and transparency.[6]

Whether EU data protection legislation applies depends on whether personal data are processed. Some AI-driven applications do not use personal data (for example, traffic data). Others use anonymised data. In these cases, data protection laws do not apply, or their applicability is not entirely clear.[7] The line between personal and non-personal data is blurred, because there is some risk that anonymised data can be 're-identified' – ie, the anonymisation can be undone. However, re-identification is usually illegal. In addition, persons re-identifying the data usually have to put in major efforts and potentially need access to additional information about individuals who might be included in an anonymised dataset for re-identification. **Section 4.2** discusses the topic in more detail, linked to the results of the interviews carried out for this report.

In addition to the EU data protection acquis, European non-discrimination law is key for safeguarding fundamental rights in the context of the use of AI and related technologies. Article 2 of the TEU provides that non-discrimination is one of the fundamental values of the EU, and Article 10 of the TFEU requires the Union to combat discrimination on a number of grounds. Moreover, Articles 20 and 21 of the Charter provide for equality before the law and non-discrimination.

Beyond this, several EU non-discrimination directives enshrine more specific and detailed provisions. They have varying scopes of application.[8] These include the Employment Equality Directive (2000/78/EC),[9] the Racial Equality Directive (2000/43/EC),[10] the Gender Goods and Services Directive (2004/113/EC),[11] and the recast Gender Equality Directive (2006/54/EC).[12]

EU Member States are also party to other international human rights conventions (see the list of conventions in the **Key Findings and FRA opinions** section). These contain legally binding standards and safeguards to comply with when they act in areas that do not fall within the scope of EU competence. The main such instrument is the ECHR, ratified by all EU Member States. It is accompanied by additional protocols, to which a great majority of EU Member States are parties. The ECHR has a wide reach: it also applies to areas not covered by EU law.

In addition, the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data[13] is another source of pan-European data protection obligations binding on all EU Member States. It was recently modernised.[14]

Sector-specific EU and national legislation also enshrines safeguards for the protection of fundamental rights. An overview of such more technical legislation is beyond the scope of this report. However, this chapter provides a few examples relevant to the use cases discussed in the report. This is complemented by a couple of examples of national laws from the five EU Member States covered.

None of the five EU Member States covered currently have horizontal AI-specific laws, although the countries are looking into the potential need for regulation. Some EU countries, such as Finland, issued recommendations for self-regulation and the development of responsibility standards for the private sector.[15] In Estonia, an assessment concluded that a separate AI-specific law will not be required in the foreseeable future, since the current legal framework is sufficient.[16] According to the relevant Estonian long-term strategy, however, the legal environment must be adapted to avoid unnecessary hindrances to implementing AI.[17]

The situation concerning sectoral legislation relevant to the use of AI in different sectors varies across EU Member States. However, active policymaking on AI has recently emerged at the national level. National action plans on AI have appeared and remain the core policy development in Member States. Some countries are working on growing entrepreneurship.[18] Others are focused on enacting market-oriented policies compatible with the UN 2030 Agenda for Sustainable Development.[19] Educational activities to promote AI and increasing public use of AI are often identified as AI-related strategy goals. Investment in research and development is also frequently outlined as a relevant goal.[20]

While domestic AI discussions on potential legislative reforms remain attentive to European initiatives, national, sector-specific fundamental rights safeguards are also being enacted. For instance, Finland began considering an overhaul of domestic human rights safeguards in the public sector by proposing a broader, across-the-board legislative update as opposed to individual AI laws.

In specific reference to the processing of personal data under immigration law, the Finnish Constitutional Law Committee has put forward a proposal to strengthen the safeguards of the Finnish Constitution, overriding constitutional law shortcomings in relation to, among others, protection under the law, accountability, as well the ambiguity of algorithms in automated decision making. Whenever public authorities automate their decision-making processes, these processes must adhere to the constitutional principle of rule of law, and may not endanger the observance of rules on good administration and due process.[21] This proposal articulated a vision on what requirements the Finnish Constitution sets for AI use and automated decision making within public administration.

The research identified other initiatives and policies linked to AI and fundamental rights in the five Member States examined. For example, the Estonian e-State charter includes a summary of citizens' rights for better communicating with agencies electronically. It also targets AI in relation to the right to know what data is collected by public authorities.[22]

Similarly, the Ministry of the Interior of the Netherlands presented a policy brief to parliament on AI, public values and fundamental rights.[23] The brief stresses a human-centric approach, where AI-applications have a strong influence on human beings or on society as a whole. It also lists the most important risks of AI for fundamental rights, such as discrimination as a result of biased data, or reduced interpersonal relations if AI takes over certain forms of interaction.

3.2. ‘USE CASE’ EXAMPLES

Social welfare (Use case 1)

When regulating social welfare, EU Member States enacted rules aiming to protect fundamental rights specifically in this area in addition to existing horizontal EU regulations (see **Section 2.1**). These mostly define rules for the processing and protection of personal data for the purpose of social benefits and insurance.

In Estonia, for example, the Insurance Activities Act, applicable to all types and forms of insurance, regulates the processing and transmission of personal data in this context. It states that public authorities, health care providers, insurance undertakings and other third parties may transmit personal data at the request of an insurance undertaking if the personal health or court data are necessary for the insurance undertaking to perform an insurance contract or if the right and obligation to disclose such data derives from law. The scope of this Act also includes data transfers for the purpose of data processing within AI systems.

The Social Welfare Act contains more specific provisions on data protection of persons in need of social assistance. They have to be notified of the processing of their data and should provide consent for further processing. Any person in the established target group has the right to opt out of data processing. The Social Welfare Act also allows local authorities to process (including using algorithms) personal data of youth between 16 and 26 years of age stored in state registries to identify the youth not in employment, education or training.

In Finland, Act No. 552/2019 on Secondary Use of Health and Social Data applies to using AI in social care and healthcare. This Act is based on the norms for securing and protecting sensitive personal data as outlined in the GDPR. It aims to establish conditions for the effective and secure “processing of, and access to, personal health and social data for certain secondary purposes, such as research and statistics, innovation and development, knowledge management, teaching and authority planning.”[24] The Act regulates the manner in which registered health data can and cannot be processed.

Several other laws apply to various types of social benefits. In France, the 2015 Code of relations between the public and the administration applies for the purpose of processing or accessing personal data related to social benefits with minor amendments after the entry into force of the GDPR. This code states “that algorithms used by public administrations must be published” and “the person subject to automated decision making has a right to be informed”.[25]

Predictive policing (Use case 2)

In the context of predictive policing, the EU’s Law Enforcement Directive contains key fundamental rights safeguards. These stipulate how law enforcement authorities should apply some of the main data protection principles set out in the GDPR.[26] These include the requirement for data controllers (i.e. the competent law enforcement authorities) to provide data subjects with information on the controller’s data processing activities, such as the identity and contact details of the data controller, the purposes of the processing and information about the right to lodge a complaint (Article 13).

In specific cases, data controllers shall provide further information – for example, the legal basis for processing – to enable data subjects to exercise their rights. The right of access (Article 14) requires the data controller to confirm, upon request of the data subject, whether there are processing operations related to them. If this is the case, the data subject shall be able to access this data and also to request additional information, including the purposes and legal basis of the processing and the categories of personal data processed. Both the right to information and the right to access can be restricted in a number of cases, including to avoid obstructing or prejudicing the prevention, detection, investigation or prosecution of criminal offences; or to protect public security and national security.[27]

In addition, Article 11 of the Law Enforcement Directive explicitly prohibits automated decision making.[28] This prohibition is limited if authorised by EU or national law which safeguards the data subject's rights, including "at least the right to obtain human intervention on the part of the controller" (for more, see **Section 4.2**).

In some cases, the scope of implementing national legislation is broader than the directive. For example, the Finnish Act on the Processing of Personal Data in Criminal Matters and in Connection with Maintaining National Security strengthens the right to information by not distinguishing between the information provided in general and in special circumstances.[29]

Healthcare (Use case 3)

As regards EU-level fundamental rights safeguards when using AI in healthcare, the GDPR empowers patients with rights to be informed, in part by granting them more control of their personal health data. Such data qualifies as 'sensitive data'[30], as found, for example, in their medical records.[31] The rights include the rights to access one's own personal (health) data; to object to the processing of own personal data; rectification and erasure of data, as well as rights in case of breach.[.32]

Under the GDPR, administrative fines for breaches of processing data, including health data, are not allowed. However, in Estonia, for instance, domestic law allows for a maximum penalty of EUR 400,000 in application of the misdemeanour procedure in such cases. The Data Protection Inspectorate can also impose similar fines in the misdemeanour procedure.[33]

In France, the Data Protection Act and the Public Health Code impose stricter requirements than those set out in the GDPR regarding health data processing. The French Data Protection Act has been amended through the Law for the Modernisation of the Health System, to allow for the processing of personal health data for various purposes, provided they fall within the scope of one of the exceptions to the general principle of prohibition of sensitive data processing under Article 9 of the GDPR.[34]

Targeted advertising (Use case 4)

When considering fundamental rights safeguards in relation to targeted advertising and the underlying mechanisms regarding profiling in particular, the EU legal framework on privacy and data protection provides the most relevant fundamental rights provisions. The protection of privacy and personal data holds a status that takes precedence over economic benefits. Hence, rules on processing of (special categories of) personal data are relevant for companies operating in the area of or applying targeted advertising in that they place companies under certain obligations.

The main legal provisions setting out rules on protecting personal data in the EU are the GDPR and the Directive on privacy and electronic communications (e-Privacy Directive), which is a *lex specialis* to the GDPR. The GDPR is directly applicable in all EU Member States whenever a company is based in the EU and processes personal data, and if a company is based outside of the EU, but processes data relating to individuals in the EU.

The e-Privacy Directive, with a strong focus on fundamental rights, concerns the processing of personal data and the protection of privacy in the electronic communications sector (e.g. when individuals use their computer, smartphone or tablet). In 2017, the European Commission proposed an e-Privacy Regulation, which would replace the current e-Privacy Directive.[35] The legislative proposal would broaden the scope of the directive, and include specific provisions concerning unsolicited marketing, cookies and confidentiality.

3.3. REQUIREMENTS FOR JUSTIFIED INTERFERENCES WITH FUNDAMENTAL RIGHTS

Chapter 4 highlights selected fundamental rights – as covered by the Charter – that are particularly affected by AI, taking into account the four use cases discussed in Chapter 2. Most of these rights are not absolute rights, so can be subject to limitations in line with Article 52 (1) of the Charter. Accordingly, before analysing to what extent the different fundamental rights are impacted by the use of AI, this section presents the general steps that need to be followed to determine whether or not a Charter right can be limited.

Fundamental rights affected by AI that are not absolute can be subject to limitations. Interferences with such fundamental rights can only be justified if they respect the requirements of the Charter and of the ECHR, in case of Charter rights corresponding to rights guaranteed in the ECHR (Article 52 (3) of the Charter).[36]

Pursuant to Article 52 (1) of the Charter, any limitation on fundamental rights must:

- be provided for by law,
- genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others,
- respect the essence of the right,
- be necessary, and
- be proportionate.[37]

The Court of Justice of the EU (CJEU) has also emphasised that any limitation on the exercise of the rights and freedoms recognised by in the Charter must respect "the essence" of those rights and freedoms.[38] This means that fundamental rights can be limited to a certain extent, but not completely disregarded.

Once it has been established that the inalienable, essential core of a right is not violated by a measure, the next step is to conduct the necessity and proportionality test outlined in the Charter in respect of non-core aspects of that right.[39] Any interference with a Charter right needs to be examined as to whether the given legitimate aim could not be obtained by other means that interfere less with the right guaranteed.[40] Similar requirements are also imposed by the ECHR, as interpreted by the European Court of Human Rights (ECtHR).[41] These include the 'essence of a right' concept, which can be derived from the object and purpose of the ECHR as a whole.[42] In respect to the use of new technologies, the ECtHR observed in *S. and Marper v. the UK* that States should "strike a right balance" between protecting fundamental rights and developing new technologies.[43]

Given the wide range of applications of AI systems in everyday life as presented in the four selected use cases, a wide range fundamental rights may have to be assessed, taking into account a variety of elements, depending on the context and the particular area of use. Most notably, the specific purpose for which AI is used, its functionality, complexity, and the scale at which it is deployed, are relevant for assessing fundamental rights implications.[44]

Endnotes

[1] See also van Veen, C. (2018), ‘**Artificial Intelligence; What’s Human Rights Got to Do with It?**’ *Data & Society: Points – blog of Data & Society Research Institute*, 14 May 2018; Barfield, W. & Pagallo, U. (2020), *Advanced Introduction to Law and Artificial Intelligence*, Cheltenham/Northhampton, MA, Edward Elgar, 2020, pp. 19-20.

[2] See also CJEU, Åklagaren v. Hans Åkerberg Fransson [GC], 26 February 2013, paras. 17, 20.

[3] **European Convention for the Protection of Human Rights and Fundamental Freedoms**, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5.

[4] For an overview of the application of the Charter, see FRA (2018a), ***Applying the Charter of Fundamental Rights of the European Union in law and policy making at national level***, Luxembourg, Publications Office.

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (**General Data Protection Regulation**), OJ L 119, 4.5.2016, pp. 1-88.

[6] For more, see FRA (2018), ***Handbook on European Data Protection Law***. *2018 Edition*, Luxembourg, Publications Office.

[7] See for example, Hacker, P. (2020), *A Legal Framework for AI Training Data. Law, Innovation and Technology* (forthcoming), available at **SSRN**.

[8] For an overview of European non-discrimination law, see FRA (2018), ***Handbook on European non-discrimination law***. *2018 Edition*, Luxembourg, Publications Office.

[9] Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation, OJ L 303, 2.12.2000, pp. 16-22.

[10] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180, 19.7.2000, pp. 22-26.

[11] Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, OJ L 373, 21.12.2004, pp. 37-43.

[12] Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast), OJ L 204, 26.7.2006, pp. 23-36.

[13] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Strasbourg, 28 January 1981 (ETS No. 108).

[14] Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Strasbourg, 10 October 2018 (CETS No. 223).

[15] The AI Finland Project’s ethics working group and the Ethics Challenge added emphasis on companies and self-regulation. AI Finland, ‘***Etiikkahaaste*** **(Ethics Challenge)’,** ***Tekoäly on uusi sähkö*** (in Finnish).

[16] Republic of Estonia (2019), **Report of Estonia’s AI Taskforce**, p. 38.

[17] The Estonian Government **launched the preparation** for a long-term strategy.

[18] For example, see the Netherlands, Ministry of Economic Affairs and Climate Policy (2019), **Strategic Action on AI Strategic Action Plan AI** (*Strategisch Actieplan AI* – SAPAI).

[19] For an example of an effort to adapt goals to the development of a sustainable market, see Spain, Ministry of Science, Innovation and Universities (2019), **National AI Strategy** (in Spanish).

[20] For a more comprehensive overview, see European Commission (2019), **National strategies on Artificial Intelligence**; or the OECD **AI policy observatory**.

[21] Finnish Constitutional Law Committee (2019), ‘Committee Opinion PeVL 7/2019 Vp — HE 18/2019 vp: Draft Proposal to Parliament for the Law on the Processing of Personal Data in the Immigration Administration and for Related Laws’.

[22] Estonia, National Audit Office and Chancellor of Justice (2018), **Everyone’s Rights in e-State: The e-State Charter**.

[23] Netherlands, Ministry of the Interior and Kingdom Relations (2019) **AI, public values and fundamental rights** (in Dutch).

[24] Elina Saxlin-Hautamäki and Johanna Lilja (2019), **Secondary use of health data – the new Finnish Act**.

[25] de Donno, M. (2017), **The French Code “Des Relations Entre Le Public Et L’Administration”. A New European Era For Administrative Procedure?**, *Italian Journal of Public Law* 2, pp. 220-260.

[26] See FRA (2018), ***Preventing unlawful profiling today and in the future: a guide***, Luxembourg, Publications Office, Tables 2 and 4.

[27] Sajfert, J. and Quintel, T. (2017), **Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities**, available at **SSRN**.

[28] Note that Article 11 of the Law Enforcement Directive seems to apply to automated decisions taken solely through automated processing. This means that this safeguard will not apply if human agency is involved. Orla Lynskey (2019), **Criminal justice profiling and EU data protection law: Precarious protection from predictive policing**, p. 21.

[29] The English translation is available via the **Finlex website**.

[30] GDPR, recital (10) and Art. 9 (1).

[31] European Patients Forum (n.d.), ***The new EU Regulation on the protection of personal data: what does it mean for patients? A guide for patients and patients’ organisations***.

[32] GDPR, Arts. 15-17, 20-21 and 34.

[33] White&Case (2019), **GDPR Guide to National Implementation: Estonia**.

[34] Merav Griguer (2019), **Processing health data in France: What to look out for after GDPR?**

[35] European Commission, **Proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)**, COM(2017) 10 final, Brussels, 10.1.2017.

[36] Charter, Art. 52 (3): “In so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention.”

[37] As also reiterated and explained by the CJEU. See, for example, C-73/07, *Satakunnan Markkinapörssi and Satamedia*, 16 December 2008, para. 56; Joined cases C-92/09 and C-93/09, *Volker und Markus Schecke and Eifert GbR and Hartmut Eifert*, 9 November 2010, para. 77; Joined cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others*, 8 April 2014, para. 52; C-362/14, *Maximillian Schrems v. Data Protection Commissioner*, 6 October 2015, para. 92; and C-419/14, *WebMindLicenses Kft. v. Nemzeti Adó-es Vámhivatal Kiemelt Adó- és Vám Főigazgatóság*, 17 December 2015, paras. 69 and 80-82.

[38] See CJEU, C-362/14, *Maximillian Schrems v. Data Protection Commissioner*, 6 October 2015, paras. 94-95, which refer to Article 52 (3) of the Charter. See also Scheinin, Martin and Sorell, Tom (2015), ***SURVEILLE Deliverable D4.10 – Synthesis report from WP4, merging the***

ethics and law analysis and discussing their outcomes, 7 April 2015, p. 9.

[39] See e.g. Brkan, M. (2019), '**The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning**', *German Law Journal* 20 (2019), p. 867; Lenaerts, K. (2019), '**Limits on Limitations: The Essence of Fundamental Rights in the EU**', *German Law Journal* 20 (2019), pp. 779-794.

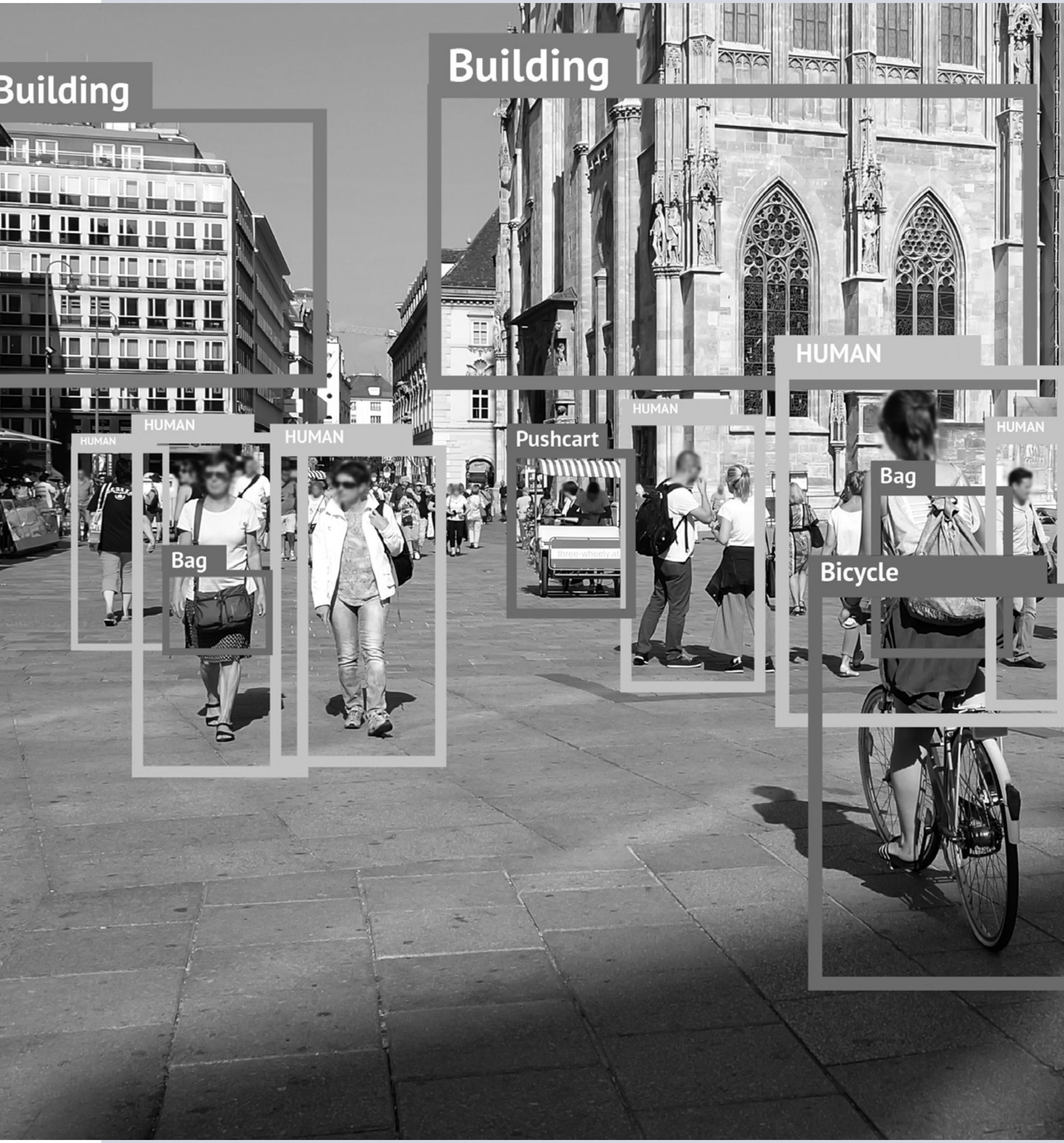
[40] CJEU, Joined cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others*, 8 April 2014.

[41] See, for instance, *Khelili v. Switzerland*, No. 16188/07, 18 October 2011; ECtHR, *S. and Marper v. the United Kingdom* [GC], Nos. 30562/04 and 30566/04, 4 December 2008; ECtHR, *K & T v. Finland*, No. 25702/94, 12 July 2001; ECtHR*, Z v. Finland*, No. 22009/93, 25 February 1997; ECtHR, *Huvig v. France*, No. 11105/84, 24 April 1990; ECtHR, *Leander v. Sweden*, No. 9248/81, 26 March 1987.

[42] Scheinin, Martin and Sorell, Tom (2015), SURVEILLE Deliverable D4.10 – Synthesis report from WP4, merging the ethics and law analysis and discussing their outcomes, 7 April 2015, p. 9.

[43] ECtHR, *S. and Marper v. the United Kingdom* [GC], Nos. 30562/04 and 30566/04, 4 December 2008, para. 112.

[44] See also Council of Europe, ***Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems***, Appendix, para. A.8.



Building
Building
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
Bag
Bag
Pushcart
Bicycle

4

IMPACT OF CURRENT USE OF AI ON SELECTED FUNDAMENTAL RIGHTS

Deploying AI systems engages a wide range of fundamental rights. As seen in **Chapter 2**, the use cases presented in this report involve a range of technologies of varying levels of complexity and automation. They are in different phases of development and applied in different contexts, for different purposes and at different scale. While the rights affected depend on these factors, a number of horizontal and sector-specific fundamental issues emerge.

The chapter begins with a general overview of risks perceived by interviewees, and their general awareness of fundamental rights implications when using AI. The chapter then highlights selected fundamental rights affected by AI-related technologies, with reference to the four use cases analysed.

The analysis takes into account and presents the views, practices and awareness of these issues expressed in the interviews conducted for this report. Interviewees were first asked about the general risks they see when using AI. They were then asked about general fundamental rights awareness when using AI and about more concrete fundamental rights implications, which were mostly linked to data protection, non-discrimination and availability of complaints mechanisms.

4.1. PERCEIVED RISKS

It is important to recognise that many issues cut across different rights. For example, a potentially biased decision made by an algorithm could involve the right to non-discrimination, protection of personal data, and the right to an effective remedy. Similarly, a particular issue can be seen from the perspective of different rights. For instance, a good explanation of a decision made by an algorithm is required under the right to protection of personal data, right to good administration, and the right to an effective remedy and a fair trial.

When asked about general risks when using AI, the interviewees did not always mention fundamental rights as the main risks, although some highlighted related topics. Private sector representatives most often mentioned inaccuracy as a risk of using AI, followed by potential bias and the proper legal basis for processing personal data. One respondent from an international retail company stated that one business risk is linked to European customers being extremely knowledgeable about their rights; namely, people do not hesitate to ask about data storage and automated decision making. If customers are not properly informed, they might complain and the company may lose a client. In addition, the interviewee continued, breaching the law, and possible fines linked to a breach, is another major business risk.

With respect to public administration, bias was most often highlighted as a risk associated with using AI. In addition, public authorities often discussed inaccuracy and data re-identification as risks of using AI. For example, interviewees working on social benefits algorithms stated that incorrect results in general are a risk. This can occur potentially due to rare cases, which are not well identified by the algorithm, or due to errors in the input data. They also highlighted the difficulties associated with moving from testing to deploying a system, including technical challenges, resources required and potential different results when deployed.

Respondents working on targeted advertising also highlighted business risks – for example, when offering irrelevant or inappropriate content. One respondent mentioned potentially losing control over automated systems.

In addition, interviewees indicate challenges linked to the difficulty of interpreting results and outputs from AI systems. One interviewee from the consultancy sector fears that the risk related to the lack or absence of sufficient AI knowledge and understanding can cause ongoing projects to be halted, due to a company's inability to explain clearly what an algorithms will perform, and for what purpose.

Another interviewee from the law enforcement sector, looking into the possible use of AI to support decisions about licence applications, explains that there are inherent risks on how and why such a system proposes a certain response. For example, when potentially using AI to support decisions about license applications for firearms, the respondent asserts that it would not only be critical to understand the reasoning behind negative decisions, but also positive decisions. Several interviews showed that a major concern is to assign properly trained staff with sufficient expertise to trace, explain and interact with the AI system.

This finding is also corroborated by the results of the European Commission survey among companies in the EU. In that survey, 85 % indicate as an obstacle to adopting AI technologies the difficulty to hire new staff with the right skills; 80 % mention the complexity of algorithms as an obstacle.[1]

With respect to the ability to explain decisions based on algorithms, an interviewee working in public administration mentioned that there are no alternatives to being completely transparent when making decisions. There should not be any room for doubt. In a similar vein, a respondent working in the area of health for the private sector mentions that 'self-learning' algorithms are forbidden in their area of work, because only fixed algorithms can be traced.

"The use of AI can bring many benefits, but also risks, it is like nuclear energy."
(Interviewee working in private sector, Spain)

Other risks reported without providing much additional information include cyber-security, data quality, excessive monitoring of people due to the use of data and algorithms, job loss due to automation, and profiling.

4.2. GENERAL AWARENESS OF FUNDAMENTAL RIGHTS AND LEGAL FRAMEWORKS IN THE AI CONTEXT

Not everyone in the EU is aware about their fundamental rights. FRA's Fundamental Rights Survey shows that slightly more than every second person in the EU (aged 16 or older) has heard about the Charter. Slightly more people, two out of three, have heard about the ECHR and the Universal Declaration of Human Rights. This might be because the ECHR is older and more established in people's common knowledge.[2]

"[Our use of AI] does not impact [human rights] in any way. In terms of the decision process, it does not matter whether the decision is made by machine or a human."
(Interviewee working for public administration, Estonia)

The majority of people interviewed for this project acknowledge that using AI can generally affect fundamental rights. Only very few mention that their use of AI does not have a potential impact on fundamental rights or that they were not aware of any such implications. Their responses are influenced by the different ways they use AI, but also their understanding of what fundamental rights are.

For example, one respondent working on the production of pension forecasts based on machine learning says that producing statistics does not have an impact on fundamental rights, apart from data protection issues, which need to be addressed. Another respondent working on social benefits algorithms argues that the impact depends on "how widely human rights are defined" – for example, the right to receive a correct pension.

"Once all the rights related to data protection are ensured, I do not see how human rights are of relevance here."
(Private company, Spain)

None of the interviewees working on targeted advertising believe that their use of AI affects fundamental rights negatively. One respondent working on targeted communication with customers stated that one reason for such a response relates to a lack of knowledge about what exactly fundamental rights are.

"We did not touch the topic because we assume that there are no human rights issues involved: all the activities are within the legal framework, all the activities are compliant with data protection and good practices, and therefore we assume that there are no human rights issues related to these systems."
(Public administration, Spain)

Practically all interviewees showed awareness about the rights to privacy and data protection as well as to non-discrimination. Other rights, such as human dignity, the right to a fair trial and to effective remedy were also mentioned, albeit very briefly.

A closer look at interviewees' responses indicates diverging views across respondent groups. Most respondents working for private companies discuss data protection and non-discrimination, but rarely mention other rights challenges. A company working on targeted advertising mentions that they are attentive to issues linked to freedom of speech and the right to information in the sense that their company promotes these rights. This is because posting adverts helps news and other websites obtain funding to continue their work, one interviewee notes.

The range of rights awareness is much broader among public sector representatives working on AI, who referred to other rights such as human dignity and the presumption of innocence.

Those working on AI systems in different fields of application also highlight that the use of the systems is also covered by sector-specific laws. For example, the system making decisions about unemployment benefits is regulated by national legislation on unemployment insurance, on administrative procedures and on data protection. However, some respondents are not aware of any legal standards that apply to their use of AI or are unsure about it.

In the absence of AI-specific regulation, several respondents mention ethics guidelines and certification schemes. Some work with existing guidelines and standards, not necessarily specifically aimed at AI. This is the case, for example, with the IT security system 'ISKE' in Estonia[3] or in the area of financial services, with the Payment Card Industry Data Security Standard.[4] Respondents also refer to the standards developed by the International Organization for Standardization (ISO), the Institute of Electrical and Electronics Engineers (IEEE) or the European Committee for Standardization (CEN).

A respondent working on targeted advertising argues that certification is not needed in their field because posting ads is not the same as issues linked to the health sector or the work of banks. Several interviewees noted that their organisations are developing (internal) guidelines.

"I do not think that we should regulate specific technology like AI. It is sufficient to have general principles and technology-neutral rules."
(Private sector, Estonia)

Some respondents mention the guidelines developed at the EU and international level, such as the guidelines from the European Commission's

High-Level Expert Group on AI, the OECD's guidelines, or the UNESCO standards. Some are aware of ongoing developments at EU and Council of Europe level.

Some refer to the need to update sector-specific regulations to be able to innovate on AI – for example, in the area of health. Yet one interviewee states that existing standards are sufficient and that AI does not need to be regulated separately.

"Yes, there are codes, and yes, there are procedures, but both these codes and procedures are out of date, because we are using something we created for the analog world in the digital world."
(Private sector, Spain)

4.3. HUMAN DIGNITY

Using AI-driven technologies broadly implicates the duty to respect human dignity, the foundation of all fundamental rights guaranteed by the Charter.[5] Article 1 of the Charter states that human dignity is inviolable and that it must be respected and protected at all times. The CJEU has confirmed in its case law that the fundamental right to dignity is part of EU law.[6]

AI-driven processing of personal data must be carried out in a manner that respects human dignity. This puts the human at the centre of all discussions and actions related to AI. Rather than the technology, the 'human being' creating and affected by the new technology needs to be the focus. Taking human dignity as the starting point can help to ensure that the use of AI benefits everyone – for example, by supporting ageing and access to healthcare in a dignified manner.

The use of AI also risks infringing on other closely connected Charter rights, such as the right to life (Article 2) and the right to integrity of the person (Article 3). In this context, it is important to consider how to avoid the harmful use of AI to prevent violations of these rights, for example when it comes to the use of AI by people engaging in criminal activities or when AI is used for weapons.[7]

Apart from such extreme cases, preserving dignity includes avoiding subjecting people to AI without their knowledge and/or informed consent, which is strongly linked to privacy and data protection. For example, when people's applications for social benefits are decided upon through the use of AI, people need to be made aware (and consent to the use when automated decisions are taken). To give another example, a certain proportion of the population does not feel comfortable being subjected to biometric identification systems. Hence, using it without allowing them to opt out could potentially violate their dignity.[8]

Only very few respondents from public administration referred to the right to dignity when discussing fundamental rights. One respondent, considering the use of AI in prisons, mentions that in this particular context it first needs to be assessed whether the risk of violating fundamental rights would be too high, such as the right to human dignity. Other interviewees made only general references to this right, without discussing it in relation to a concrete use of AI.

4.4. RIGHT TO PRIVACY AND DATA PROTECTION – SELECTED CHALLENGES

The right to respect for private life and the protection of personal data (Articles 7 and 8 of the Charter) are at the core of fundamental rights discussions around the use of AI. While closely related, the rights to respect for private life and the protection of personal data are distinct, self-standing rights. They have been described as the "classic" right to the protection of privacy and a more "modern" right, the right to data protection.[9]

Both strive to protect similar values, i.e. the autonomy and human dignity of individuals, by granting them a personal sphere in which they can freely develop their personalities, think and shape their opinions. They thus form an essential prerequisite for the exercise of other fundamental rights, such as the freedom of thought, conscience and religion (Article 10 of the Charter), freedom of expression and information (Article 11 of the Charter), and freedom of assembly and of association (Article 12 of the Charter).[10]

Given that these two rights are not absolute rights, they can be subject to limitations. However, any interference needs to be adequately justified[11] and cannot compromise the essential, inalienable core of that right,[12] as explained in **Section 3.3.**

The concept of "private life" or "privacy" is complex and broad, and not susceptible to an exhaustive definition. It covers the physical and psychological integrity of a person, and can, therefore, embrace multiple aspects of the person's physical and social identity.[13] There is also a zone of interaction of a person with others, even in a public context, which may fall within the scope of "privacy". In other contexts, the ECtHR has used the concept of "reasonable expectation of privacy" – referring to the extent to which people can expect privacy in public spaces without being subjected to surveillance – as one of the factors, albeit not necessarily a conclusive one, to decide on a violation of the right to respect for private life. Its relevance and scope of application, however, appears to be limited.[14] Similarly, according to the UN Human Rights Committee, the mere fact that participants in assemblies are out in public does not mean that their privacy cannot be infringed. The same applies to the monitoring of social media to glean information about participation in peaceful assemblies.[15]

The widespread use of AI-technologies may, as the technologies continue to develop, raise unchartered issues and novel concerns about the right to respect for private life. AI-driven technologies may change the way we think about privacy. Algorithmic tools can predict, and reveal information about, people's behaviour in unprecedented ways – without people even realizing that they are giving away such information. Personal data obtained from the internet may, for instance, then be used for targeted advertising, raising many fundamental rights concerns.[16] Issues linked to personal data sharing via smart-phone apps particularly raises significant concerns, including a variety of potential harmful effects, such as manipulation and exploitation of vulnerabilities, discrimination, security issues and fraud (e.g. identity theft) and reduced trust in the digital economy.[17]

Using AI-driven technologies often implies computerised processing of large amounts of personal data. This constitutes an interference with the right to protection of personal data set out in Article 8 of the Charter (embodying pre-existing EU data protection law), as well as the right to private life under Article 7 of the Charter and Article 8 of the ECHR.

Awareness of data protection issues and use of personal data

In the EU, 69 % of people have heard about the GDPR.[18] By contrast, virtually all interviewees are aware of the GDPR and discussed data protection issues. Data protection rules deriving from the GDPR and national law are clearly the most well-known and applied rights in the area of AI. Other fundamental rights are less known.

"We were a little anxious when the GDPR was implemented, but in the end it meant managing datasets and access rights [...] It is a good reminder that not everything can be or should be done."
(Public administration, Finland)

When discussing the legal framework governing the use of AI, most respondents only mentioned data protection rules, as well as some sectoral laws. Some clearly say that there is no other legal framework apart from the data protection laws. An interviewee working for a Spanish public administration notes: "we rely on the data protection regulation and norms, which is all that is available at the moment".

"Actually, I'm concerned that the GDPR might hinder AI research. I'm afraid that some large databases that we have used previously cannot be used for our research anymore."
(Private company, Netherlands)

One interviewee, reflecting on an image-based diagnostic tool, expressed the view that the GDPR could hinder research. The hospital using the tool to support diagnosis after strokes had clear rules on data protection, the interviewee indicated, although they did not know whether data protection certification was requested.

"There is the GDPR but it does not give you specific rules. It gives principles but it comes down to ethical issues and interpretation."
(Private company, Estonia)

Others referred to more general data protection guidelines or indicated that they were not aware of such documents.

All respondents working in target advertising are aware of privacy and data protection issues. Although not all are responsible for data protection issues in their companies, they are all aware of efforts to protect the data and privacy. One interviewee mentioned that, contrary to earlier years, personal data are now stored much more securely and handled with more care. Attention was given to properly handling consent for data processing. As a consequence, there is a high level of awareness about data protection and privacy issues linked to AI use.

However, data protection law only applies when personal data are processed. For example, using anonymised data to develop AI tools (i.e. as training data) is most likely permissible in many instances and would not trigger the GDPR.

Research shows that data can often be de-anonymised.[19] However, such efforts often require expert knowledge and potentially additional information, and are illegal. While the illegality of de-anonymisation does not necessarily preclude the applicability of the GDPR, it is more important to consider if re-identification of anonymised data is reasonably likely.[20] Anonymising data is only one aspect of protecting the privacy of data subjects. When assessing risks of re-identification, other aspects are also important to consider when disseminating anonymised data. These include who will use the data, for what purpose, and what outputs will be produced.[21]

In the interviews, respondents were not always entirely clear about their use of personal data. They often only superficially described the data used, as mentioned in Chapter 2. In several instances, interviewees indicated that they use non-personal data or anonymised data, arguing that data protection was not relevant in such cases. For example, a semi-public organisation working on environmental management uses aggregated data on water consumption for machine learning-based predictions of water consumption. These data are not available at individual level.

Other interviewees said they did not use personal data, although the data originally stem from individuals. The tool supporting restaurant inspection by collecting data from online sources does not use any personal data – the interviewee indicated. However, they indicated the need to be careful when mining data online, because, even if publicly available, it might include personal data, such as usernames.

"It would be great to retrieve some data from another service so that the client wouldn't have to repeat it all, but where does the line go in reusing data?"
(Public administration, Finland)

In another example, an insurance company is using a chatbot to make client contact more effective. The data used to train the system are chat protocols (conversation logs), which are not linked to any personal data. However, in this example, linking these data to personal data might be possible in the future, according to the respondent.

Companies working on targeted online advertising indicate using (pseudo-) anonymised data. This is done, for example, after excluding names and social security keys and encrypting data. The identity of the consumers is not relevant to the company, an interviewee mentioned.

While some indicate that they use non-personal or anonymised data, for others this is not possible because the data are used to make predictions or decisions about specific individuals. For example, an interviewee from a company working on credit rating mentioned they need to know the identity of consumers for their assessments. In this case, this is even more important than the right to be forgotten, according to an interviewee.

An exhaustive discussion of data protection issues is not possible in this report. However, two aspects clearly emerged during the interviews: automated decision making linked to the right to human review, and the right to obtain meaningful information when decisions are automated.

Automated decision making

Article 22 of the GDPR and Article 11 of the Law Enforcement Directive generally forbid automated decision making, meaning any "decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Under Article 22 of the GDPR, explicit consent is needed when decisions are solely automated and have a legal or similarly significant effect on people and if such automated decision making is not authorised by law. The authorisation by Union or national law is the sole precondition under the Law Enforcement Directive (Article 11) for such processing. For a decision not to be considered fully automated, both instruments require human review by the controller.[22]

However, the concept of 'automated' decision making is elusive and requires further discussion and research. For example, in some cases, human intervention might be limited to 'signing-off' on the outcomes of the AI system, rendering it virtually automated.[23] Importantly, human review must not mean a human just signing off the recommendations or outputs from an algorithm. It must be done by someone who has the "authority and competence to change the decision", considering all relevant data at hand.[24] If humans review and potentially override outcomes of the system, this must also be evaluated.

Research indicates that humans overrule outcomes from algorithms mainly when the result from the algorithm is not in line with their stereotypes. [25] This behaviour threatens the possible added value of automated processing in being potentially more accurate or even fairer than humans. It may also put minority groups at a disadvantage, and is therefore also relevant for non-discrimination issues (discussed below).

Overall, there is disagreement about the exact scope of these provisions of the EU data protection *acquis*, and whether they impose a general ban on certain types of automated decisions, or provide data subjects with some rights in the context of certain types of AI-driven decision making.[26]

Using algorithms in the area of social benefits, health and predictive policing clearly have potential legal or other significant consequences. The interviews suggest that those working in these areas are well aware of the concept of human review before decisions are taken with the support of AI.

Many interviewees indicate that no automated decisions are taken. One exception is an automation of unemployment benefits, which is, based on national law, fully automated for decisions that do not involve any discretion. In another example, from another country, only positive decisions, based on pre-defined rules, are automated for student benefits. In this case all negative decisions are made by humans. Both cases refer to rule-based decisions, not involving the use of statistics or machine learning.

Another respondent, testing the use of AI systems, including machine learning in the area of social benefits, mentions that equality could be negatively impacted. This is because automation makes human behaviour visible, including existing biased practices. This makes precautions necessary and, as a consequence, the organisation would only allow decisions made by humans.

Interviewees working in health highlighted risks linked to the automation of decisions. An interviewee discussing the tool to support stroke diagnosis feels it is important not to rely on the system to avoid the risk of automation or confirmation bias. They caution that early positive experiences with the application could prompt users to rely on it too easily and devote less attention to their own assessment of the images. Other interviewees raised similar concerns. One interviewee, discussing a tool that analyses images to provide a probability for the presence of a certain type of lesion, notes that the technology supports the diagnosis of simple cases, but that the expertise of doctors is particularly important – and trusted – in more complex cases.

Targeted advertising is often considered not to have a significant effect on people. However, this may be the case if, for example, an individual's vulnerabilities are used for successful advertising. Considering vulnerabilities is particularly important for people from disadvantaged groups, who may not be aware that they can opt out of direct marketing (see box) or of their right to have a say when decisions are automated.

In the absence of case law in this area, more information and research is needed to identify the impact of such automated decisions (i.e. which advertisement will be delivered to whom, when, how and why). Answering these questions is challenging, as targeted advertising is based on highly complex technology and at scale.

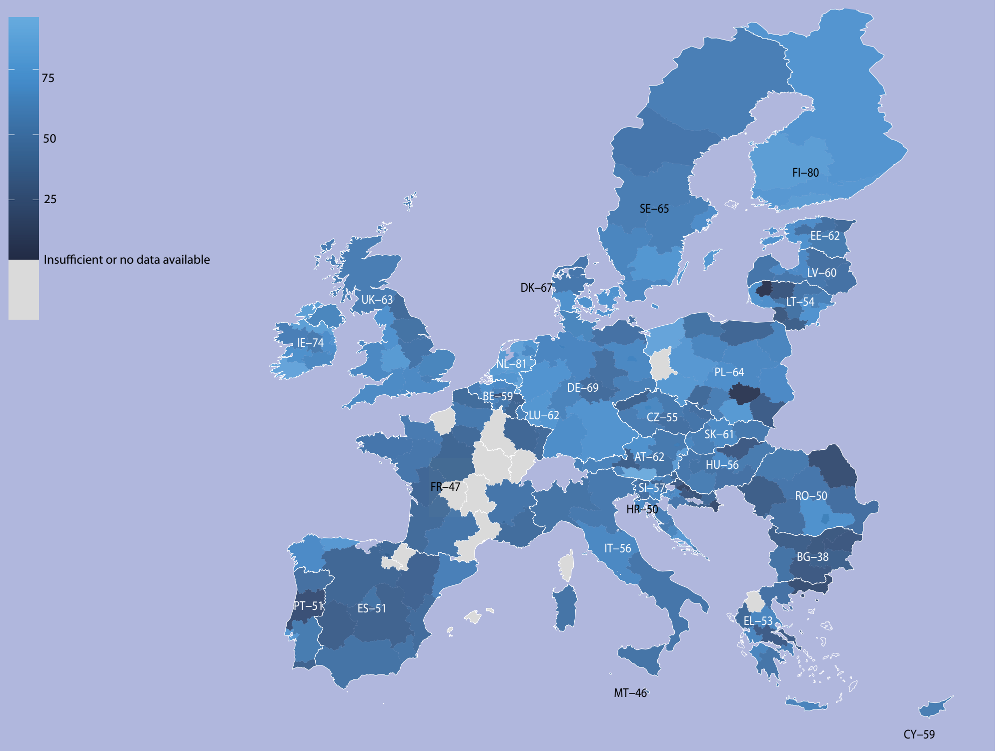
Awareness of right to opt out from direct marketing among general population

In 2019, a Eurobarometer survey asked people in the EU if they are aware of their right to opt out from direct marketing. Overall, only 59 % of EU citizens have heard about this right (with 24 % having exercised it). But people can only exercise their right if they are aware of it – which becomes even more important when direct marketing is made much more efficient through machine learning.

Awareness levels strongly vary across the EU. The percentage of people who know about their right to opt out from direct marketing ranges from 38 % in Bulgaria to 81 % in the Netherlands. Figure 4 shows the percentages. It also highlights – based on FRA's analysis of Eurobarometer data – that there is a strong variation within countries, when broken down by regions.

In some regions, fewer than one in four have heard about their right. These are areas with higher shares of people at risk of poverty. This indicates the general problem that people who are more disadvantaged in society tend to be less aware of this right. The data show that people who are not working, who more often struggle to pay their bills, who are living in rural areas, or who are older, are less aware of this right.

FIGURE 4: AWARENESS OF GDPR RIGHT TO OPT OUT FROM DIRECT MARKETING, IN THE EU AND UNITED KINGDOM, BY COUNTRY AND REGION (%)



Note: Map does not show non-EU countries other than the UK. Light shading = more aware of right. Dark shading = less aware of right. Results for regions within countries represented by light grey spaces were excluded because there were fewer than 20 respondents, meaning the numbers of observations were too low for reliable results. N = 26,503. Question: "The General Data Protection Regulation (GDPR) guarantees a number of rights. Have you heard of each of the following rights? […] 18.2 The right to object to receiving direct marketing."

*Source: FRA, 2020 [Calculations and presentation based on European Commission (2019), **Eurobarometer, 91.2**]*

Experiences from use cases

In general, interviewed experts highlighted that data protection law is difficult to interpret and lacks clarity when it comes to the meaning of automated decision making. One expert from France felt that, because automated decision making is so difficult to explain, all automated decision making should be banned, meaning the exceptions in the GDPR that allow for some automated decision making should be removed. They pointed out that AI can only be used as a decision support tool.

Another expert, an independent lawyer from the Netherlands, views current laws and standards as sufficient, but says they need to be concretised per sector. Particularly, the expert mentions that the scope of the existing rules on permissible automated decision making are not clear and that it remains unclear to him what a comprehensive assessment, or a 'human in the loop' means. This was also raised in relation to the SyRI case, where it remained unclear to what extent the decisions were reviewed.

Another expert working at a supervisory authority, more generally, sees no need for adapting data protection laws as "the legislation is quite comprehensive. It is more about the organisation of the supervision thereof, and also the political will behind it".

These concerns reflect the findings of other research, which also raise serious issues concerning the right to human review. For example, the responsible officers questioned the results of an algorithmic system built to profile unemployed people in Poland in less than one percent of cases. This essentially makes a supporting tool an automated decision making tool.[27]

Linked to the question of reviewing decisions or outputs from AI systems is the challenge of a clear lack of knowledge about how AI works. Interviewees often could not explain in detail how the system they use works or which data it uses, be it due to the lack of knowledge or lack of transparency. Meaningful information about the logic involved, or explaining outcomes from algorithms, is essential for several fundamental rights. It is crucial not only for the processing of personal data, but also for ensuring that the algorithms are fair and do not discriminate. It is also necessary to enable people to properly challenge decisions and AI systems.

One interviewee working for public administration explains that the complexity differs depending on the tasks. Licence administration systems can be relatively straightforward. Crime prevention analysis uses more data sources, which makes it harder to understand. Another interviewee working in law enforcement says that the current AI used by police organisations is not yet so complex that it would make explanations difficult, but that this might be the case in the future.

A respondent working on financial data transactions indicates that traditional models were straightforward to understand. However, new methodologies are more difficult to explain, and the company has to invest resources into making these models more explainable. Still, the level of explainability required by the GDPR is not clear to the respondent.

"There is a risk of having too much trust in the machine."
(Public administration, France)

"There is a huge tension surrounding the GDPR. So we want to do well, but might in fact be worse off, because interpretation of the data then turns out to be impossible."
(Public administration, Netherlands)

"If we had to explain the model, we wouldn't be able to. The model is statistical and not very explainable."
(Public administration, France)

"Internally we can explain the decisions of the machine learning models and we have several means to do that."
(Private sector, Estonia)

"If the systems do not have black boxes of information or processes, we already take a step forward in the defence of human rights."
(Public administration, Spain)

"We are strongly attached to the idea that AI has to be explainable."
(Public administration, France)

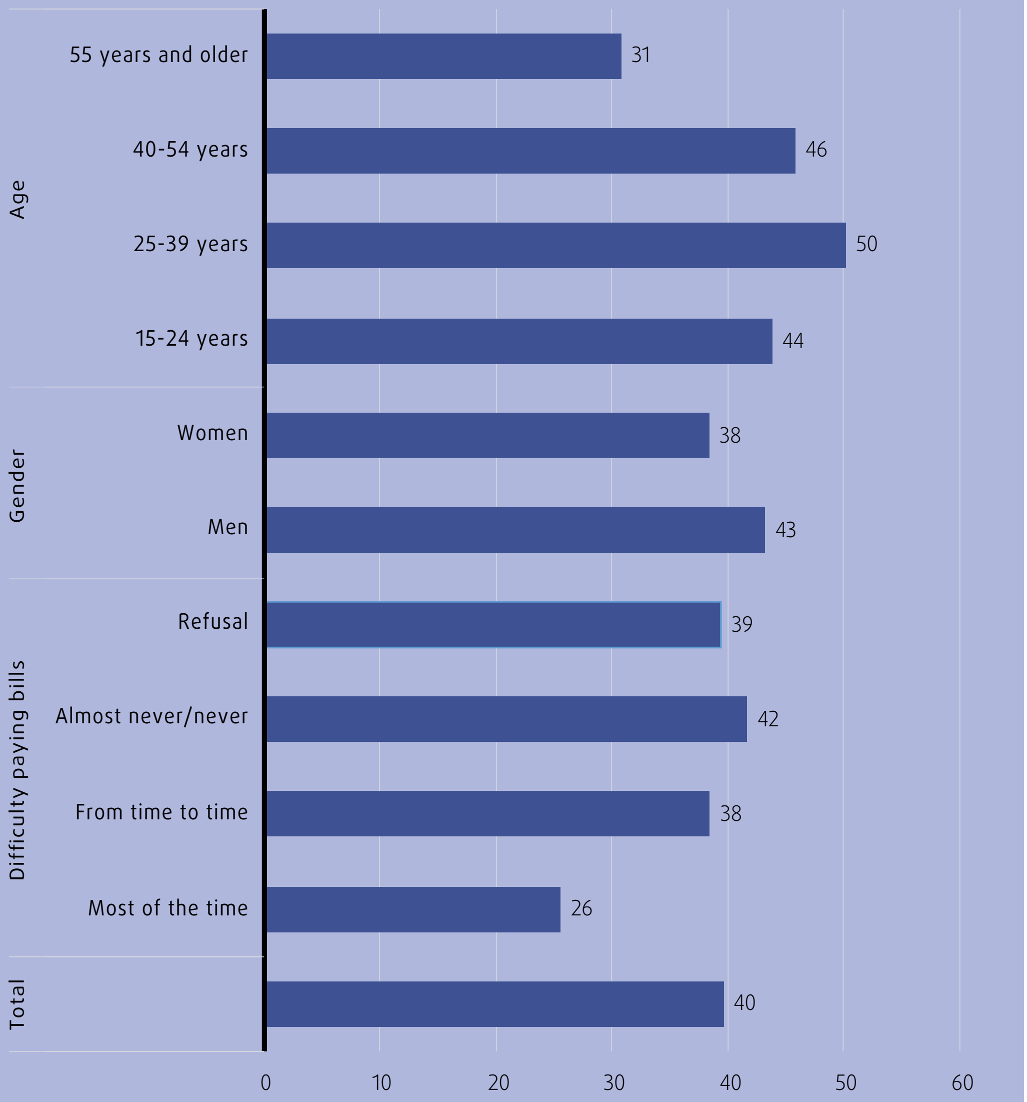
Awareness of right to have a say when decisions are automated

Most people are not aware that they have the right to have a say when decisions are automated, evidence suggests. A Eurobarometer survey showed that 40 % of Europeans know about their data protection rights.

FRA's analysis of the Eurobarometer survey shows that this figure drops considerably among people with lower socio-economic status. Only 26 % of EU citizens who report that they are struggling to pay their bills most of the time know about this right. This lack of rights awareness among those socially disadvantaged could contribute to further social exclusion if those already disadvantaged are less aware that they can challenge (automated) decisions about them (see Figure 5).

Gender differences are small, yet women are even less aware of this right (38 % of women and 43 % of men). Older people are considerably less aware (31 % among those aged 55 and older).

FIGURE 5: AWARENESS OF RIGHT TO HAVE A SAY WHEN DECISIONS ARE AUTOMATED, BY AGE, GENDER AND DIFFICULTY IN PAYING BILLS (%)



Notes: N = 26,503. Question: "The General Data Protection Regulation (GDPR) guarantees a number of rights. Have you heard of each of the following rights? […] 18.5 The right to have a say when decisions are automated (e.g. an algorithm decides if you will be granted a loan or not)."

Source: FRA, 2020 [calculations and presentation based on European Commission (2019), **Eurobarometer, 91.2***]*

4.5. EQUALITY AND NON-DISCRIMINATION

Equality before the law and non-discrimination are enshrined in Articles 20 and 21 of the Charter. Discrimination is "where one person is treated less favourably than another is, has been or would be, treated in a comparable situation" based on a perceived or real personal characteristic[28] (called 'protected grounds/characteristics'). Article 21 of the Charter prohibits any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.

The Charter prohibition reflects corresponding rights in the ECHR (Article 14) and in Protocol No. 12 to the ECHR (Article 12), but is even broader, as it establishes a non-exhaustive, open list extending protection to a wide range of new grounds. Unlike Article 14 of the ECHR, the Charter right to non-discrimination is a freestanding right that applies to situations that do not need to be covered by any other Charter provision.[29]

Main challenges

Discrimination is a crucial topic when it comes to the use of AI, because the very purpose of machine learning algorithms is to categorise, classify and separate. As one interviewed expert points out, making differences is not per se a bad thing. According to this expert, when deciding to grant a loan, credit history can be used to differentiate between individuals, but not on the basis of protected attributes, such as gender or religion. However, many personal attributes or life experiences are often strongly correlated with protected attributes. The credit history might be systematically different for men and women due to differences in earnings and job histories.

Interviewees often mention efficiency as the main purpose for using AI-related technologies. Yet it is important to note that this cannot not justify unfair, differential treatment.

Often, protected attributes might be highly correlated with risks. For example, differences in life situations among men and women might often be linked to different insurance risks. This is, however, not acceptable, as the *Test-Achats* ruling, [30] shows. In that case, the CJEU put an end to gender discrimination in insurance pricing.[31]

Under certain circumstances and in some areas, using algorithms could positively contribute by reducing bias and stereotyping. Algorithmic data analysis may produce results that could dispel prejudicial attitudes. For example, predictive policing might, in some contexts, lead to more equitable and non-discriminatory policing by reducing reliance on subjective human judgments.[32] Predictive techniques may be used to identify so-called 'white-collar crimes', such as financial crimes that are historically under-policed.[33]

Nevertheless, direct or indirect discrimination[34] through the use of algorithms that involve big data is considered as one of the most pressing challenges in the use of AI-driven technologies.[35] Bias and discrimination, including gender-based discrimination, in data-supported algorithmic decision making can occur for several reasons and at many levels in AI systems. They are difficult to detect and mitigate.[36] Often, the quality of the data and biases within it are the source of potential discrimination and unfair treatment.[37]

The discriminatory effects generated on certain groups are, in practice, very difficult for individuals to challenge.[38] So far, only a limited number of court cases have dealt with discrimination relating to AI systems.[39]

UK Court of Appeal: police use of facial recognition violates human rights

A first instance decision of the Divisional Court of Cardiff in 2019 dismissed a claim concerning the lawfulness of the South Wales Police's use of the "AFR Locate" face recognition system. The Court of Appeal overturned that decision.

It found that the facial recognition programme used by the police was unlawful. The Court of Appeal ruled that "too much discretion is currently left to individual police officers". It added that "[i]t is not clear who can be placed on the watch list, nor is it clear that there are any criteria for determining where [the technology] can be deployed".*

The court also held that the police did not sufficiently investigate if the software in use exhibited race or gender bias.

This judgment is the first in-merit specifically on this matter in Europe. It considerably narrows the scope of what is permissible and what law enforcement agencies need to do to fully comply with human rights law.**

** UK, Court of Appeal,* ***R (Bridges) v. CC South Wales****, [2020] EWCA Civ 1058, 11 August 2020.*

*** Ars Technica, '****Police use of facial recognition violates human rights, UK court rules****', 11 August 2020.*

Studies have highlighted the potential for discrimination prompted by the use of AI-systems across the areas covered by the report.[40] In the area of predictive policing, for example, a particular risk relates to the potential for automated decision making tools to reproduce and entrench existing discriminatory practices that undermine equality before the law (Article 20 of the Charter). The historical crime data that underpins predictive policing may be biased,[41] reflecting inherent data gaps (e.g. chronic underreporting for certain types of crime), alongside issues with how data is recorded (e.g. human error, but also bias by individual officers).

Crime victimisation surveys consistently show that a large proportion of crime is never reported to the police by the public – particularly crimes involving physical and/or sexual violence, and hate crimes. For example, FRA's survey on violence against women – with 42,000 respondents – showed that only one in five women who experienced violence, by their partner or anyone else, brought the most serious incident to the attention of the police.[42] FRA's

EU-MIDIS II survey of 25,500 respondents across the EU showed that only three in ten reported incidents of racially motivated hate crime to the police or any other organisation.[43]

Compared with violent crime and hate crime, property crime – such as burglary – has a higher rate of reporting to the police, particularly in developed countries. This may be because this is a requirement when claiming on an insurance policy.

In sum – relying on official crime statistics (that are based on reported crime) when looking to develop AI models in the field of predictive policing is particularly problematic when it comes to specific crimes and specific groups.

Some variables used in AI modelling can be proxies for race, ethnicity, gender and other protected categories. The complexity of the algorithms makes it harder to identify and remove such biases. Instead of providing objective analysis, predictive policing software may turn into an 'echo chamber' cementing existing systemic flaws and injustices with the 'stamp' of what appears to be scientific legitimacy.[44]

The use of predictive policing may also make law enforcement responses less equitable by focusing on certain crimes or areas.[45] Predictive policing is currently focused on property crimes such as theft and burglaries, which are often associated with certain demographics and neighbourhoods. This can result in certain demographics and neighbourhoods – and the individuals living in them – being further stigmatised.[46] Meanwhile, white-collar crime – typically committed by different demographics – is less prioritised.[47] These patterns of policing – whereby certain neighbourhoods or communities are disproportionately policed – predates the use of AI. However, the 'promise' that AI is more 'objective' and can, in turn, be used to counteract discriminatory policing, needs to be verified in practice.

Oxford University researcher Sandra Wachter highlights that discrimination may occur due to information linked to protected attributes in targeted advertising. Newly created profiles for the purpose of advertising might amount to indirect discrimination and potentially even require new characteristics to be added to non-discrimination legislation, and extend for its scope to be expanded to other areas.[48]

Experiences from use cases

Many interviewees noted that the use of AI, in general, can discriminate, but that the systems they are working with do not. Many indicated a belief that excluding information on protected attributes is sufficient protection against discrimination. However, discrimination can occur due to other information contained in datasets that may indicate protected attributes. Traces of protected groups are often hidden in other information.

"If you want the machine not to discriminate on the basis of sex, do not put the variable of sex, as easy as that, or make the examples symmetrical if you notice that sex has certain relevance."
(Public administration, Spain)

An example from a public authority, which uses AI in tax and customs, shows the challenges linked to identifying possible bias and potential discrimination when using algorithms. When scrutinising their algorithms, a public administration body found a higher degree of errors in tax declarations among recently issued national identification numbers, which have almost always been attributed to immigrants. This prompted further research into the correlation. It turned out that the outputs of people with recent identification numbers more often contained errors because they had never filed their taxes before, and did not know how to do so (which was also the case for non-migrants). This is also an example of proxy information, where parts of a number could indicate immigrant status.

Another interviewee working on the potential use of AI for detecting benefits fraud mentioned in this respect: "If you want to prevent discrimination based on ethnicity, for instance, it does not suffice to just remove the 'ethnicity label', because the neighbourhood composition is often also determined by ethnicity, or ethnicity plays a role in it. So [preventing discrimination] often goes beyond the 'direct' characteristics".

"[I]f you do not have access to sensitive personal data, it is impossible to check if you are profiling on that basis."
(Public administration, Netherlands)

Even if most of the respondents were aware of the general potential for discrimination when using AI, they often ruled out that their system discriminates against people based on protected characteristics. Some respondents also believe that their tools have a positive impact in terms of non-discrimination. One respondent, testing AI for social benefits decisions, regrets not being able to use AI for data protection reasons, even though, in the respondent's view, automation could process big datasets effectively and without discrimination. While noting that protection of personal data needs to be observed, the respondent feels it hinders prompt decision making and non-discrimination – "if it can be automated, it should be automated".

Some respondents were not clear or not sure about whether their use of AI could discriminate. Respondents repeatedly stated that their system cannot discriminate because it does not include data on protected characteristics. For example, several interviewees working in predictive policing and law enforcement felt that there was no potential for discrimination, as the AI systems did not use data on, or return outcomes related to, protected grounds, or because the system does not aim to identify people.

Others working on predictive policing felt that discrimination could occur, in particular because of issues in the training data. In relation to the predictive policing 'heat map' case, for example, one interviewee noted that – because the dataset is never fully neutral, representative or complete – there is a strong risk of bias and possible discrimination towards particular groups. They identified sharing datasets to increase the amount of data available as one way to mitigate this risk, but felt that this was impeded by data protection regulations. They also indicated that multi-level teams with the task to travel to different police authorities and check on the quality of the systems used are being set up.

In the area of targeted advertising, some interviewees mentioned discrimination as a potential problem, mainly after being asked directly about it. Overall respondents do not think that their systems discriminate. Three respondents mention that information on gender and age is not used and consequently no discrimination in this respect can occur. Another interviewee is not sure if this information is included or not.

"For discrimination, it's complicated because some diseases are more present in certain ethnic groups. Predictions take into account the sexual, ethnic, genetic character. But it is not discriminatory or a violation of human rights."
(Private sector, France)

A respondent working on a breast cancer detection tool highlighted that age, gender and ethnicity are relevant factors as some population groups are more likely to develop certain types of cancer. Respondents working in health highlighted that the potential for discrimination is also linked to who uses the system, suggesting that this could become a greater challenge if the system were used by non-medical staff.

A different, but related, example comes from a respondent working on credit rating for a private company, selling credit scores of individuals created by an algorithm. The company uses information about gender, age and citizenship in its credit risk models. This information has some impact on the outcome of the credit scores. For example, younger people or non-citizens have a higher credit risk score, but the influence of demographics is much smaller compared to credit history data. According to the interviewee, their system "certainly does not impact on the right to non-discrimination, because we do not make any decisions, we sell data and data analytics. Creditors have to monitor that they do not discriminate".

Another interviewee working on the data strategy for a financial institution in the private sector, using AI to analyse financial transactions, clearly mentions the challenges of understanding what non-discrimination constitutes for their work. The interviewee mentions, for example, that it is not clear to what extent it is illegal to exclude older people from receiving credit if their life expectancy is expected to be lower than the mortgage repayment period they asked for.

These findings point to uncertainty and ambiguity in the financial sector with respect to how Article 21 of the Charter – on non-discrimination – translates into real life situations.[49]

Vulnerable groups

Much of the discussion and research about discrimination when using AI is linked to biased results with respect to ethnic origin, gender, and to some extent, age. Although it is important to analyse potential discrimination against these groups, the Charter covers several other grounds of discrimination, which are less often part of discussions or research.

These other grounds include, for example, political opinion, sexual orientation, and disability. The Charter provides particular rights to some special groups (beyond Articles 20 and 21), including the rights of the child (Article 24), the rights of the elderly (Article 25), and the rights of persons with disabilities (Article 26).

The question of age – with respect to older age groups and younger adults – came up during the interviews, notably when it comes to insurance and credit (see above).

However, none of the interviewees or experts directly mentioned the rights of the child. This might be linked to some extent to the nature of the use cases investigated, but it clearly reflects the fact that this topic is not high on the agenda of many of those working in AI.

Article 24 of the Charter emphasises that the best interests of the child must be the primary consideration in all activities of public authorities and private actors that concern children, which applies of course – equally – to the field of AI.[50]

Only two respondents from public administration mentioned possible use of AI in the area of child custody and the distribution of children in schools. But they did not address this in consideration of the rights of the child. These respondents did not wish to go into more detail concerning these use cases – potentially reflecting the sensitivity of this topic.

Finally, issues linked to the integration of people with disabilities were not raised in any of the interviews.

Awareness among general population of potential for AI to lead to discrimination

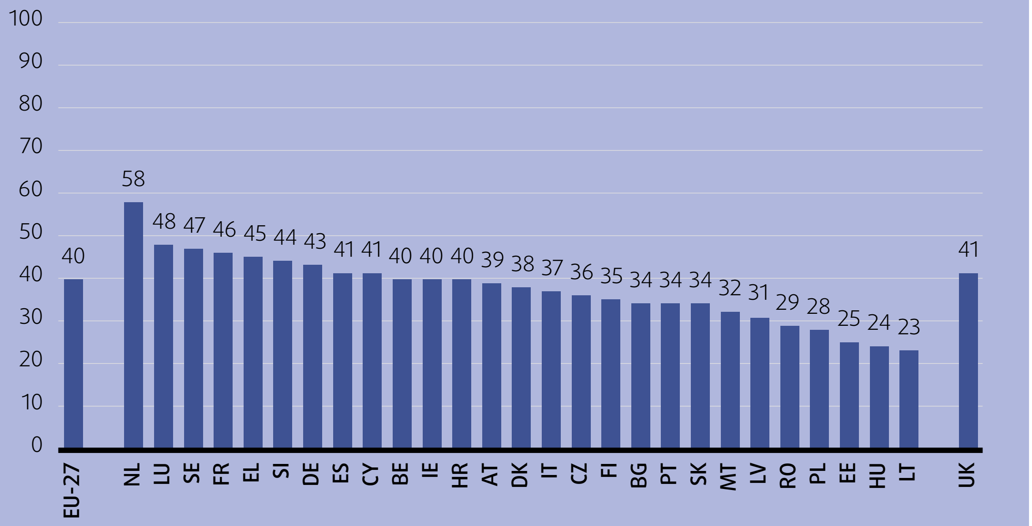
A Eurobarometer survey that included questions on AI asked respondents about the areas they are mostly concerned about when it comes to the use of AI, including discrimination in decision making, unclear responsibility, and that there is nobody to complain to.

Only around 40 % of EU citizens indicated that they are concerned that using AI could lead to discrimination in terms of age, gender, race or nationality – for example, in taking decisions on recruitment, credit worthiness, etc.

Results vary across countries. Higher proportions of people are concerned about discrimination in the Netherlands (58 %), Luxembourg (48 %) and Sweden (47 %). Lower proportions expressed concern in Estonia (25 %), Hungary (24 %) and Lithuania (23 %) (see Figure 6).

However, from this question it is not clear if people do not know that discrimination can happen, or if they are aware that it can happen but do not think it is a problem.

FIGURE 6: AWARENESS ABOUT THE RISKS OF DISCRIMINATION WHEN USING AI, BY COUNTRY (%)



Notes: Includes people who indicated that they are concerned that AI could lead to discrimination among three possible issues, or all of the three issues.

*Source: FRA calculations based on European Commission (2019), **Eurobarometer, 92.3***

Tackling gender inequality in the design and use of AI

The Charter stipulates that equality between women and men must be ensured in all areas, including employment, work and pay (Article 23). Gender discrimination is a major concern when it comes to the design and use of AI and related technologies.*

On the development side, the European Economic and Social Committee notes that the development of AI is taking place within a homogenous environment principally consisting of young white men. This results in cultural and gender disparities, which are being embedded in AI technologies. For example, training data are prone to manipulation, may be biased, reflect cultural, gender and other prejudices or preferences, and contain errors.** This is also reflected in this research, where, despite efforts to achieve gender balance, the majority of interviewees were men.

Disparities at the design and deployment stage are linked to the systematic disadvantages affecting women in the labour market and the potential lack of awareness of gender biases. A recent study showed that the increased use of industrial robots could widen the gender gap, despite both genders benefitting from increased automation, as the analysis indicated that men in medium- and high-skill occupations would benefit disproportionally.***

Looking ahead, using data and algorithms could help to better mainstream gender equality into policies and processes by paying attention to gendered datasets. Drawing on discussions around gender inequalities and the use of data ('data feminism')**** could help to raise awareness that the male point of view should not be taken as the default view, which also then finds its way into datasets.*****

* *See also European Commission,* ***White Paper On Artificial Intelligence – A European approach to excellence and trust****, COM(2020) 65 final, Brussels, 19 February 2020, p. 1.*

** *European Economic and Social Committee, Artificial intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion), 31 May 2017, JO C 288, p. 43.*

*** *Aksoy, C., Özcan, B. and Philipp, J. (2020),* ***Robots and the Gender Pay Gap in Europe****, IZA Discussion Paper No. 13482.*

**** *See the* ***webpage on data feminism*** *on the datasociety's website.*

***** *Criado Perez, C. (2020), Invisible Women. Exposing data bias in a world designed for men, London.*

4.6. ACCESS TO JUSTICE

The right to an effective remedy before a tribunal and to a fair trial (Article 47 of the Charter) is one of the most often used Charter right in legal proceedings. This highlights its importance in upholding fundamental rights and the rule of law. This right of horizontal character empowers individuals to challenge a measure affecting any right conferred to them by EU law, not only in respect of those guaranteed in the Charter.[51] The CJEU has underlined that Article 47 of the Charter constitutes a reaffirmation of the principle of effective judicial protection and that the characteristics of a remedy must be determined in a manner that is consistent with this principle.[52]

The right to an effective remedy also covers decisions taken with the support of AI technologies. EU data protection law reconfirms that the right to an effective judicial remedy must be provided in relation to decisions by the controller or the processor[53] as well as the supervisory authority.[54] Data processed by AI-driven technologies is no exception.

It is crucial to note that the possibility to lodge an administrative complaint before a supervisory authority as provided for by the GDPR and the Law Enforcement Directive[55] is not considered an effective judicial remedy under Article 47 of the Charter. This is because no court is involved in such a review. Judicial review should always remain available and accessible, when internal and alternative dispute settlement mechanisms prove insufficient or when the person concerned opts for judicial review.[56]

Using AI can challenge the right to an effective remedy in different ways. One prominent concern is the lack of transparency in the use and operation of new technologies. Algorithmic decision making is notoriously opaque: data collection, algorithm training, selection of data for modelling or profiling, the situation around individual consent, effectiveness and error rates of the algorithm and other aspects are often not transparently reported.[57]

Without access to this information, individuals may not be able to defend themselves, assign responsibility for the decisions affecting them,[58] appeal any decision negatively affecting them or have a fair trial, which includes the principle of equality of arms and adversarial proceedings as established by the ECtHR.[59] These requirements also form part of the corresponding Charter right (Article 47) in view of Article 52 (3) of the Charter.

Main challenges

These issues are reflected in the specific challenges to the right to an effective remedy and a fair trial that the interviewed experts outlined. Generally, experts indicate a difference in accessing remedies at private companies and public administration. Public authorities are more often forced to be transparent about their use of AI. Meanwhile, companies appear to be more secretive, the assessment of several experts suggests. However, an expert from the Netherlands said that people might more readily complain to companies, but be reluctant to complain to public authorities. This is because public services often concern vulnerable people, in need of social benefits, who would be less inclined to complain about any decisions.

Opportunities to successfully complain about the use of AI and challenge decisions based on AI are essential for providing access to justice. The interviews emphasised the following as important in this respect:

- Making people aware that AI is used
- Making people aware of how and where to complain
- Making sure that the AI system and decisions based on AI can be explained

First, everyone needs to know if they are dealing with an AI system. If a taken decision affects people, e.g. on social benefits, people concerned might complain in general – but they will not be able to complain about the use of AI if they do not know AI is involved.

An expert explained that, while there is general willingness to complain, the biggest problem is that people often do not know that AI is being used, because organisations are not transparent about this, even though this is required by the GDPR. Several interviewees indicate that informing people that any decision made about them is based on (partly) automated tools is the very first step for providing access to complaints.

Second, everyone needs to know how and where to complain. It may be difficult for people to know which body deals with what type of complaints. One expert pointed out that consumers often do not know how to complain – for example, to a bank that might use algorithms for deciding on financial matters. A public administration that issues automated decisions decided to add names of employees to the decisions to provide contact persons to those potentially challenging the (automated) decision. Most interviewees indicated that there are ways and procedures for complaints in place, which are the same procedures as those for any other complaints not linked to the use of AI. Only few companies or organisations that use AI on anonymised or aggregated data indicate that they do not have any complaint mechanisms in place.

Finally, those complaining need enough information to challenge the underlying decision. Only thorough information about the AI systems provides equality of arms to meaningfully challenge decisions. However, this is not straightforward when it comes to the use of AI, particularly because of:

- potential intellectual property rights issues, and
- because complex systems are difficult to explain.

Intellectual property rights form one hurdle to providing enough information about how a decision was made, or how a system works. Algorithms can be part of an implemented software, or technical invention, that may be subject to intellectual property rights – a right protected under Article 17 (2) of the Charter. Actors often seek out copyright, patent and trade secret protection to safeguard their knowledge on AI.[60]

One interviewee from the insurance sector claims that, due to the highly competitive market, "one may not share too much about the workings of a used technology" as to, for instance, why a particular price was given to a customer. This is essentially because competitors could benefit from this knowledge if the underlying software were subject to scrutiny. Another respondent using AI to handle visa applications notes that using systems developed by external providers whose algorithms are covered by intellectual property rights can hinder the necessary transparency at a later stage.

Another challenge for successfully complaining about automated decisions or the use of AI in general is the **challenge to explain the decisions based on complex systems**. The interviewees working for public administration suggest that there is usually clear guidance on how to complain against an administrative decision, an area where interviewees highlight the importance of detailed explanations. For example, for the systems that automatically provide unemployment benefits for cases that do not involve discretion, clients can ask for the reasoning behind automated administrative acts. An interviewee indicates that if clients wish to see the calculations behind financial decisions, they may do so in the self-service system on the organisation's website or in their publications, which contain detailed descriptions of the calculations used.

Interviewees recognise that an open and transparent logic is essential for providing explanations regarding AI-supported decisions, but that this is often challenging or impossible to achieve. One interviewee working for a bank mentions that more complex machine learning solutions cannot be used for certain decision making, because the reasoning of the system cannot be explained easily, and this is why such systems are only used for other purposes. However, an interviewee working for another bank indicates that such systems are used, but they use simpler methods in addition to the complex ones to get an idea of the probable reasons for the decisions.

One expert raised the problem that companies internally might not have enough information about the way algorithms work themselves. The lack of expertise and knowledge appears to be a major hindrance in practice when seeking access to effective remedy.[61]

"The topic of transparency is very important nowadays, there are many procedures on how to publish the information, many automatic means that help to upload the information on the portals, and there has been a lot of work done in terms of transparency."
(Public administration, Spain)

Experiences from use cases

Respondents discussing **predictive policing** tools highlighted transparency as important.

In the gender-based violence use case, they felt that sending both the police file and the outcome of the AI system to the judge, and informing the victim of the level of risk attributed to the case and the police measures that will apply as a result, enhances transparency.

Interviewees discussing the heat map example referred to numerous requests to the police to explain the system's purpose and how it works, and highlighted transparency as a way to reduce public anxiety.

A number of interviewees pointed to the possibility for individuals affected by the system to make complaints to the police, the courts or the

Ombudsinstitution. With reference to the domestic violence case, however, the interviewee indicated that there is no procedure in place to question a system of police protocol.

In terms of measures to protect fundamental rights in the **health services** use cases, several interviewees referred to ethics committees, as well as general legal safeguards and data protection rules. Checks and controls were primarily mentioned to take place through external actors. No specific complaints procedures were in place in the organisations of those interviewees who responded to this question.

Some interviewees highlighted that doctors ultimately take responsibility for decisions, and that patients often do not know about the use of an AI tool in the first place. For example, in the breast cancer detection example, the interviewee indicated that there is no possibility for legal recourse against the developer of the tool, as the radiologist makes the decision on diagnosis and is liable for any errors.

The safeguards in place for the **targeted advertising** cases mainly follow data protection requirements, such as ensuring that consent is obtained and respected. One company makes sure not to have clients engaged in illicit practices and rejects clients from certain sectors, such as political advertising.

Complaints received

Few of the organisations interviewed received any complaints challenging their use of AI. In some cases, interviewees claim to have received complaints by complainants not aware that AI was used, who noticed incorrect outputs in decision making.

For example, individuals lodged complaints regarding traffic fines, whereby a police officer stopped a car driver, and upon hearing the car driver's explanation as to why the fine was wrongfully administered, proceeded to manually correct the information in the system, without being able to update the system's historical data. In these cases, such fines will remain visible throughout the system, and this particular person would continue to be profiled as a high risk on each occasion.

Even though organisations rarely received any formal complaints with respect to their use of AI, interviewees often state that this is due to the early stages of their AI implementation. Nonetheless, interviewees reported repeated requests for access to or rectification of personal data, and some people requested their information to be removed, as well as explanations as to why a certain recommendation was made.

"The number of the complaints about data use is miniscule, rather people may have asked to delete some information about them."
(Private company, Estonia)

The majority of interviewees claim that procedures are the same as to if a decision had been processed or undertaken by a human. On the other hand, a few other interviewees showed interest in opening new channels to analyse, explain and redress decisions involving their AI solutions.

Other rights linked to access to justice set out in the Charter are also impacted, most notably by the use of AI in law enforcement. These include, for example, the presumption of innocence (Article 48 of the Charter).

When identifying people who are suspected of having committed a crime, the police may target their activities specifically against one person or put them under suspicion based on flawed and fragmented data and algorithmic profiling.[62] Uncritical reliance on automated tools, without proper human review that takes into account other information, might contribute to discrimination in decision making.

4.7. RIGHT TO SOCIAL SECURITY AND SOCIAL ASSISTANCE

The right to social security and assistance enshrined in Article 34 of the Charter is a classic social right,[63] inspired by various international and European legal standards.[64] This provision, combining both elements of a right and of a principle,[65] has a great significance in the EU in view of the free movement of people within the Union.

Instead of tying issues of social protection to the labour market, this Charter right takes a new, communitarian approach when broadly referring to "providing [social] protection in cases such as maternity, illness, industrial accidents, dependency or old age, and in the case of loss of employment" (Article 34 (1)).[66]

It is, however, a primarily programmatic statement that does not prescribe any minimum standard of protection. It is in principle up to EU Member States to determine the conditions of entitlement and access to social benefits, with further clarification needed from the CJEU.[67] Yet, Article 34 (1) of the Charter provides protection against measures restricting or abolishing existing social security rights.[68]

In addition, access to social rights is guaranteed to all individuals legally residing within the EU who exercise their right to free movement, regardless of their nationality, subject to EU and national laws (Article 34 (2)). This thus creates justiciable rights before national courts and the CJEU.[69]

It is becoming increasingly apparent that the impact of AI technologies on social protection systems and the lives of the many individuals who rely upon them can be far-reaching and – potentially – very problematic. Introducing AI-driven technologies in social welfare systems risks creating barriers to access to this right.[70]

For example, using AI in social security needs to account for potential negative – and discriminatory – effects on non-nationals (both EU citizens and third-country nationals) exercising their right to freedom of movement in the EU. They could be negatively affected, for example, if a system relies on data about job histories, which are not available for those moving from other EU Member States.

Only one respondent addressed the 'right to receive a correct pension' as an aspect of a wider definition of human rights. Meanwhile, none of those interviewed referred to the fundamental right to social security and social assistance. This could partly reflect the nature of the use cases. However, the lack of references to social rights among public sector interviewees was notable.

4.8. CONSUMER PROTECTION

The Charter stipulates that EU policies must ensure a high level of consumer protection, which is based on Article 169 of the TFEU. EU institutions and other bodies needs to observe this principle, as do Member State authorities when implementing EU law.[71]

This Charter principle provides only for the guarantee of a particular goal ("a high level of consumer protection"). Article 169 of the TFEU is more concrete, as it also determines the means of how to achieve the stated aim – for example, protecting the health, safety and economic interest of consumers, as well as promoting their right to information and education. [72]

Among the use cases, the use of AI for targeted advertising, and the use of medical records by companies, are of particular importance.

When it comes to targeted advertising, consumers need to be aware that they can opt out from being targeted. If they are not aware, they might be subjected to advertising they do not want. This is particularly problematic in combination with highly sophisticated AI systems for advertising, which can amount to some sort of manipulation of consumer preferences.[73]

Consumer protection is also of major relevance for the use of health data (EHRs). The European Consumer Organisation (BEUC) noted that AI in the area of health brings challenges for consumers. It recommends that AI technologies must fully respect data protection rules, be transparent to the consumer, and avoid discrimination. BEUC has also called for updated regulation and legislative measures for market surveillance, law enforcement, and efficient redress concerning digital health products and services to fully protect EU consumers.[74]

BEUC carried out a survey among consumers about their views on AI in selected EU Member States. It shows that more than one in two respondents agree that companies are using AI to manipulate consumer decisions. In addition, almost half of respondents believe that personalised content and adverts on e-commerce platforms do not have an added value (44 %). Slightly more than half of the survey respondents expressed low trust that governments effectively control AI.[75]

In the interviews conducted for this study, consumer protection was only mentioned at the margins, when discussing risks of using AI and fundamental rights. However, some respondents from businesses refer to consumer protection legislation as a relevant framework also applying to their use of AI. Moreover, some respondents deem consumer protection authorities potentially relevant oversight bodies when AI is used.

In general terms, many interviewees in the business sector stress the importance of consumer satisfaction. For example, a company using video surveillance for the security of customers at their premises mention that consumer protection regulations are relevant for such technical solutions, and that the use of the systems should aim to improve the situation of consumers while also preserving their rights. Several AI tools are built to understand and profile consumers to enable businesses to improve their services and marketing.

Data protection is an important aspect for business. This is also linked to the fact that breaching data protection rules is considered a business risk, as mentioned above. One major concern of companies is obtaining and managing consent from consumers and customers to process their data, when using AI tools for marketing purposes. Interviewees report that the GDPR has had an impact, improving their systems to handle consent.

4.9. RIGHT TO GOOD ADMINISTRATION

The right to good administration is a well-established general principle of EU law elaborated by the CJEU. As such, it is binding on all EU Member States.[76] It is also a fundamental right enshrined in Article 41 of the Charter, although only for actions of EU institutions, bodies and agencies.[77]

As a general principle of EU law, it requires EU Member States to apply the requirements of the right to good administration in all public action. This right includes, but is not limited to, the right of an individual to have access to their file and the obligation of any public authority to give sufficient reasons for its decisions.[78]

Access to the file facilitates understanding of the evidentiary basis on which a decision has been made, and/or of the reasons underlying it. This places the individual in a better position to put forward counter-arguments when exercising the right to be heard and the right to an effective remedy.[79]

The obligation to give reasons makes, from the perspective of the individuals affected, the decision-making process more transparent, so that the person concerned can know and understand why a measure or action has been taken. Transparency is also an enabling principle that provides foundations for other rights,[80] including the exercise of the right to an effective remedy.

According to the CJEU, the context in which individual decisions are made is important in determining the extent of the duty to give reasons.[81] In France, for instance, the Code on the Relations between the Public and the Administration requires written explanations of the factual and legal considerations on which a decision has been based.[82]

The right to good administration also applies when AI systems process personal data and support decision making by public authorities. Although the right to good administration may be subjected to certain limitations, the question arises of how to ensure that the potentially huge number of individuals all have access to their files (personal data used in AI systems). Another question is how to make sure that public authorities always give sufficient reasons when the operation of AI-driven technologies cannot be fully explained due to their inherent opacity and complexity.

The use of a system to categorise unemployed people, set up in Poland, highlighted problems linked to public administration and the use of algorithms. Based on questions answered by unemployed people, a categorisation was developed through a statistical algorithm. The system received a lot of criticism from civil society with respect to the lack of opportunities to complain and potential discrimination.[83] In the end, a complaint by the Ombudsinstitution – based on administrative grounds – led to a constitutional court ruling that put an end to the system's use.[84]

The intent to increase efficiency drives the use of AI in the public sector – an aim that directly speaks to improving administration and benefiting citizens. Respondents in public administration by far most often indicate efficiency as the reason for considering the use of AI or for presently using AI. One respondent, who advises ministries on digital strategies and their use of AI, said that the main reasons for adopting AI are to improve the service to citizens and to reduce the costs of these services for public administration.

Interviewees also indicate that public administration has particular requirements, meaning AI cannot be used for all purposes and needs particular attention when it comes to decision making. However, the efficiency of a system is also considered an important added value.

In this sense, a respondent working on the digitalisation of migration management indicates that building too complex AI systems is a risk, because afterwards it would require a lot of work to understand the system in retrospect. The interviewee indicates that their team needs to be careful not to allow AI to make final decisions, which have to be taken by a human – because society and clients are not ready for this, according to the interviewee. Although some systems are appealing, they do not work effectively, and this could result in extra work and negative results. However, the interviewee also indicates that the dimension of efficiency "is often side-lined when discussing data protection".

The requirements for good administration also directly link the issues raised above with respect to data protection, non-discrimination and the right to an effective remedy and fair trial. Public administration can only process data on a legal basis. Decisions need to be fair and transparent and pathways to challenge decisions need to be available and accessible. As a result, the requirements for good administration are directly linked to the discussion and analysis above with respect to the legal processing of data (under data protection), fair decisions (linked to the discussion about non-discrimination), alongside transparency and ways to challenge and explain decisions (with respect to access to justice).

Endnotes

[1] See European Commission, **European enterprise survey on the use of technologies based on artificial intelligence**, Luxembourg, July 2020.

[2] FRA (2020), ***What do fundamental rights mean for people in the EU***, Luxembourg, Publications Office, p. 28.

[3] See **webpage on Three-level IT Baseline Security System ISKE** on ther website of Estonia's Information System Authority.

[4] See the **website of the PCI Security Standards Council**.

[5] Barak, A. (2019), 'Human dignity as a framework right (motherright)', in Barak, A., *Human Dignity: The Constitutional Value and the Constitutional Right*, Cambridge, Cambridge University Press, 2015, Ch. 9 (pp. 156-169).

[6] CJEU, C-377/98, *Netherlands v. European Parliament and Council*, 9 October 2001, paras. 70-77.

[7] For a discussion on the malicious use of AI, see for example, Brundage, M. et al. (2018), **The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation**.

[8] FRA (2019), ***Facial recognition technology: fundamental rights considerations in the context of law enforcement***, Luxembourg, Publications Office, November 2019.

[9] CJEU, Joined Cases C-92/09 and C-93/09, *Volker und Markus Schecke and Eifert GbR and Hartmut Eifert*, Opinion of Advocate General Sharpston, 17 June 2010, para. 71.

[10] FRA, Council of Europe and EDPS (2018), ***Handbook on European data protection law. 2018 Edition***, Luxembourg, Publications Office, June 2018, p. 19.

[11] See also *Ibid*., pp. 35-52.

[12] ECtHR (2019), ***Guide on Article 8 of the European Convention on Human Rights – Right to respect for private and family life, home and correspondence***, Strasbourg, Council of Europe, updated on 31 August 2019, paras. 133 and 136.

[13] ECtHR*, López Ribalda and Others v. Spain*, Nos. 1874/13 and 8567/13, 17 October 2019, para. 87. For a comprehensive legal analysis of the meaning and content of 'privacy', see also Koops, B.-J. et al. (2017), '**A Typology of Privacy**', *University of Pennsylvania Journal of International Law*, Vol. 38, Issue 2, pp. 483-575.

[14] Vermeulen, M. (2015), ***SURVEILLE Deliverable D4.7 – The scope of the right to private life in public places***, July 2014, p. 2.

[15] UN, Human Rights Committee, ***General Comment No. 37 (2020) on the right of peaceful assembly (article 21)***, CCPR/C/GC/37, 17 September 2020, para. 62.

[16] Costello, Róisín Áine (2020), **The Impacts of AdTech on Privacy Rights and the Rule of Law**, Technology and Regulation.

[17] Norwegian Consumer Council (2020), **Out of Control. How consumers are exploited by the online advertising industry**.

[18] FRA (2020), *Your rights matter: Data protection and privacy – Fundamental Rights Survey*, Luxembourg, Publications Office.

[19] Rocher, L., Hendrickx, J. M. and de Montjoye Y. (2019), **Estimating the success of re-identifications in incomplete datasets using generative models**, *Nature Communications* 10, No. 3069.

[20] Hacker, P. (2020), *A Legal Framework for AI Training Data. Law, Innovation and Technology* (forthcoming), available at **SSRN**; Article 29 Data Protection Working Party (2014), **Opinion 05/2014 on Anonymisation Techniques**; see also Finck, Michèle and Pallas, Frank, *They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data Under the GDPR* (October 1, 2019), Forthcoming, International Data Privacy Law, 2020, Max Planck Institute for Innovation & Competition Research Paper No. 19-14, available at **SSRN**; and Sartor G. and Lagioia F. (2020), *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, study prepared for the panel for the Future of Science and Technology (STOA) of the European Parliament.

[21] See, for example, the UK Data Service's **blog on "Access to sensitive data for research: 'The 5 Safes'**"; see also the discussion in Ohm, P. (2010), "Broken promises of privacy: responding to the surprising failure of anonymization", *UCLA Law Review*, p. 1701.

[22] GDPR, Art. 22 (3); and Law Enforcement Directive, Art. 11 (1).

[23] Veale, M. and Edwards, L. (2018), 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling', *Computer Law & Security Review*, Vol. 34 (2), April 2018, pp. 398-404.

[24] Article 29 Data Protection Working Party (2018), **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679**, Adopted on 3 October 2017, as last Revised and Adopted on 6 February 2018.

[25] Green, B. And Chen, Y. (2019), 'Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments', *In FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19)*, January 29-31, 2019.

[26] González Fuster, G. (2020***), Artificial Intelligence and Law Enforcement – Impact on Fundamental Rights***, European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, PE 656.295, July 2020, p. 17; Brkan, M. (2019), 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond', *International Journal of Law and Information Technology*, Vol. 27 (2), p. 98; Article 29 Working Party, ***Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679***, Adopted on 3 October 2017, as last Revised and Adopted on 6 February 2018, WP251rev.0, p. 19.

[27] Misuraca, G., and van Noordt, C. (2020), **Overview of the use and impact of AI in public services in the EU**, European Commission Joint Research Centre, Luxembourg.

[28] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180, 19.7.2000, pp. 22-26, Art. 2; and Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation, OJ L 303, 2.12.2000, pp. 16-22, Art. 2.

[29] FRA and CoE (2018), ***Handbook on European non-discrimination law. 2018 edition***, Luxembourg, Publications Office, June 2018, p. 35.

[30] CJEU, C-236/09, *Association Belge des Consommateurs Test-Achats ASBL and Others v. Conseil des ministres*, 14 January 2011.

[31] European Commission (2012), EU rules on gender-neutral pricing in insurance industry enter into force, Press release, **IP/11/1581**, 20 December 2012.

[32] Elizabeth E. Joh (2015), '**The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing**', *UC Davis Legal Studies Research Paper No. 473*, pp. 17-18.

[33] Aleš Završnik (2019), 'Algorithmic justice: Algorithms and big data in criminal justice settings', *European Journal of Criminology*, p. 14. DOI: 10.1177/1477370819876762.

[34] See also European Commission, ***White Paper On Artificial Intelligence – A European approach to excellence and trust***, COM(2020) 65 final, Brussels, 19 February 2020, p. 1.

[35] FRA (2018), ***#BigData: Discrimination in data-supported decision making***, Luxembourg, Publications Office, June 2018, p. 3.

[36] *Ibid.*

[37] FRA (2019), ***Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights***, Luxembourg, Publications Office.

[38] Korff, D. and Browne, I. (2013) '**The use of the Internet & related services, private life & data protection: trends, technologies, threats and implications**', Council of Europe, T-PD(2013)07.

[39] See National Non-discrimination and Equality Tribunal of Finland, **decision no. 216/2017** from 21 March 2018. See also the SyRI case discussed above and UK, Court of Appeal, ***R (Bridges) v. CC South Wales***, [2020] EWCA Civ 1058, 11 August 2020.

[40] See also Equinet (2020), ***Regulating for an equal AI: A new role for equality bodies***, Brussels, report prepared by Allen R. and Masters D.

[41] Tolan S., Miron M., Gomez E. and Castillo C (2019), 'Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia', Best Paper Award, International Conference on AI and Law, 2019; Richardson R., Schultz J. and Crawford K. (2019), Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, *94 N.Y.U. L. REV. ONLINE 192 (2019)*, available at **SSRN**.

[42] FRA (2014), **Violence against women: an EU-wide survey. Main results report**, Luxembourg, Publications Office, p. 61.

[43] FRA (2018), ***Second European Union Minorities and Discrimination Survey. Main results***, Luxembourg, Publications Office, p. 66.

[44] Erik Bakke (2018), "Predictive policing: The argument for public transparency", *New York University Annual Survey of American Law*, Vol. 74, pp. 139-140; Andrew G. Ferguson (2017), '**Policing Predictive Policing**', *Washington University Law Review*, Vol. 94, pp. 1146-1150; andCouncil of Europe Committee of experts on internet intermediaries (MSI-NET) (2017), **Algorithms and Human Rights**, Council of Europe DGI(2017)12, p. 11.

[45] Elizabeth E. Joh (2015), '**The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing**', *UC Davis Legal Studies Research Paper No. 473*, p. 18.

[46] Gstrein, O. J., Bunnik, A., and Zwitter, A. (2019), 'Ethical, Legal and Social Challenges of Predictive Policing', *Católica Law Review* 3:3, pp. 77-98; Albert Meijer & Martijn Wessels (2019), 'Predictive Policing: Review of Benefits and Drawbacks', *International Journal of Public Administration* 42:12, p. 1036, DOI: 10.1080/01900692.2019.1575664.

[47] Aleš Završnik (2019), 'Algorithmic justice: Algorithms and big data in criminal justice settings', *European Journal of Criminology*, pp. 8-9. DOI: 10.1177/1477370819876762.

[48] Wachter, Sandra (2020), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising', *Berkeley Technology Law Journal*, Vol. 35, No. 2, 2020, (forthcoming), available at **SSRN**.

[49] On the use of AI in financial industries leading to unequal access to financial services, see in the legal literature e.g. Boyd, D., Levy K. & Marwick, A. (2014), 'The Networked Nature of Algorithmic Discrimination' in Gangadharan, S. P., Eubanks, V. & Barocas, S. (eds), ***Data and Discrimination: Collected Essays***, Open Technology Institute, pp. 53-62.

[50] For an overview of child rights issues, see UNICEF Innovation, Human Rights Center, UC Berkeley (2019), **Artificial Intelligence and Children's Rights**.

[51] EU Network of Independent Experts on Fundamental Rights, ***Commentary on the Charter on Fundamental Rights of the European Union***, June 2006, p. 360. See also: FRA and CoE (2016**), *Handbook on European law relating to access to justice***, Luxembourg, Publications Office, June 2016, p. 92.

[52] CJEU, C-432/05, *Unibet (London) Ltd, Unibet (International) Ltd v. Justitiekanslern*, 13 March 2007, para. 37; CJEU, C-93/12, *ET Agrokonsulting-04-Velko Stoyanov v. Izpalnitelen direktor na Darzhaven fond 'Zemedelie' – Razplashtatelna agentsia*, 27 June 2013, para. 59; CJEU, C-562/13, *Centre public d'action sociale d'Ottignies-Louvain-la-Neuve v. Moussa Abdida*, 18 December 2014, para. 45.

[53] Law Enforcement Directive, Art. 54; and GDPR, Art. 79.

[54] Law Enforcement Directive, Art. 53; and GDPR, Art. 78.

[55] Law Enforcement Directive, Art. 52; and GDPR, Art. 77.

[56] Council of Europe, ***Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*** (adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies), Appendix, para. B.4.5.

[57] Andrew G. Ferguson (2017), '**Policing Predictive Policing**', 94 *Washington University Law Review*, pp. 1165-1167.

[58] Gstrein, O. J., Bunnik, A., & Zwitter, A. (2019), Ethical, Legal and Social Challenges of Predictive Policing', *Católica Law Review*, 3:3, pp. 80-81; Yeung K. (2019), **A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework**, Prepared by the Council of Europe Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT).

[59] Council of Europe, *Algorithms and human rights*, pp.11 and 24.

[60] International Technology Law Association (2019), '**Responsible AI: A Global Policy Framework**', pp. 258-282.

[61] The lack of expertise on AI is also reflected in the survey among companies in the EU, where the lack of skills among existing staff and difficulties in hiring new staff are the most prominent obstacle for further AI adoption (European Commission (2020), **European enterprise survey on the use of technologies based on artificial intelligence**, Luxembourg, July 2020, p. 11).

[62] See a detailed assessment of the impact of predictive policing on the presumption of innocence in Mendola, Marco (2016), **One Step Further in the 'Surveillance Society': The Case of Predictive Policing**.

[63] See e.g. Egorov, A. and Wujczyk, M. (eds.) (2016), ***The Right to Social Security in the Constitutions of the World: Broadening the moral and legal space for social justice***, Geneva, ILO Global Study, Vol. 1: Europe, pp. xv-xvii and 1-6.

[64] These include Arts. 153 and 156 of the TFEU; Arts. 12 and 13 of the 1961 European Social Charter; as well as points 2 and 10 of the 1989 Community Charter on the Fundamental Social Rights of Workers (see ***Explanations relating to the Charter of Fundamental Rights***, OJ C 303, 14.12.2007, pp. 17-35).

[65] ***Explanations relating to the Charter of Fundamental Rights*** (OJ C 303, 14.12.2007, pp. 17-35), Explanation on Article 52 — Scope and interpretation of rights and principles.

[66] Łukasz Bojarski, Dieter Schindlauer and Katrin Wladasch (2014), ***The European Charter of Fundamental Rights as a Living Instrument – Manual***, Rome/Warsaw/Vienna, pp. 61-62.

[67] De Becker, E. (2016), '**The (Possible) Role of the Right to Social Security in the EU Economic Monitoring Process**', *German Law Journal*, Vol. 17, No. 3, pp. 297, 304; Paju, J. (2017), *The European Union and Social Security Law*, Oxford, Hart Publishing, sub-section 7.5.2.

[68] *Ibid.*, pp. 297-298; Peers, S. & Prechal, S. (2014), 'Scope and Interpretation of Rights and Principles', in Hervey, T., Kenner, J., Peers, S. and Ward, A. (eds.), *The EU Charter of Fundamental Rights. A Commentary*, Oxford and Portland, Oregon; Hart Publishing, 2014, pp. 1455, 1508.

[69] With the exception of Poland and the United Kingdom, see Protocol (No. 30) on the application of the Charter of Fundamental Rights of the European Union to Poland and to the United Kingdom (OJ C 115, 9.5.2008, pp. 313-314), Art. 1 (2).

[70] Christiaan van Veen and Ben Zevenbergen, '**Conference on Social Protection by Artificial Intelligence: Decoding Human Rights in a Digital Age**', *Freedom to Tinker – Research and Expert Commentary on Digital Technologies in Public Life*, 29 May 2019.

[71] Art. 51 (1) of the Charter; see also ***Explanations relating to the Charter of Fundamental Rights*** (OJ C 303, 14.12.2007, pp. 17-35), Explanation on Article 52 — Scope and interpretation of rights and principles.

[72] Łukasz Bojarski, Dieter Schindlauer and Katrin Wladasch (2014), ***The European Charter of Fundamental Rights as a Living Instrument – Manual***, Rome/Warsaw/Vienna, p. 67.

[73] Sartor, Giovanni (2020), **New aspects and challenges in consumer protection**, study for the committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg.

[74] The European Consumer Organisation (BEUC) (2018), **Digital Health, Principles and Recommendations**.

[75] BEUC (2020), **Artificial Intelligence: what consumers say. Findings and policy recommendations of a multi-country survey on AI**.

[76] In recent case law, see CJEU, C-604/12, *H. N. v. Minister for Justice, Equality and Law Reform, Ireland, Attorney General*, 8 May 2014, para. 49.

[77] Also confirmed by the CJEU, Joined Cases C-141/12 and C-372/12, *YS v. Minister voor Immigratie, Integratie en Asiel, and Minister voor Immigratie, Integratie en Asiel v. M, S*, 17 July 2014, paras. 66-70).

[78] These components, initially developed by the CJEU case law, have been codified in Article 41 (2) of the Charter. For more on this right in leading academic literature, see Craig, P. (2014), 'Article 41 – Right to Good Administration', in Hervey, T., Kenner, J., Peers, S. and Ward, A. (eds.), *The EU Charter of Fundamental Rights. A Commentary*, Oxford and Portland, Oregon; Hart Publishing, 2014, pp. 1069-1098.

[79] *Ibid*., p. 1082.

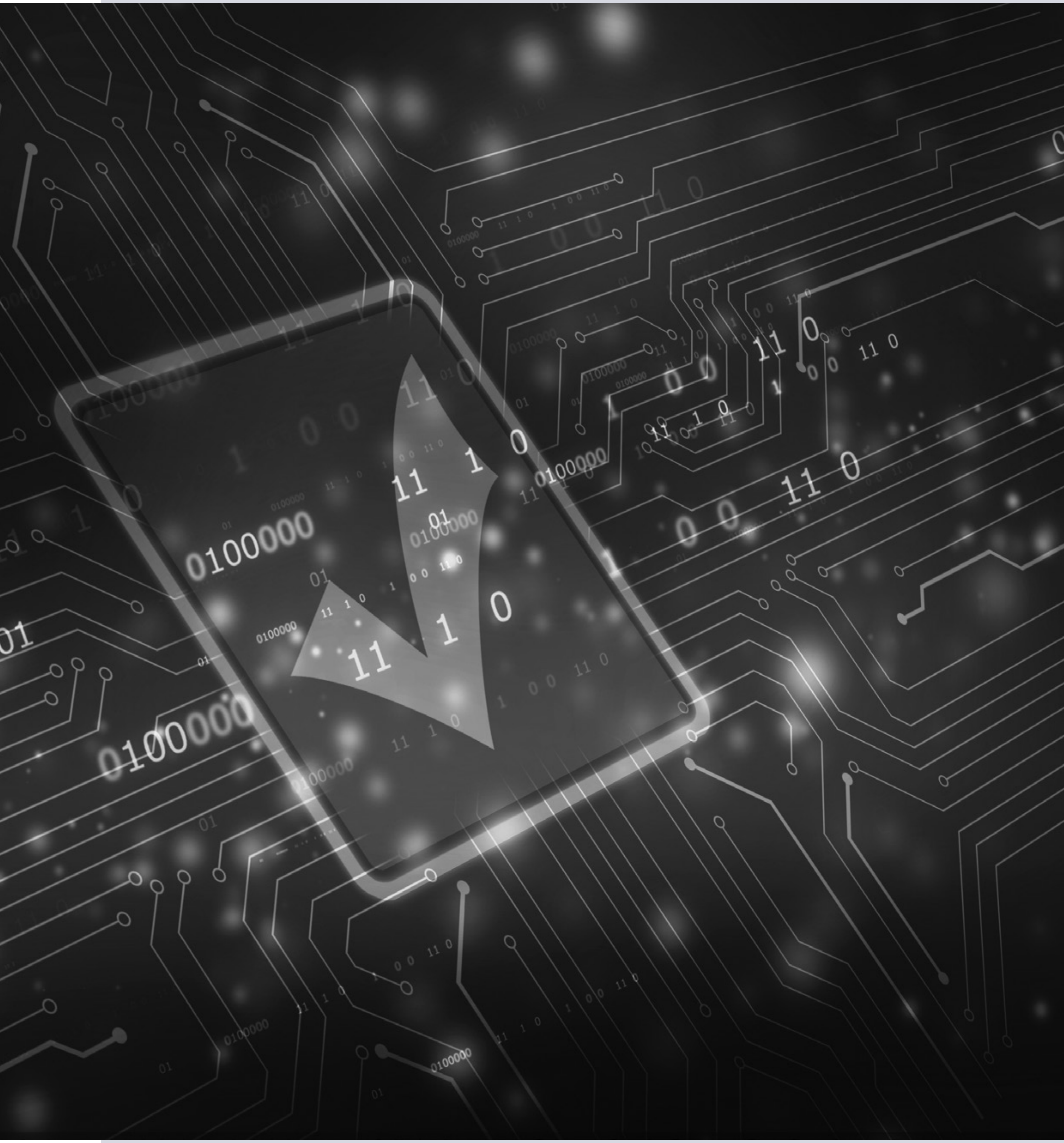
[80] Finck, M. (2019), '**Automated Decision-Making and Administrative Law**', *Max Planck Institute for Innovation & Competition Research Paper No. 19-10*, p. 8.

[81] Craig, P. (2014), 'Article 41 – Right to Good Administration', in Hervey, T., Kenner, J., Peers, S. and Ward, A. (eds.), *The EU Charter of Fundamental Rights. A Commentary*, Oxford and Portland, Oregon; Hart Publishing, 2014, pp. 1086-1087.

[82] France, ***Code des relations entre le public et l'administration***, Article L2111-5.

[83] Panoptykon Foundation (2015), **Profiling the unemployed in Poland: Social and political implications of algorithmic decision making**; see also Algorithm Watch (2019), ***Poland to scrap controversial unemployment scoring system***.

[84] See **Decision K 53/16**, available on the Constitutional Tribunal's website.



5

FUNDAMENTAL RIGHTS IMPACT ASSESSMENT – A PRACTICAL TOOL FOR PROTECTING FUNDAMENTAL RIGHTS

Chapter 4 illustrated the extent to which using AI affects different fundamental rights. This chapter analyses how fundamental rights impact assessments (FRIA) could reduce the negative impacts that using AI can have on fundamental rights.

Section 5.1 provides a brief overview of the current discussion on the need for fundamental rights impact assessments in this field. **Section 5.2** analyses current practices in addressing fundamental rights implications, based on the interviews conducted for this report. Interviewees were asked about what sort of testing was done before the system was used, and who controls the tasks affected by the use of the technology.

The chapter ends with suggestions on how to assess the fundamental rights impact when using AI and related technologies.

5.1. CALLING FOR A FUNDAMENTAL RIGHTS IMPACT ASSESSMENT – AVAILABLE GUIDANCE AND TOOLS

International organisations,[1] academics[2] and civil society[3] have called for fundamental rights impact assessments to be conducted when using AI or related technologies.

For example, the Committee of Ministers of the Council of Europe's guidelines on addressing the human rights impacts of algorithmic systems recommend that states should conduct "impact assessments prior to public procurement, during development, at regular milestones, and throughout their context-specific deployment in order to identify the risks of rights-adverse outcomes".[4]

There is a need for flexible impact assessments that can adapt to different situations given that fundamental rights violations are always contextual. Scholars exemplify this based on EU anti-discrimination law, where equality is always contextual and depends on the case at hand.[5]

Fundamental rights compliance cannot be automated and hard-coded into computer software. Rather, each use case needs separate examination to determine whether any fundamental rights issue arises. Nevertheless, assessments can follow a systematic approach and provide similar information.

Existing standards provide guidance on how to do a fundamental rights impact assessment of AI and related technology. These include hard law, soft law instruments (such as recommendations or declarations), and practical tools (e.g. guidelines and checklists).

Beyond the requirements flowing from data protection legislation (see box), there are few examples of laws requiring mandatory assessments of the effects of AI in general. In view of the increasing uptake of AI, the Canadian government has issued guidelines, including mandatory requirements for assessing AI for use by public administration. It applies to any system, tool, or statistical model used to recommend or make an administrative decision about a client.[6]

Learning from data protection impact assessments

European data protection law requires a data protection impact assessment (DPIA).[a] The CoE Modernised Convention No. 108 provides for a general obligation to examine the likely impact of data processing on individuals' rights and fundamental freedoms before their use. Following the assessment, controllers should design the processing in such a manner to prevent or minimise identified risks.[b]

EU law imposes a similar, more detailed, obligation. The GDPR foresees a Data Protection Impact Assessment (DPIA) for data processing that is likely "to result in a high risk to the rights and freedoms of natural persons."[c] Therefore, where required by law, a DPIA for an AI technology could potentially also address the broader fundamental rights implications, besides the impact on the right to privacy,[d] and be used as a tool to further investigate algorithms and their impacts.[e]

However, under the GDPR (Article 35), the DPIA is limited to 'high risk' cases processing personal data. It therefore may miss other high risk cases that are not primarily or obviously related to protection of personal data. At the same time, the GDPR is delimited to its specific field of application, with accompanying expertise in this field. This means that the potential extension of the scope of a DPIA to other fundamental rights might be limited.

The GDPR also gives some indications about the modalities to undertake a DPIA. First, a DPIA should be conducted before any high risk processing.[f] Second, a DPIA should provide for a systematic description of envisaged operations, the purpose and the legitimate interests pursued. It must also assess the necessity and proportionality of the processing and the possible risks to the rights of individuals. In addition, it must contain the planned security measures to address the risks identified.[g]

While pointing out that different methodologies can apply, the Article 29 Working Party (WP 29) proposes – in a check list form – minimum criteria that a controller should use to assess if the DPIA comprehensively complies with the GDPR.[h]

Finally, the GDPR foresees prior mandatory consultation of the relevant supervisory authority, if the impact assessment indicates that processing presents risks that cannot be mitigated. [i] This gives a crucial role to DPAs, as independent bodies established by law.[j]

The European Data Protection Supervisor (EDPS) provides guidance on carrying out DPIAs.[k] Data protection authorities have also discussed, and provide guidance on, how to assess AI technologies.[l]

[a] *For more information on Data Protection Impact Assessment, see: FRA, Council of Europe and EDPS (2018),* ***Handbook on European data protection law. 2018 edition****, p. 179-181.*

[b] ***Council of Europe Modernised Convention No. 108****, Art. 10 (2).*

[c] *GDPR, Art. 35 (1).*

[d] *GDPR, Recitals (2) and (75); Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA), wp248rev.01, 13 October 2017.*

[e] *Edwards and Veale (2018); FRA (2018),* ***#BigData: Discrimination in data-supported decision making****, Luxembourg, Publications Office, June 2018.*

[f] *GDPR, Art. 35 (1). The WP29 specifies that 'carrying out a DPIA is a continual process, not a one-time exercise.'*

[g] *GDPR, Art. 35 (7), as well as recitals (84) and (90).*

[h] *Article 29 Working Party,* ***Guidelines on Data Protection Impact Assessment (DPIA)****, wp248rev.01, 13 October 2017, Annex 2.*

[i] *GDPR, Art. 36.*

[j] *GDPR, Art. 35.*

[k] *EDPS (2019),* ***Accountability on the ground Part II: Data Protection Impact Assessments & Prior Consultation****, v.1.3, July 2019.*

[l] *See, for example, the* ***Declaration on ethics and data protection in AI****, adopted by the 40th International Conference of Data Protection and Privacy Commissioners (ICDPPC), in 2018.*

There are many more examples of non-binding guidelines. At the global level, the United Nations Guiding Principles on Business and Human Rights recommend that enterprises integrate the findings from human rights impact assessments across relevant internal functions and processes, and take appropriate action.[7] Although they do not refer specifically to AI, the guidelines are relevant in supporting the development of AI technology in a rights compliant manner.[8]

At the EU level, the Ethics Guidelines for Trustworthy AI prepared by the European Commission's High-Level Group on Artificial Intelligence[9] also recommend performing a FRIA, before a system's development, when "there are risks that fundamental rights can negatively be affected by the technology".[10] They also emphasise the need to put in place mechanisms to receive external feedback on AI systems that potentially infringe on fundamental rights.

In addition, private companies, [11] associations of private companies[12] or of both public and private interests,[13] as well as NGOs[14] and other organisations[15] have developed different types of guidance to support AI impact assessments. These documents do not usually contain clear guidelines on impact assessment. Instead, they highlight the different aspects and criteria that should be taken into account when developing and carrying out an impact assessment.

Broad categories include the purpose of the system, the description of the technology, the assessment of the impact and targeted population/individual, evaluating fairness and diversity, the description of the audits planned or performed, as well as accountability. Some explicitly refer to applicable international human rights law standards.[16]

Various codes of ethics or conducts,[17] standards,[18] as well as certification schemes are also in place.[19]

Several practical tools are available to assess the impact of AI technologies and mitigate risks, developed by a wide range of actors. These include checklists,[20] lists of questions,[21] online self-evaluation tools,[22] and risk management frameworks.[23]

Some focus specifically on assessing fundamental rights risks.[24] Others focus on ethical, societal or economic implications.[25] These can be useful references when performing a thorough fundamental rights impact assessment of AI technologies.

In July 2020, for example, the High-Level Group on Artificial Intelligence issued an "Assessment List for Trustworthy AI" (ALTAI),[26] after a six month pilot involving more than 350 stakeholders. ALTAI helps organisations self-evaluate – on a voluntary basis – the reliability and trustworthiness of AI and reduce potential risks for users. It supports businesses and public administrations to ask the right questions around the seven requirements for responsible AI identified in the Ethics Guidelines for Trustworthy AI.[27] ALTAI specifically refers to the need to perform a fundamental rights impact assessment. It includes examples of questions to assess impact on non-discrimination/equality; the right to privacy; the rights of the child; the freedom of expression; as well as the freedom of information and association.[28]

Several online assessment tools target the use of AI by public authorities. The Canadian Government developed an Algorithmic Impact Assessment tool (AIA),[29] pursuant to the Canadian Directive on Automated Decision-Making.[30] The AIA represents an automated assessment consisting of more than 50 questions that unfold the requirements of the directive. Questions relate to fundamental rights concerns – such as an AI system's impact on the freedom of movement, on the likelihood of incarceration of an individual, on the legal status, on access to funding or benefits, or on indigenous people. A score is attributed to each reply and a final impact scoring is provided and made publically available on the government's website.

As another example, the Ethics Toolkit[31] is a freely accessible tool designed for local governments. Based on a risk management approach, it supports fair automated decisions and minimising unintentional harm to individuals in the field of the criminal justice, higher-education, social media and other areas.

Among national human rights bodies, the Danish Institute for Human Rights proposed a human rights compliance "quick check.[32] This involves an interactive online computer programme that allows companies to select and modify the information in a database to suit their type of business and area of operations to check rights compliance. The quick check is based on the Human Rights Compliance Assessment tool,[33] which runs on a database of over 350 questions and 1,000 corresponding human rights indicators. It uses international human rights law standards as benchmarks. Applying to all fields of operations, it can provide guidance when developing impact assessment for AI technology.

Academic work has also suggested operational frameworks for assessing risks in using AI technology. Some focus specifically on identifying and addressing fundamental rights implications by the private sector.[34] Some focus on developing ethical and values-oriented models (analysing the societal impact of the data used) with the creation of ad hoc expert (review) committee.[35]

Others have developed guidance frameworks for specific case studies. For example, in the field of criminal justice, the ALGO CARE framework[36] introduced a step-by-step assessment to evaluate the key legal and practical concerns that should be considered in relation to police using algorithmic risk assessment tools.

Some have argued for participatory ways to involve and consider the views of the affected rights-holders and other stakeholders communities when developing an impact assessment and publically engage with them from the start.[37] Others have joined cross-discipline expertise of science and law to design practical frameworks.[38]

5.2. IMPACT ASSESSMENTS AND TESTING IN PRACTICE

Virtually all the systems discussed in the interviews were subject to some sort of testing, which included elements of impact assessment. However, these were mainly technical and data protection (impact) assessments. These rarely address potential impacts on other fundamental rights.

Some interviewees argue against conducting a fundamental rights impact assessment, because in their view the system does not negatively affect fundamental rights or because they are unsure about it. For example, a respondent working on traffic management, using cameras for monitoring traffic, indicated that they only tested for accuracy of the system, but not fundamental rights, apart from respecting data protection rules.

Some respondents simply did not know if fundamental rights were assessed as part of a general impact assessment that was carried out.

Testing and stages of development

Much testing is done before any new AI system is used. As respondents highlighted, moving an AI system into production is a very challenging task. As mentioned, public administration as well as private companies are usually careful when using AI. Many projects that interviewees refer to are still in development or in the pilot phase, and some had not started concrete testing.

"When testing the system, we did not really look at the legal aspects, we looked at whether the system is profitable."
(Private company, Estonia)

Testing can be done in several stages. These include the development stage (so-called proof-of-concept), pilot stages before deployment, and tests during and after deployment. If possible, live experimentation is carried out at the initial stages, which often involves staged deployment.

For example, the organisation interviewed that tests different applications to support job seekers conducts continuous, step-by-step testing. Selected members of the organisation test the tool in real situations, using check lists. The interviewee mentioned that it is challenging to move to the deployment stage and it is planned to supervise the tool in real time.

In another example, involving automated rule-based granting of social benefits, different assessments were carried out. Before implementation, a group of lawyers, data protection specialists, compensation specialists and accountants performed a general impact assessment. After this, the department responsible for using the system conducted tests to decide whether the system could be used.

Following this, the system was monitored in its implementation, using a step-by-step approach. In a first step, about half of the decisions were taken by the system. In a next step, the decisions taken automatically were expanded to all negative decisions. After this, another area of decisions was added, including all decisions on ending compensation payments. At the time the interviews were conducted, about 95 % of decisions were automated. The interviewee indicated that, after carrying out these tests, they feel sure that the system is secure, and that there are no outstanding risks.

A company working on a fraud detection system replaced their rule-based system with a machine learning tool. Before changing the system, the old and new system were run in parallel to see if the machine learning system performs better than the rule-based one. The interviewee mentioned that "[there] was rigorous analysis behind it and direct feedback where we saw what would be the impact on losses versus how many good customers we were impacting negatively". The interviewee added that, when they "were comfortable that [the machine learning system] was better [than the static rule system] in all aspects, we deployed it in its entirety".

In other use cases, no previous automated system existed and tests were reviewed by humans. For example, an automated transcription service was tested during court hearings, when allowed by the judge. This included regular feedback on the correctness of the transcription services from judges.

One interviewee from law enforcement, working on a tool to detect domestic violence, identifies issues with precision and accuracy when using the system. If a police officer does not have sufficient training and knowledge about the system, the indicators required by the system cannot gather the required information, which could lead to miscalculation. They highlight that the robustness of the system is tested annually to assure the quality of the two questionnaires used, the completeness of the data, and the training of the police officers using the AI system. This process also considers how personal data protection laws and protocols are applied. The tests discussed focus strongly on technical aspects and general operations.

Fundamental rights and data protection impact assessments

Apart from data protection, which all respondents mentioned, other fundamental rights were typically not considered. Respondents only reflected about other potential impacts on fundamental rights, or mentioned that these aspects were considered, when prompted by the interviewer.

Many respondents are generally aware of discrimination issues – but often discussed this only after being explicitly asked about discrimination. Yet they gave no information about any formal, in-depth tests for discrimination. Generally, respondents ruled out the possibility that their system discriminates based on protected attributes. For example, one interviewee states that they test the system against data protection laws and specific applicable legal acts, but not fundamental rights. However, the interviewee did consider potential discrimination, but ruled it out. It needs to be kept in mind for future technologies, the interviewee stated.

However, there are cases where non-discrimination was generally considered during the testing phase of AI systems. One respondent from a municipal authority mentioned that they cannot assess the fairness of a model, because they cannot access data needed due to data protection reasons. According to the interviewee, "there is a huge tension surrounding the GDPR. So we want to do well, but might in fact be worse off, because interpretation of the data then turns out to be impossible".

Most respondents reported that a data protection impact assessment, as required by law, was conducted, although these took different forms. A bank tested a tool for analysing speech from customer calls to find out about reoccurring problems, and carried out a data protection impact assessment

"Yes, we assess the legality of personal data protection and the conformity with their specific legal acts."
(Public administration, Estonia)

(DPIA) specifically for testing the tool. The outcome was that the system can be tested if data were only used for the testing phase and are deleted after a certain period after the test, and if access to the data by employees is restricted to the testing phase and supervised. For the deployment of the tool, another DPIA is required in this case.

There is sometimes a lack of clarity as to what extent the use of AI and related technologies, most notably the use of algorithms, belongs to an DPIA. In the area of predictive policing, for instance, some DPIAs were done for the underlying architecture of the system, rather than the specific AI tool. Another interviewee using algorithms in financial services also mentioned not assessing the machine learning tool as such within the framework of a DPIA, because of the belief that it does not apply to the machine learning system (but the underlying data).

One interviewee felt that the data protection impact assessment for the crime heat map example was not sufficiently in depth to safeguard the quality of the model, and that the system was not equipped to deal with cross-sectoral use of data, where different rules might apply. They indicated that further standards were required.

A respondent working on migration management indicated having data protection officers involved in their analysis. The legal service has a specialised quality control AI tool to study the data protection aspects of their system. However, the respondent also mentioned that more guidance is needed.

The companies working on targeted advertising all looked into data protection issues, although not all respondents were sure if an impact assessment was conducted. The companies assessed, for example, whether only people who consented are approached in targeted communication. For targeted ads, they assessed whether information on possible re-identification is deleted, including whether cookies and trackers are anonymised.

With respect to DPIAs generally, some respondents did not know, as this was not their area of responsibility. Others knew they had a positive DPIA, but were not aware of any details. It appears that the legal assessment is sometimes detached from the technical side, with technical people not knowing about legal assessments. One interviewee from a private company working on credit risk scoring mentioned: “I make suggestions how some system could be developed and then the compliance manager tells me if it is in conformity with the laws”.

Audits and working with external (oversight) bodies

The public administrations and private companies involved in FRA’s research all carry out tests before deploying any AI. These are often linked to existing internal and external oversight processes. The use of AI is frequently subjected to internal review processes within companies and public administration, although these are not necessarily formalised review processes. Some interviewees mentioned that they are working on formalising existing internal review processes for overseeing AI systems.

Interviewees from the public sector say that they have to be particularly cautious before using any AI to support decisions. A representative working on migration management at a public administration indicates that "[i]n the private sector, [wrong results] might cause business-related losses, in the police it impacts people's lives and their fundamental rights".

Yet it is not always clear to public administration, or to businesses, who is responsible for checking and overseeing the use of AI. Public administrations appear to be under stronger scrutiny when it comes to oversight of their AI systems. Such oversight is often done through regular audits, for example connected to budgetary review.

Some interviewees, from public and private organisations, report that their AI systems are currently checked in the framework of an existing IT review (e.g. regular database checks), in the absence of review processes that specifically look into the use of AI. In addition, interviewees report about sector-specific certification schemes that also look into the use of AI – for example, in the area of health or financial services.

Several interviewees mentioned that they were in contact with data protection authorities. Some companies and public administrations sought permission from the data protection authorities before using their AI system or at least were generally in contact with them. For example, one company working on targeted advertising mentioned discussing their use of personal data with the national data protection authority.

Experts interviewed for this report further highlighted the relevance of data protection authorities for overseeing AI systems with respect to the use of personal data. However, experts strongly highlighted that data protection authorities are under-resourced for this task for two reasons. Data protection authorities often do not have relevant AI-related expertise.[39] Additionally, their budgets are overstretched and their workload heavy.

Experts' views differ with respect to the need of additional oversight bodies, and the potential creation of an AI specific institution. However, they agree that existing bodies all have to work on topics linked to AI within their mandates.

Equality bodies, as well as other human rights institutions, are mentioned by some interviewed experts as providing oversight concerning possible discrimination when using AI. They highlighted that these institutions need to build up expertise in this area to better contribute to the oversight of AI. However, similar to data protection authorities, this is a challenging task for equality bodies given their lack of resources.

"We are proactive not only among ourselves to mitigate risks, but we also get additional audits. We also see sometimes that some regulatory audits are quite sloppy. For us that is not good because we have lots of customer data."
(Private company, Estonia)

Several interviewees mentioned consumer protection authorities as potentially providing relevant oversight on the use of AI. One respondent, working for a retail company, would like to have an advisory agency that could be consulted about possible use of AI for innovation without being investigated right away. At the moment, the company prefers to consult consumer authorities over data protection authorities about potential future marketing campaigns. This is because data protection authorities might start an investigation into their efforts.

When discussing oversight, those developing and using AI, as well as experts, repeatedly mention the challenge to really understand the impact when using AI. Despite the need to engage existing oversight bodies, responsibilities to oversee the use of AI from a fundamental rights perspective remain unclear.

5.3. FUNDAMENTAL RIGHTS IMPACT ASSESSMENT IN PRACTICE

Many key actors in the field of fundamental rights have called for conducting fundamental rights impact assessments before using any AI-driven systems. This section highlights some of the elements that could be incorporated into such an assessment.

Fundamental rights impact assessments are needed given that a contextualised assessment is required. This is because uses of AI vary considerably in terms of complexity, level of automation, potential errors and harm, scale of application, as well as area of use. The more complex an AI system is, the more difficult it is to assess its potential impact.

While the fundamental rights implicated will vary depending on the area of application, the full spectrum of rights needs to be considered for each use of AI. However, uses of AI are likely to involve some of the rights most often affected by AI systems. The discussion in the preceding chapter makes clear that issues linked to data protection, non-discrimination, as well as access to effective remedies and a fair trial, are relevant for all uses of AI.

Thus, the following horizontal points could be a basic starting point for considering the impact of AI on selected rights:

- The **legal processing of data** needs to be confirmed in line with data protection laws.

If personal data are used, the full data protection framework applies. This ensures that processing is legal and does not violate a person’s rights to respect for a private and family life, and data protection.

- The processing should **not lead to unfair treatment or discrimination of protected groups**.

Assessing non-discrimination needs to be at the core of assessing AI. Even apparently miniscule differences can scale up and create risks contravening the principle of non-discrimination. The disadvantage to people depends on the nature (kind of harm), severity (strength of harm) and significance (how many people are put at a disadvantage compared to another group of people). Statistical assessments on group differences are an important tool to assess unfair and discriminatory uses of AI.[40]

- People subjected to AI and related technologies should **be able to complain and receive effective remedies**.

There should be accessible ways for people to complain about potential decisions being made and to effectively access remedies. This includes availability of information that allows the explanation of decisions.

In addition, other relevant rights in the Charter apply. Public administrations using AI need to consider good administration principles. Businesses have to take consumer protection into account.

Other rights are relevant depending on the area of application. Some examples include:

- the right to social protection, when working with social benefits;
- the right to freedom of expression and information, when using AI to support online content moderation;
- the right of assembly and of association, when considering the use of facial recognition technology in public spaces;
- the right to education, when using AI in the education sector;
- the right to asylum, when using AI to support migration management;
- the right of collective bargaining and action, when using AI in the 'gig-economy';
- the right to fair and just working conditions, when using AI at the workplace;
- the right to access preventive health care, when using AI in health services;
- and the right to the presumption of innocence and the right to defence, when using AI in the justice sector or for law-enforcement purposes.

Information needed to assess the potential impact on fundamental rights before implementing AI

Given the variety of tools, purposes and area of application, assessments are contextual. To be able to meaningfully respond to the horizontal points raised above, and to assess specific rights linked to different use cases, at least the following information needs to be available:

- A description of the purpose and context of the system, as well as the legal basis.
- A description of possible harm of using the system, including questions around false positives, false negatives, and other possible harm due to the automation and scale of use.
- A description of the technology used. This includes information on the data used for building the system and its legal basis for processing. A description of relevant information to include is provided in FRA's paper on data quality and AI.[41]
- An evidence-based description of the accuracy of the AI system in terms of outcomes based on training data and possible tests and experiments in real life situations, if appropriate. Here, false positives and false negatives should be considered separately. These should include breakdowns for as many groups as possible to allow for checking potential discrimination (e.g. differences in the accuracy between women and men).
- Where already available, the provision of information about compliance with existing standards and potential certifications obtained.

Ex-post assessments and safeguards

Lastly, envisaging ex-post safeguards further contributes to the fundamental rights compliant use of AI. These could include:

- Regular repetition of assessments after deployment, where appropriate. This is important to learn about potential feedback loops and in case rules are updated. This also requires recording information on the use and outcomes of the system to the extent data protection is respected.
- Making people subjected to AI systems aware that they are subjected to this technology, as they can otherwise not challenge any decision affecting them.
- Making available easily accessible channels for effectively complaining about decisions made based on the AI system.

Engaging external experts, stakeholders and oversight bodies

The above information could be the basis for consultation with different stakeholders and experts before a particular AI system is used. Depending on the nature of the application and its legal basis, a consultation with relevant stakeholders would ensure that no potential harm has been omitted and different perspectives are brought into the assessment. Stakeholders could include civil society; different public and private organisations; as well as experts from different fields of fundamental rights, including data protection.

As the ten experts interviewed for this report highlighted, existing oversight bodies are also responsible for AI oversight within their mandates. Sector-specific bodies and certification schemes are doing this to some extent, the interviews suggest – for example, in health care and financial oversight.

To monitor, comprehend and effectively respond to the potential impacts of AI on a wide spectrum of fundamental rights, data protection authorities, equality bodies, ombuds institutions and national human rights institutions could play an important role, providing input and oversight from their various points of expertise. However, as interviews indicated, extensive upskilling and resource allocation is needed to underpin this.

Endnotes

[1] Council of Europe, Commissioner for Human Rights (2019), *Unboxing Artificial Intelligence: 10 steps to protect Human Rights – Recommendation*, Council of Europe, Strasbourg, May 2019.

[2] Heleen L Janssen (2020), 'An approach for a fundamental rights impact assessment to automated decision-making, International Data Privacy Law', *International Data Privacy Law*,Vol. 10, Issue 1, February 2020, pp. 76-106; Alessandro Mantelero, 'AI and Big Data: A blueprint for a human rights, social and ethical impact assessment', *Computer Law & Security Review* Vol. 34, Issue 4, August 2018, pp. 754-772; Edwards, Lilian and Veale, Michael (2017), **Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For** (May 23, 2017), 16 *Duke Law & Technology Review* 18.

[3] AccessNow (2020), **Access Now's submission to the Consultation on the "White Paper on Artificial Intelligence - a European approach to excellence and trust".**

[4] Council of Europe, **Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems,** 8 April 2020, para 5.2. (human rights impact assessment).

[5] For a detailed discussion with respect to non-discrimination, see: Wachter S., Mittelstatt, B., and Russel C. (2020), **Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI**.

[6] Government of Canada, (2019), **Directive on Automated Decision-Making**.

[7] United Nations, ***UN Guiding principles on Business and Human Rights***, endorsed by Human Rights Council Resolution 17/4, A/HRC/RES/17/4, 6 July 2011, Principles 18, 19, 20.

[8] Heleen L Janssen, ***An approach for a fundamental rights impact assessment to automated decision-making, International Data Privacy Law***, Vol. 10, Issue 1, February 2020, pp. 76-106.

[9] High-Level Expert Group on Artificial Intelligence, ***Ethics Guidelines for Trustworthy AI***, 8 April 2019, Chapter III.

[10] *Ibid*, p. 15.

[11] See for example: IBM, ***Everyday Ethics for Artificial Intelligence***, 2019; Sony, ***Sony Group AI Ethics Guidelines***, 2019; Vodaphone, ***Vodaphone's AI framework***, 2019; Arborus International and Orange, ***International Charter for Inclusive AI***, 21 April 2020, signed by more than 40 private companies, including Camfil, Danone, EDF, L'Oréal, Metro, Sodexo, etc.

[12] Information Technology Industry Council (ITI), ***ITI AI Policy Principles***, 2017.

[13] ECP Platform for the Information Society, ***Artificial Intelligence Impact Assessment***, The Netherlands, 14 November 2019.

[14] Amnesty International, Access Now, Human Rights Watch, Wikimedia Foundation, ***The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems***, 16 May 2018 at **RightsCon Toronto**; University of Montreal, ***Montreal Declaration Responsible AI***, 2018.

[15] Electrical and Electronics Engineers (IEEE), Global Initiative on Ethics of Autonomous and Intelligent Systems, ***Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems***, 2019; Future of Life Institute, ***Asilomar AI Principles***, Conference outcome of the Future of Life Institute's second conference on the future of artificial intelligence, 2017.

[16] See for example: ECP Platform for the Information Society, ***Artificial Intelligence Impact Assessment***, The Netherlands, 14 November 2019; IEEE Initiative.

[17] Association for Computer Machinery (ACM), ***ACM Code of Ethics and Professional Conduct***, 22 June 2018.

[18] Future of Humanity Institute, University of Oxford, ***Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development***, April 2019.

[19] ISO, ***Standards by iso/iec jtc 1/sc 42, Artificial intelligence,Sstandard and/or project under the direct responsibility of iso/iec jtc 1/sc 42 secretaria***t, ISO, ***ISO/IEC TR 24028:2020 standard Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence***, May 2020. It establishes, among others, the "approaches to assess and achieve availability, resiliency, reliability, accuracy, safety, security and privacy of AI systems." Other ISO standards under development as of September 2020: ISO/IEC CD 23894 Information Technology — Artificial Intelligence — Risk Management, ISO/IEC AWI TR 24027 Information technology — Artificial Intelligence (AI) — Bias in AI systems and AI aided decision making, or ISO/IEC AWI TR 24368 Information technology — Artificial intelligence — Overview of ethical and societal concerns, more information available on **ISO's website**; Electrical and Electronics Engineers (IEEE), ***IEEE P7003™ Algorithmic Bias Considerations***; German AI Federal Association (*KI Bundesverband*), *German AI Federal Association: seal of quality (KI Bundesverband Guetesiegel)*, 22 March 2019.

[20] Article 29 Working Party, ***Guidelines on Data Protection Impact Assessment (DPIA)***, wp248rev.01, 13 October 2017, Annex 2 – Criteria for an acceptable DPIA.

[21] High-Level Expert Group on Artificial Intelligence, ***Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment***, 17 July 2020.

[22] Government of Canada, ***Algorithmic Impact Assessment Tool***, 2019; Danish Institute for Human Rights, ***Human rights compliance assessment quick check***, 7 June 2016.

[23] Center of Government Excellence, Johns Hopkins University, ***Ethics & Algorithm toolkit***, 2018; Government of Canada, ***Algorithmic Impact Assessment Tool***, 2019.

[24] Article 29 Working Party, ***Guidelines on Data Protection Impact Assessment (DPIA)***, wp248rev.01, 13 October 2017, Annex 2 (data protection focus); Danish Institute for Human Rights, ***Human Rights Impact Assessment Guidance and Toolbox***, 2016; AI Pulse, **Creating a Tool to Reproducibly Estimate the Ethical Impact of Artificial Intelligence**, 26 September 2019.

[25] Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), ***Principles for Accountable Algorithms and a Social Impact Statement for Algorithms***, 2019; FAT/ML, ***Social Impact Statement for Algorithms***, 2019.

[26] High-Level Expert Group on Artificial Intelligence, ***Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment***, 17 July 2020.

[27] See e.g. human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; accountability. High-Level Expert Group on Artificial Intelligence, ***Ethics Guidelines for Trustworthy AI***, 8 April 2019.

[28] High-Level Expert Group on Artificial Intelligence, ***Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment***, 17 July 2020, p.5.

[29] Government of Canada, ***Algorithmic Impact Assessment tool***, 2019.

[30] Government of Canada, **Directive on Automated Decision-Making**, 2019, Article 6 an Appendix C.

[31] Center of Government Excellence, Johns Hopkins University, ***Ethics & Algorithm toolkit***, 2018.

[32] Danish Institute for Human Rights, ***Human rights compliance assessment quick check***, 7 June 2016.

[33] Danish Institute for Human Rights, ***Human Rights Impact Assessment Guidance and Toolbox***, 2016.

[34] Heleen L Janssen, ***An approach for a fundamental rights impact assessment to automated decision-making, International Data Privacy***

Law, *International Data Privacy Law*,Vol. 10, Issue 1, February 2020.

35 Alessandro Mantelero, ***AI and Big Data: A blueprint for a human rights, social and ethical impact assessment****, Computer Law & Security Review,* Vol. 34, Issue 4, August 2018, pp. 754-772; AI Pulse - Program on Understanding Law, Science, and Evidence (PULSE), UCLA School of Law, ***Creating a Tool to Reproducibly Estimate the Ethical Impact of Artificial Intelligence****,* 26 September 2019. This model includes a series of questions for Assessing the Human Rights Impact of AI-Enabled Projects.

36 Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes, **Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality**, Journal**,** Vol. 27, 2018 – **Issue 2**.

37 AINOW, **Algorithmic Impact Assessments: a Practical Framework for Public Agency Accountability**, April 2018.

38 The Institute for Ethical AI & Machine Learning (Ethical ML Network (BETA)), **The Machine Learning Maturity Model,** 2019.

39 Brave (2020), **Europe's governments are failing the GDPR**.

40 See Wachter S., Mittelstatt, B., and Russel C. (2020), **Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI**.

41 FRA (2019), ***Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights***, Luxembourg, Publications Office, June 2019.

6

MOVING FORWARD: CHALLENGES AND OPPORTUNITIES

This report is published amidst ongoing European legislative and policy developments on artificial intelligence and the global fight against the coronavirus. The Covid-19 pandemic has potentially quickened acceptance of innovative technologies. Yet it has also shown that AI is not the panacea to all problems, and comes with various challenges.

This report clearly shows that using AI systems engages a wide range of fundamental rights. It also shows that many businesses and public administrations are already using or planning to use AI and related technologies. However, these technologies involve different levels of complexity. Most examples use relatively simple algorithms. The level of automation also varies. Most – but not all – decision making is subject to human review.

The applications currently used are also often only in the development stage. EU and national legislators and policymakers should keep this reality in mind – especially when presented with optimistic expectations of AI's potential vis-à-vis the challenges related to using new technologies and the need to regulate them.

"We try to look into the future. We will automate more and more."
(Private company, Estonia)

"The next steps are related to transparency and open data: that is to say, publish not only information in pdf, but also information in reusable formatting so that it could be reused internally and by the private sector."
(Public administration, Spain)

"AI is a great thing but we must learn to use it."
(Private company, Spain)

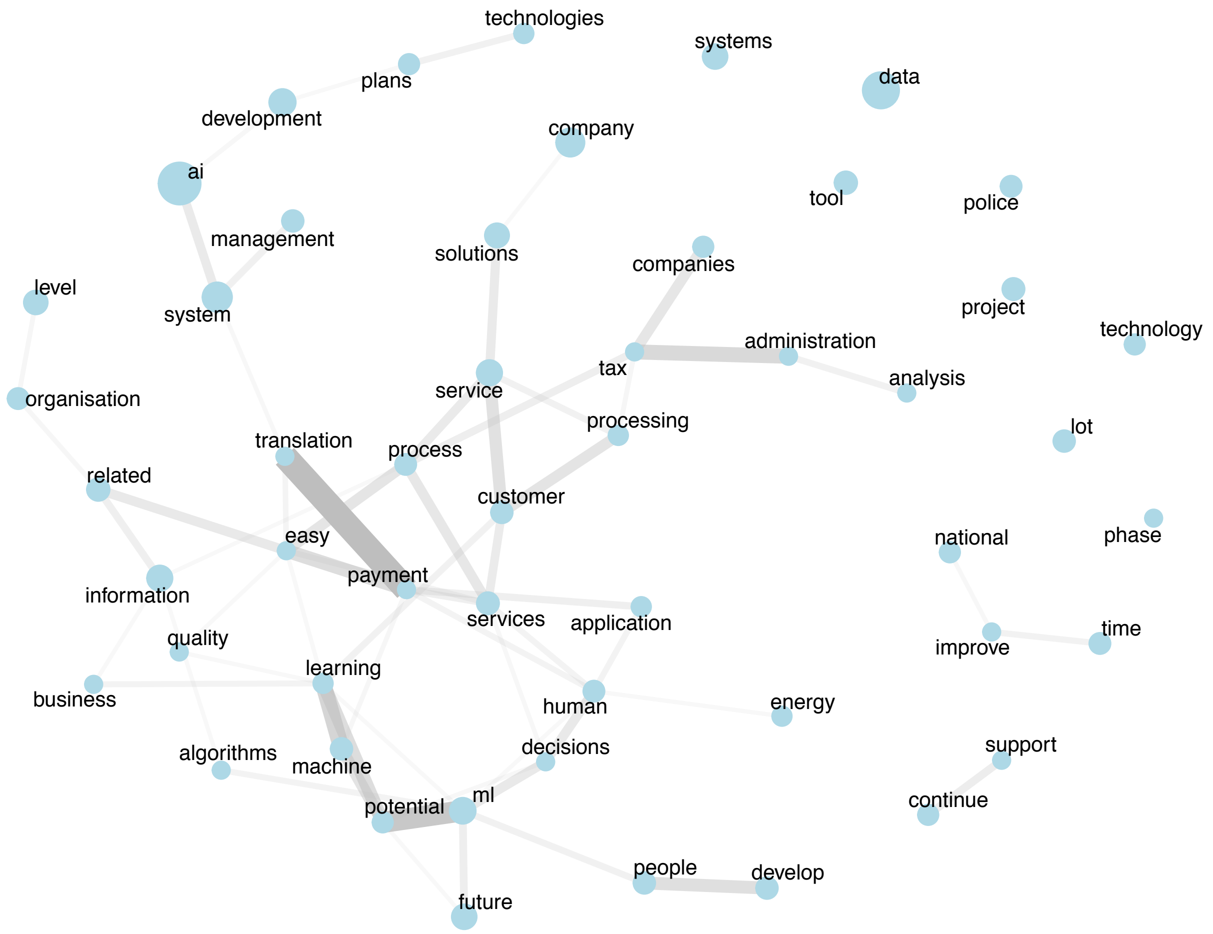
The vast majority of public administrations and businesses interviewed plan to keep working on or using AI. Only two interviewees indicated that they will not further use or develop AI. Another two interviewees are cautious. They plan to wait and see what others are doing, including because of a lack of resources for further work on using AI.

However, most said that they will further develop or continue to test tools and (data) infrastructure with respect to the use of AI. This includes starting new or continuing ongoing pilots, evaluating existing efforts, sharing data and results with others, increasing data quality, or trying to obtain other data sources.

Some interviewees mentioned that they are engaged in ongoing debates and expressed the desire to contribute to the further development of legislation. They still see the current situation – the absence of harmonised law in the area – as an obstacle to the further use of AI. In addition, some respondents said they are working on issues linked to the interpretability of AI. This means that they are working on methods that enhance understanding and explanation of decisions based on more complex AI. Some indicated a desire to look more closely into ethical and legal matters.

Figure 7 shows correlations of words interviewees often use when talking about their future use of AI. The figure indicates topics that are often raised. For example, interviewees often used the term 'data' when discussing future developments.

FIGURE 7: CORRELATIONS OF WORDS RESPONDENTS MOST OFTEN MENTION WHEN DISCUSSING FUTURE PLANS TO USE AI



Notes: *Based on text from interview summaries, when respondents spoke about their future use of AI, including words mentioned at least ten times. The lines connecting words indicate the strength of word correlations within text passages. The size of the dots indicate the frequency of the words used.*

Source: *FRA, 2020*

Effectively and adequately protecting fundamental rights in the EU is a key objective of the current efforts to better regulate the use of AI. In the context of upcoming EU legislation on AI, the European Commission's White Paper addresses current gaps, helping to mitigate the uncertainty around the use of AI with respect to fundamental rights, and making the use of AI more transparent and accountable in terms of fundamental rights. It includes requirements for AI use that directly link to the information needed to assess the impact of AI on fundamental rights, as discussed above.

Requirements linked to the description of training data, data and record keeping, information to be provided to those subjected to AI, robustness and accuracy, as well as human oversight are all highly relevant when assessing and protecting fundamental rights. In this respect, the body of evidence presented in this report offers general insights into how different technologies can affect fundamental rights and what safeguards are needed to ensure fully fundamental rights-compliant use of AI in practice.

At the same time, further research into the fundamental rights implications of the use of AI in specific areas will further support policy and legislative efforts at the EU level aiming to shape Europe's digital future more widely.

FRA will continue to look into the fundamental implications of AI by carrying out more focussed analysis of specific use cases. To increase knowledge on what can potentially go wrong and consequently help mitigate and prevent fundamental rights violations, FRA will look into potential simulation studies. These can showcase how biased algorithms can negatively affect fundamental rights.

The use of AI often involves automating tasks that were previously carried out by humans. Here we need to acknowledge that human behaviour is sometimes not in line with fundamental rights, both when using AI and when not using AI. For example, the police might engage in unlawful profiling. Decisions by public administration or companies might sometimes be driven by negative stereotypes.

Current developments in the use of AI need to acknowledge the potential for discrimination with respect to the data on which an AI system is built, and with respect to the underlying assumptions that humans in turn may feed into the development and deployment of a system. Automating certain tasks without fully understanding what is being automated could lead to unlawful processing of data, the use of technology that treats people unfairly, and might make it impossible to challenge certain outcomes – to name some challenges.

However, the increased availability of data and technological tools can also be used to better understand where and how unequal treatment occurs. Current technological developments and the increased availability of data also provide a unique opportunity to better understand the structures of society, which can be used to support fundamental rights compliance. The opportunities created by AI can also contribute to better understanding and consequently mitigation of fundamental rights violations.

Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: **https://europa.eu/european-union/contact_en**

On the phone or by email
Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: **https://europa.eu/european-union/contact_en**

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: **https:// europa.eu/european-union/index_en**

EU publications
You can download or order free and priced EU publications at: **https://op.europa.eu/en/publications**. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see **https://europa.eu/european-union/contact_en**).

EU law and related documents
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR- Lex at: **http://eur-lex.europa.eu**

Open data from the EU
The EU Open Data Portal (**http://data.europa.eu/euodp/en**) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

FRA EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

PROMOTING AND PROTECTING YOUR FUNDAMENTAL RIGHTS ACROSS THE EU

Artificial intelligence (AI) already plays a role in deciding what unemployment benefits someone gets, where a burglary is likely to take place, whether someone is at risk of cancer, or who sees that catchy advertisement for low mortgage rates. Its use keeps growing, presenting seemingly endless possibilities. But we need to make sure to fully uphold fundamental rights standards when using AI.

This report presents concrete examples of how companies and public administrations in the EU are using, or trying to use, AI. It focuses on four core areas – social benefits, predictive policing, health services and targeted advertising.

The report discusses the potential implications for fundamental rights and analyses how such rights are taken into account when using or developing AI applications. In so doing, it aims to help ensure that the future EU regulatory framework for AI is firmly grounded in respect for human and fundamental rights.

EU Charter of Fundamental Rights

Access to justice

Non-discrimination

Information society

FRA – EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS
Schwarzenbergplatz 11 – 1040 Vienna – Austria
TEL. +43 158030-0 – FAX +43 158030-699

fra.europa.eu

facebook.com/fundamentalrights
twitter.com/EURightsAgency
linkedin.com/company/eu-fundamental-rights-agency

Publications Office
of the European Union