Delmi

Report
2026:3

Artificial Intelligence and Asylum Decision-Making:

Any Role for Human Rights Law?

Artificial Intelligence and Asylum Decision-Making: Any Role for Human Rights Law?

Vladislava Stoyanova

Report 2026:3

Delmi

SWEDISH GOVERNMENT INQUIRIES

The Migration Studies Delegation Ju 2013:17

Delmi Report 2026:3
Order: www.delmi.se
E-mail: ju.delmi@regeringskansliet.se
Cover: By Paul Campbell on Unsplash
Stockholm 2026
ISBN: 978-91-89993-56-3

Preface

Artificial intelligence (AI) is reshaping migration, asylum and border policies by enabling automated decision-making. While AI promises efficiency, its use in asylum procedures raises serious human rights concerns, particularly because affected individuals are often in vulnerable positions.

The EU AI Act classifies use of AI in asylum procedures as "high risk". This report examines whether AI-supported decision-making in asylum cases aligns with human rights law, especially the rights to privacy and protection from refoulement.

Key concerns include the possible *procedural harm*, since AI may compromise fairness, transparency and the involvement of applicants; *privacy*, since AI use must meet legal standards of legality and necessity amongst others. And, last but not least, *proportionality*. Even if AI complies with the EU AI Act, it may still violate human rights if the risks outweigh the benefits, especially when harms are serious and hard to justify.

A fundamental challenge is that asylum decisions lack feedback mechanisms, making it difficult to test and validate AI systems. As a result, AI could shift discretion from individual decision-makers to system designers – changing not just how decisions are made, but how asylum law functions in practice.

The author of this research report is Vladislava Stoyanova, Associate Professor of Public International Law at the Faculty of Law, Lund University. The report has been reviewed by external researchers at a scientific seminar. Niovi Vavuola, Associate Professor and Chair in Cyber Policy at Université du Luxembourg and Lena Enqvist Associate Professor in Law, at Umeå University.

The work on this report has been followed by Anna Lindblad, Acting Chief of Staff and member of Delmi's Board of Directors. At the secretariat of Delmi, Deputy Director Henrik Malm Lindberg and Delegation Secretaries Suzanne Planchard and Constanza Vera-Larrucea have contributed to the review process of the report.

As usual with Delmi-publications, the authors are fully responsible for the report's contents, including its conclusions and policy recommendations.

Stockholm, November 2025

Joakim Palme
Chair, Delmi

Agneta Carlberger Kundoori
Director, Delmi

Summary

Artificial intelligence-supported decision-making will play a defining role in the future of migration, asylum, and border policy. Assumptions that AI enables more efficient decisions will shape the trajectory of asylum law. These developments are part of a new era where automated decision-making, digitalisation and AI have increasingly important role in our lives, including for public administration. Migration, asylum and border management are, however, marked by certain specific features - most notably, the fact that affected individuals are often in a position of vulnerability, particularly in relation to the host State from which they are seeking protection. For this reason, the EU Artificial Intelligence Act (EU AI Act) classifies the use of AI for immigration, asylum and border control as 'high-risk.' Another specificity is the wide scope of discretion that normally States enjoy in this area.

In light of these developments, the report engages with the following central question: To what extent are technologies of AI supported decision-making used in the context of asylum procedures, compatible with human rights law?

This compliance question could only be answered by first identifying the benefits that AI systems might offer and the harms to fundamental interests that AI systems might pose. As to the benefits, EU AI Act and with the Council of Europe AI Framework Convention that are tools for interpretation of human rights law, rest on the assumption that AI systems generate benefits and that, while they entail risks, such risks are to be managed. The harm can be initially circumscribed to harm related to the procedure for assessing protection claims, i.e. harm to the fundamental interests protected by the right not to be subjected to *refoulement.*

The study clarified the difficulties in establishing causation between this harm and the use of AI systems. These difficulties can be however overcome if the harm is conceptualised as a *procedural harm*. This justifies focus on procedural guarantees (i.e. quality of the decision-making process, timeliness, effectiveness, independence, involvement of affected individuals, clarity of the reasoning behind decisions). If an AI system is involved, these guarantees should be complied with, including the involvement of the affected individuals and clarity of the decisions that have affected them.

Another concern is privacy. In relationship to the right to private life, the harm can be conceptualized in the following way – the use of the systems constitutes an interference with private life and therefore a limitation of private life. If this

is the case, the interference (i.e. the usage of the AI system in the decision-making process) has to comply with the tests of legality, suitability and proportionality to be in compliance with States' negative human rights obligations.

The application of these tests means that AI systems might fully comply with, for example, the concrete requirements of the EU AI Act, and still be contrary to human rights law. The latter demands contextual proportionality assessment. This prompts an analysis whether any risks (including residual risks) necessarily posed by high-risk AI systems are proportionate to any benefits. This assessment implies consideration of the cost of precaution: the cost for addressing the risks. In the context of the asylum determination procedures, the cost of precaution might be considered more difficult to justify. This in turn might imply shifting risks to the asylum-seekers. Such a shift might become more and more difficult to substantiate with the increase of the seriousness of the potential harm. The bigger role an AI system has in the decision-making process, the easier it might be to expose the causal link between the system and the harm, which necessarily affects the proportionality analyses. When there is a higher risk that *both* the right to private life *and* the right to *non-refoulement* will be compromised, the proportionality test is more likely to weigh in favour of the individual.

Importantly, the proportionality analysis is also influenced by the scale of benefits offered by AI systems. The broader and more concretely defined these benefits are, the more likely it is that potential risks will be considered proportionate.

The study stressed some inherent characteristics as to the nature of the refugee status determination procedure. The study highlights how *AI supported decision-making presents distinctive problems for applying the legal standards in the procedure about assessment of protection needs.* A major issue is that asylum decision-makers typically lack the means to verify whether their decisions (e.g. granting or rejecting protection) were correct. This absence of feedback means there is no reliable test data to evaluate AI systems during development or after being placed in operation. Moreover, the historical data for the development of the system might not be relevant for predicting future risks in applicants' countries of origin.

This brings us to the following final insight that the report offers: the news *technologies themselves can change the practice of asylum law.* Such a change seems possible, given the increased importance of data, the selection of data and the role of programmers in the design of the algorithms. This in turn implies a *shift away from discretion of individual decision-makers and in favour of discretion in the design of the systems themselves.*

Sammanfattning

Beslut som fattas med hjälp av artificiell intelligens kommer att få en allt större betydelse för framtidens migrations-, asyl- och gränspolitik. Antaganden om att AI kan möjliggöra mer effektiva och träffsäkra beslut kommer i hög grad att påverka hur asylrättens formas. Dessa förändringar äger rum i en tid då automatiserat beslutsfattande, digitalisering och AI spelar en allt viktigare roll i våra liv, även inom den offentliga förvaltningen. Samtidigt präglas migration, asyl- och gränskontroll dock av vissa specifika drag – framför allt att de personer som berörs ofta befinner sig i en utsatt situation, särskilt i förhållande till gentemot den stat från vilken de söker skydd. Av denna anledning klassificerar EU:s AI-förordning om artificiell intelligens (EU AI Act) AI-system som används inom invandring, asyl och gränskontroll som "högrisk". Ett annat kännetecken är det stora utrymme för skönsmässig bedömning som staterna normalt har på detta område.

Mot denna bakgrund behandlar rapporten följande centrala fråga: I vilken utsträckning är AI-stödda beslutsprocesser som används i samband med asylförfaranden förenliga med mänskliga rättigheter?

För att besvara denna fråga krävs först en förståelse för både de potentiella fördelar som AI-system kan erbjuda och de risker och skador som systemen kan medföra. När det gäller fördelarna bygger EU:s AI-förordning och Europarådets ramkonvention om AI, som är verktyg för tolkning av människorättslagstiftningen, på antagandet att AI-system genererar betydande nytta, men att de risker som följer måste hanteras. I denna kontext handlar den primära potentiella skadan om brister i själva asylprövningsprocessen, det vill säga risker för de grundläggande intressen som skyddas av principen om *non-refoulement* – den princip som förbjuder att skicka asylsökande tillbaka till ett land där deras liv, frihet eller säkerhet är hotad.

Studien belyser de svårigheter som finns med att etablera ett orsakssamband mellan en sådan skada och användningen av AI-system. Dessa svårigheter kan dock hanteras genom att förstå skadan som en *förfarandemässig skada*. Detta innebär att fokus läggs på förfarandemässiga garantier dvs. kvaliteten på beslutsprocessen, snabbhet, effektivitet, oberoende, involvering av berörda individer samt att tydlighet i motiveringen till besluten får en viktig betydelse. När AI-system används måste dessa garantier fullt ut upprätthållas samt att man involverar de berörda personerna och är tydlig i med beslut som har påverkat dem.

Användningen av AI i beslutsprocessen utgör ett ingrepp i rätten till privatliv och måste därför uppfylla kraven på *laglighet, lämplighet* och *proportionalitet*. Det innebär att ett AI-system kan uppfylla samtliga krav i EU:s AI-förordning och ändå vara oförenligt med mänskliga rättigheter. Den människorättsliga proportionalitetsbedömningen är kontextuell och kräver en analys av om riskerna – inklusive kvarstående risker från systemen – står i proportion till de fördelar som systemen erbjuder. Denna analys omfattar även kostnaden för försiktighetsåtgärder, det vill säga vad det kostar att hantera riskerna från systemen (d.v.s. de negativa konsekvenserna för asylsökande). Inom asylprocessen kan en hög kostnad för att iaktta försiktighet vara svår att rättfärdiga, vilket i praktiken kan leda till att risker förs över på asylsökande. Ju allvarligare risken för skada är och ju större roll AI-systemet har i beslutsprocessen, desto svårare blir det att motivera en sådan riskövervältring. Detta påverkar proportionalitetsbedömningen – särskilt då både rätten till privatliv och rätten att inte utsättas för refoulement står på spel.

Tillämpningen av dessa kriterier innebär att AI-system kan uppfylla, till exempel, de konkreta kraven i EU:s AI-förordning fullt ut, men ändå strida mot rätten om mänskliga rättigheter. Det senare kräver en sådan proportionalitetsbedömning som också sätter in frågan i sin kontext. Detta föranleder en analys av de risker (inklusive kvarstående sådana), som AI-system med hög risk nödvändigtvis medför, står i proportion till eventuella fördelar. Denna bedömning innebär att man måste beakta kostnaden för försiktighetsåtgärder alltså kostnaden för att hantera riskerna. I samband med procedurer för asylprövning kan kostnaden för försiktighetsåtgärder anses vara svårare att motivera. Detta kan i sin tur innebära att riskerna vältras över på de asylsökande. En sådan övervältring bli svårare att motivera om den potentiella skadan är än allvarligare. Ju större roll ett AI-system har i beslutsprocessen, desto lättare kan det vara att påvisa orsakssambandet mellan systemet och skadan, vilket nödvändigtvis påverkar proportionalitetsanalyserna. När det finns en högre risk att *både* rätten till privatliv *och* rätten till *non-refoulement* äventyras, är det mer sannolikt att proportionalitetstestet faller ut till individens fördel.

Proportionalitetsbedömningen påverkas även av hur omfattande och konkret definierade de förmodade fördelarna med AI-systemen är. Ju större och tydligare nyttor som kan identifieras, desto lättare blir det att motivera att riskerna ska anses vara proportionerliga.

Studien betonade vissa strukturella särdrag när det gäller hur man fastställer flyktingstatus. En iakttagelse är att *AI-stödda beslut medför särskilda problem för tillämpningen av de rättsliga normerna när man ska bedöma skyddsbehoven.* Ett grundläggande problem är att beslutsfattare i asylärenden ofta saknar

möjlighet att fastställa om deras beslut – att bevilja eller avslå skydd – var korrekta. Denna brist på återkoppling innebär att det inte finns några tillförlitliga möjligheter för att utvärdera AI-system både under utveckling och efter att de tagits i drift. Dessutom är de historiska data som använts för att utveckla systemet kanske inte relevanta för att förutsäga framtida risker i sökandenas ursprungsländer.

Ett sista viktigt medskick från rapporten är att AI-teknik *i sig kan förändra tillämpningen av asylrätten.* Detta är möjligt eftersom dataval, datatillgång och programmerares utformning av algoritmer får allt större betydelse. Detta innebär i sin tur en *förskjutning bort från enskilda beslutsfattares godtycke till ett system där godtycket i stället flyttar till hur de tekniska systemen utformas.*

Table of Contents

Tables

1. Introduction

Artificial intelligence-supported decision-making will play a defining role in the future of migration, asylum, and border policy.

Assumptions that it enables more efficient decisions will shape the trajectory of asylum law. As artificial intelligence (AI) reshapes public administration, one of the most politically charged frontiers is asylum decision-making – a domain where technology meets the lives of those seeking protection, and where human rights law faces a critical test. This tension is particularly evident in the EU, which actively promotes the use of new technologies in the area of migration, asylum and border management.[1] Technologies involving the use of AI might have been already tested.

In the asylum process, AI technologies are used in many different ways. Some examples include:

- Forecasting tools that try to predict future migration and displacement towards Europe.[2]
- Automated systems for processing asylum and residency applications.[3]
- Speech and dialect recognition to help determine where someone might be from.
- Risk assessment tools, such as those estimating the likelihood that someone might leave or "abscond."[4]

[1] Artificial intelligence in asylum procedures in the EU (European Parliament Briefing, July 2025); See European Commission, Opportunities and Challenges for the Use of Artificial Intelligence in Border Control, Migration and Security Vol. 1, Main Report 2020, written by Deloitte. Brussels: European Commission. https://op.europa.eu/en/publication-detail/-/publication/c8823cd1-a152-11ea-9d2d-01aa75ed71a1/language-en

[2] See the 2019 UNHCR Project Jetson, https://jetson.unhcr.org/ CB Casagran, et al., 'The Role of Emerging Predictive IT Tools in Effective Migration Governance (2021) Politics and Governance 133; Helga Valk, et al. 'How to Predict Future Migration: Different Methods Explained and Compared' in Peter Scholten (ed) *Introduction to Migration Studies* (Springer 2022) 463. See also the Early Warning and Preparedness System developed by the European Asylum Support Office.

[3] See, e.g. the German Federal Office for Migration and Refugees that considers itself to be is a pioneer when it comes to digitalising asylum. BAMF - Bundesamt für Migration und Flüchtlinge - Digitalising the asylum procedure

[4] See the pilot EU 'Horizon 2020' project I BorderCtrl. Germany has introduced this technology to classify languages and dialects so that the asylum-seekers' countries of origin can be identified. The report issued after the dialect analysis is taken into account in the overall assessment made by the caseworker.

- Automated distribution of welfare benefits for asylum seekers.
- Systems for assigning people to reception centres based on availability or other criteria.[5]
- Automated match tools for screening similar asylum applications.
- AI tools for summarising cases – used to automatically summarize asylum interview transcripts, AI tools for transcribing interviews.[6]
- Search tools that help find relevant documents quickly – these can assist in assessing whether it's safe to return someone to their country of origin.[7]
- Language analysis tools to establish someone´s county of origin based on how they speak.[8]
- AI powered chatbots could provide real-time multilingual information to asylum seekers to help them in the procedure.[9]

These examples suggest that AI technologies might be useful. They also illustrate the breadth of AI integration in asylum processes, which in turn raises important questions for law and policy addressed in this report.

These developments are part of a new era where automated decision-making, digitalisation and AI have increasingly important role in our lives, including for public administration. Migration, asylum and border management are, however, marked by certain specific features – most notably, the fact that affected individuals are often in a position of vulnerability, particularly in

[5] See GeoMatch - Immigration Policy Lab (immigrationlab.org) K Bansak, 'Improving refugee integration through data-driven algorithmic assignment' (2018) *Science* 359.

[6] See The Ministry of the Interior - CEDAT85 , where it is noted that 'The automatic transcription takes into account all semantic aspects, not excluding the recognition of dialects, accents, foreign terminology, and spontaneous speech, with an accuracy level of no less than 95% and a very high security rate.'

[7] See here AI in the UK Asylum system: Innovation or injustice? - Seraphus for the AI Case Summarisation and AI Policy Search tested in the UK. See Opportunities and challenges for the use of artificial intelligence in border control, migration and security . Volume 1, Main report - Publications Office of the EU'

[8] See EUAA Strategy on Digital Innovation in Asylum Procedures and Reception Systems at 28, where it is stated that the EUAA 'has been actively engaging in discussions with Member States on combining first-line artificial intelligence and second-line human analysis for better quality results in the field of LADO [language analysis for determination of origin].'

[9] Report Artificial Intelligence and Migration (CoE Parliamentary Assembly, Committee on Migration, Refugees and Displaced Persons, 2025) 1680b67b8a

relation to the host State from which they are seeking protection.[10] For this reason, the EU Artificial Intelligence Act (EU AI Act) classifies the use of AI for immigration, asylum and border control as 'high-risk.' Another specificity is the wide scope of discretion that normally States enjoy in this area.

In simple terms, these technologies aim at the automation of processes that are normally done by human beings by using algorithms and coding of data, and some form of autonomy.[11] The idea is that state officials, lawyers, case workers etc., can be relieved from mundane and repetitive tasks. The algorithms can be fed with vast amount of data, link different databases (interoperability)[12] to identify patterns and correlations that can support problem-solving.[13] In this way the process of decision-making includes a 'machine' to help (and possibly replace) human decision-making that includes review of a large amount of information and routine tasks.[14] The decision-making process can therefore become more efficient, which suggests a benefit from the use of the technologies.

Why are these technologies used in the area of asylum? An argument that is normally invoked concerns efficiency and saving of administrative resources.[15]

[10] Evelien Brouwer, 'Large-Scale Databases and Interoperability in Migration and Border Policies: The Non-Discriminatory Approach of Data Protection' 26 (2020) *European Public Law* 71; Niovi Vavoula, 'The 'Puzzle' of EU Large-Scale Information Systems for Third-Country Nationals: Surveillance of Movement and Its Challenges for Privacy and Personal Data Protection' 45 (2020) *European Law Review* 348; Charly Derave, Nathan Genicot and Nina Hetmanska, 'The Risks of Trustworthy Artificial Intelligence: The Case of the EuropeanTravel Information and Authorisation System' 13 (2022) *European Journal of Risk Regulation* 389; Valsamis Mitsilegas, 'Interoperability as a Rule of Law Challenge', EUPLANT (May 6, 2020).

[11] See the discussion of the definitions of AI in Sections 4.1. and 5.1. that refer to autonomy.

[12] Evelien Brouwer, 'Large-Scale Databases and Interoperability in Migration and Border Policies: The Non-Discriminatory Approach of Data Protection' 26 (2020) European Public Law 71; Niovi Vavoula, 'The 'Puzzle' of EU Large-Scale Information Systems for Third-Country Nationals: Surveillance of Movement and Its Challenges for Privacy and Personal Data Protection' 45 (2020) European Law Review 348; Charly Derave, Nathan Genicot and Nina Hetmanska, 'The Risks of Trustworthy Artificial Intelligence: The Case of the EuropeanTravel Information and Authorisation System' 13 (2022) European Journal of Risk Regulation 389; Valsamis Mitsilegas, 'Interoperability as a Rule of Law Challenge', EUPLANT (May 6, 2020).

[13] See Explaining decisions made with AI (UK Information Commissioner's Office and the Alan Turing Institute 2020) Explaining decisions made with AI | ICO

[14] The term 'machine' might not be entirely appropriate since these are algorithms based on coding of data performed by humans and it is humans that decide what data to use and how to code the data. However, once the algorithms are created, they might subsequently 'learn' and in this sense their operation might become more and more independent from humans.

[15] Ludivine Sarah Stewart, 'Fair and Efficient Asylum Procedures and Artificial Intelligence: Qua Vadis Due Process?' (2024) 55 *Computer Law and Security Review* 1.

In this sense, these technologies are intended to help States make, for example, faster and better decisions as to who might be eligible for international protection, who might be vulnerable,[16] and thus in need of certain reception conditions. Once protection granted, the technologies might also facilitate integration in the host society.[17] For example, in Norway, it has been reported that in light of the automated allocation of asylum-seekers to reception centres, 'caseworkers do not need to go through each reception centre individually and can find the most suitable centre for the applicant automatically.'[18]

The future public debate and policy making about migration, asylum and border management will be inevitably shaped by the possibilities offered by AI supported decision-making. The proposition that these technologies can make better and faster decisions or facilitate better and faster decisions will be important for the application and the future development of asylum law. For these reasons, the report is highly relevant not only for policy-makers and legislators, but also for the general public. Its significance is wide as AI supported decision-making is poised to become a central issue in national and international legal and policy debates concerning administrative and judicial decision-making. This is especially important in light of the increasingly blurred boundaries between human judgment and technological intervention.

The focus of this report is on AI as defined in the EU AI Act and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CoE AI Convention).[19] While the legal implications of AI are now being explored across a wide range of fields, its use in asylum procedures warrants particular attention due to the high stakes involved and the vulnerability of the individuals affected. Asylum is not only complex, but also a very politicized issue. This has at least two implications. First, there is a risk that these technologies might be developed and applied with less safeguards.[20] Human rights law accommodates States' migration

[16] Automated vulnerability assessment might lead to affording asylum seekers access to special reception conditions and procedural safeguards. See Francesca Palmiotto 'Procedural Fairness in Automated Asylum Procedures: Fundamental Rights for Fundamental Challenges'55 (2024) *Computer Law and Security Review 1*.

[17] K Bansak, J Ferwerda, J Hainmueller, A Dillon, D Hangartner, D Lawrence, & J Weinstein, 'Improving Refugee Integration through Data-driven Algorithmic Assignment' (2018) *Science* 359(6373), 325.

[18] 'Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe' (Refugee Studies Centre 2023) 63.

[19] See Sections 4.1 and 5.1. Not all automated decision-making includes AI. See, for example, *Ligue des Droits Humains* C-817/19, 27 January 2022, para 194.

[20] This risk relates to the fact that in general human rights law imposes less robust standards in the area of migration and asylum in comparison with its more general standards as applied to citizens.

control interests, which might explain the lesser stringency of the standards. The widely accepted discretion afforded to States in the area of migration control also extends to how decisions are made. This includes, potentially, the discretion to adopt AI systems. Once that choice is made, a series of subsequent decisions inevitably follow - such as which data sets to use for training, how to design the AI system, and how to structure the coding process. These downstream choices can be understood as falling within the scope of the initial discretion granted in the field of migration control.

Second, because these new technologies are often perceived as impartial and accurate,[21] they may make it easier to persuade the public that, for instance, refugees identified by an algorithm as having strong integration prospects should be accepted. In this way, the perceived objectivity can be leveraged for political purposes. Indeed, according to international refugee law and human rights law, international protection (including refugee status) is granted when a person might be exposed to risk in his/her country of origin, and in this sense, prospects for better integration are *not* relevant. However, such prospects can be taken into consideration when refugees access destination countries via resettlement.[22] Prospects for integration (understood as, for example, labour and educational opportunities) are also highly relevant if individuals in need of protection, access destination countries via the so-called complementary legal pathways. Such pathways have been recently very vocally promoted as an idea by the EU and the UN.[23]

In addition to the above mentioned two considerations, it should also be taken into account that the application of new technologies for the purposes of border control and for processing asylum and migration cases, can be regarded as a test case. States have already created various databases with information in the area of migration.[24] Databases have been already inter-linked. There is already data that can be fed into for the creation of algorithms.[25] The availability of this data enables the development of these technologies.

[21] Frida Alizadeh Westerling, 'Technology-related Risks to the Right to Asylum: Epistemic Vulnerability Production in Automated Credibility Assessment' (2022)13(2) *European Journal of Law and Technology.*

[22] T de Boer and M Zieck, 'The Legal Abyss of Discretion in the Resettlement of Refugees' (2020) International *Journal of Refugee Law* 61.

[23] For a detailed examination, see V Stoyanova, 'Addressing the Legal Quagmire of Complementary Legal Pathways' (2023) *European Journal of Migration and Law* (25) 164.

[24] Niovi Vavoula, *Immigration and Privacy in the Law of the EU. The Case of Information Systems* (Brill, 2022).

[25] Niovi Vavoula, 'The Tr-AI-nsformation of EU Migration, Asylum and Border Management: The Roles of the AI Act, Interoperable Large-scale IT System and Migration Agencies' *Computer Law & Security Review* (forthcoming).

Once the technologies are developed in the context of asylum and migration and their possibilities explored,[26] these technologies might be applied in other areas. This in turn is also a test case of human rights law. If human rights-compliant AI systems in the area of asylum and migration cannot be achieved, perhaps this should be a signal that it cannot be achieved in other areas either.

[26] Here it should be also noted that the development of such technologies is possible in the context of asylum and migration since vast data is already collected and it possible to collect since people who cross international borders are an object of various stages of data collection. See for example, the development of ETIAS. In addition, large IT systems in the area of migration are not entirely outside of the scope of the AI Act but, according to its Art 111(1), benefit from an extended compliance deadline in 2030. See Niovi Vavoula, 'The Tr-AI-nsformation of EU Migration, Asylum and Border Management: The Roles of the AI Act, Interoperable Large-scale IT System and Migration Agencies' *Computer Law & Security Review* (forthcoming).

2. Questions addressed

In light of these considerations, it is crucial to engage with the following questions:

> To what extent are technologies of AI supported decision-making used in the context of asylum procedures, compatible with human rights law? Which specific human rights may be affected?
>
> Can these technologies be developed in compliance with human rights law? If yes, how? How could these technologies help in the development of improved decision-making in the area of asylum, which might increase the overall quality of the procedure thus making it more likely to comply with human rights law?
>
> If technologies can be developed in compliance with human rights law, how could the technologies themselves change the practice of asylum law? What other problems could ensue, if the technologies are used, and are these problems possible to review against human rights law standards?

It is relevant to note here that this study is limited to the use of AI systems that are relevant to the process of seeking asylum. Admittedly the distinction might be subtle, but technologies that more generally relate to migration and movement across international borders, are excluded. This will ensure the clear focus of the study.

It is also relevant to underscore that the objective of the study is to assess *both* the advantages and the disadvantages from the new technologies from a human rights law perspective. It has been documented that human decision making and the assessment of evidence in the area of asylum, faces serious challenges, which has given grounds for questioning its objectivity.[27] AI systems could be helpful in addressing flaws that inhere in human decision-making. In addition, long waiting times that might characterise human decision making,

[27] See G Noll, *Proof, Evidentiary Assessment and Credibility in Asylum Procedures* (Martinus Nijhoff 2005).

are also highly problematic from a human rights law perspective,[28] and are neither in the interest of the affected individuals, nor in the interest of the host State.

AI systems can be indeed helpful in addressing these challenges.[29] It then follows that in some respects and from the perspective of certain rights, the new technologies might bring advantages. From the perspective of other rights, they might be problematic.[30] This implies a nuanced approach in the effort to identify and distinguish advantages and disadvantages. In addition, whether the effects of AI systems are viewed as advantages or disadvantages may depend on the perspective adopted - whether that of the affected individuals or that of the host State.

There are already reports that aim to map out the use of AI in the area of asylum.[31] These mapping out exercises are very useful starting point for this report. However, there has not been a legal analysis, i.e. an examination of the role of human rights law in terms of imposing any restraints on how these technologies should be developed and once developed how they should be used. It is precisely this gap that the report sets out to examine and bridge. There is already some scholarship that has tried to invoke human rights law for assessing the use of these new technologies.[32] However, a more robust

[28] Especially if these long waiting times imply, for example, restrictions upon the right to liberty and the right to freedom of movement.

[29] See EUAA Strategy on Digital Innovation in Asylum Procedures and Reception Systems at 33: 'Case officers should benefit from the assistance of innovative tools which could eliminate burdensome tasks, thus sparing capacity for high added-value targets. A digital casework assistant could be helpful in several aspects, such as: extracting case relevant COI and case-law from large amount of data, supporting risk analysis, filtering open source intelligence or applicants' mobile devices.'

[30] The current literature seems to focus on the problematic and negative aspects. See, for example, R Reyes, 'Artificial Intelligence Technologies and the Right to Seek and Enjoy Asylum. An Overview' in A Quintavalla and J Temperman (eds) *Artificial Intelligence and Human Rights* (Oxford University Press 2023) 311.

[31] 'Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe' (Refugee Studies Centre 2023 Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe — Refugee Studies Centre (ox.ac.uk); 'Refugee Protection in the Artificial Intelligence Era' Chatham House 2022 Refugee protection in the artificial intelligence era | Chatham House – International Affairs Think Tank

[32] P Molnar, 'Robots and refugees: the human rights impacts of artificial intelligence and automated decision-making in migration', in McAuliffe and Wilson (eds) *Research Handbook on International Migration and Digital Technology* (Edward Elgar 2021) 134; N Vavoula 'The "Puzzle" of EU Large-Scale Information Systems for Third-Country Nationals: Surveillance of Movement and Its Challenges for Privacy and Personal Data Protection (2020) *European Law Review* 372; N Vavoula, 'Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism (2021) *European Journal of Migration and Law* 457.

analysis that truly understands and appreciates also the limits of human rights law, is lacking. For example, human rights law does allow States certain scope of discretion what decisions to take and how to take these decisions. This can be also framed as procedural autonomy. Therefore, any simplistic invocation of the right to privacy and the right to *non-refoulement*,[33] might not very helpful and insightful.

Additionally, the issue of how these technologies might transform the practice of asylum law has rarely been explored. This line of inquiry goes beyond examining how human rights law might constrain or shapes the development of new technologies. It also requires asking what happens to asylum law once it becomes embedded within AI systems. More fundamentally, it invites reflection on how AI might alter the very structure and function of law itself. This report might be useful for addressing these emerging challenges.

[33] The right to *non-refoulement* implies the right not to be returned to a place where there is a risk of persecution or risk of other serious forms of ill-treatment.

3. Method, limitations and structure

This study has a number of limitations considering the novelty of AI and the lack of transparency that sometimes might characterized public management. In particular, it might not be easy to establish whether AI has been used in the decision-making process. National authorities might not have made this information public.[34] Even if some level of transparency exists, it is difficult to fully grasp how the algorithms actually function. There are at least five aspects of this difficulty.

First, the input data based on which algorithms are trained, might be inaccessible due to, for example, intellectual property rights. Here it is relevant to also note that these technologies might be developed by private companies, which creates further legal complications. Second, the input data might suffer from its own biases and deficiencies given that the creator of the technologies (i.e., the provider) has a high level of discretion what data to use and how to code it. Third, once the algorithm created, it is difficult to fully comprehend how it actually reaches a specific decision.[35] Fourth, algorithms might be created in such a way that they are meant to train themselves. This means that once an AI supported decision-making process established, it is difficult to predict how it will evolve with time. Fifth, according to the EU, automated algorithms are planned to connect different databases, which implies processing and connecting massive amounts of personal information about migrants and asylum seekers.[36] The implications from this interoperability are difficult to predict.

All of the above implies that this study's aim to apply human rights law to new factual scenarios that have arisen or might arise due to the use of AI systems, will be quite difficult. More specifically, if I have limited information about the factual empirical reality (i.e., how the technologies are developed, what are

[34] See European Commission. 2022. EMN-OECD Inform. The Use of Digitalisation and Artificial Intelligence in Migration Management. https://www.oecd.org/migration/mig/EMN-OECD-INFORM-FEB-2022-The-use-of-Digitalisation-and-AI-in-Migration-Management.pdf

[35] This is generally referred to as black box or opacity.

[36] See Regulation (EU) 2019/817 of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of borders and visa; Regulation (EU) 2019/818 of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration.

their capabilities, which databases are interlinked and how, how the coding is done, what codes are used), it is difficult to apply the law. If I have limited information about the empirical reality, it is also difficult to address the question how the law should shape the way this reality should be unfolded in the future (i.e. how human rights law should shape how States choose to develop and apply AI systems in partnership with the private sector).

For these challenges to be comprehensively addressed, empirical studies that lead to the collection of new data are necessary. Empirical studies mean studies that expose, for example, whether national asylum authorities use AI and how, what data has been used for training the algorithms, how the data is coded, whether there been any tests whether decisions proposed by the algorithms are correct. Such studies would demand the expertise of tech specialists so that they can provide insights as to the details how the algorithms are built and how they operate. Such studies have not been undertaken within the context of this study. This study is not meant to be empirical. It is rather meant to be analytical, where traditional legal analytical method is applied. The research questions asked in this study call for this method. However, to the extent possible they study takes note of literature that offers empirical information about available technologies.

The aim of the study is to explain whether and how human rights law imposes any limits that might make the technologies or certain aspects of them problematic. A reflection is also due as to whether some aspects of the technologies are actually beneficial from a human rights law perspective for any affected individuals. This type of assessment - referred to in legal scholarship as the legal-analytical method - requires a solid understanding of human rights law standards as developed by relevant authorities, such as the European Court of Human Rights (ECtHR). It involves carefully explaining these standards, as articulated in case law and other authoritative sources, and exploring how they might apply to a new and emerging area.

The focus will be on EU and Council of Europe law. Since these two legal systems often overlap and influence each other, both are considered. In some areas, EU law is more relevant, while in others – especially those focused on human rights – the emphasis is on the European Convention on Human Rights (ECHR) and how it is interpreted by the European Court of Human Rights (ECtHR). At the time of writing the ECtHR has not issued a judgment that concerns the application of the ECHR to possible harm arguably caused by an AI system. The Court has issued various judgments regarding new technologies, such as

mass surveillance and retention of data,[37] facial recognition,[38] use of GPS,[39] publishing information on internet,[40] or usage of applications on mobile phones.[41] There is also a well-developed body of case law by the Court regarding data protection.[42] The methods and the principles of reasoning developed in these areas will be an important point of reference. Such methods and principles include legality, suitability, necessity and proportionality that will be explained in Section 6 below.

References to the more detailed and specific EU law is also warranted given the Court's argumentative approach whereby the Court uses external legal frameworks (notably EU law) to interpret the ECHR.[43] The ECtHR also uses other CoE treaties to interpret the ECHR. Here it should be observed that this study does not offer an analysis of the role of the EU Charter of Fundamental Rights. This could be an object of a separate study.

Many studies have looked at AI systems through the lens of **data protection laws**, which set rules for how personal data can be collected and used.[44] This is important because AI systems often rely on large amounts of personal

[37] See e.g. *Centrum för Rättvisa v Sweden* [GC] App no 35252/08, 25 May 2021; *Big Brother Watch and Others v the United Kingdom* App no 58170/13, 62322/14 and 24960/15.

[38] *Glukhin v Russia* App no 11519/20, 4 July 2023. See also D Murray, 'Facial Recognition and the End of Human Rights as we Know Them?' (2024) *Netherlands Quarterly of Human Rights* 1.

[39] *Ben Faiza v France* App no 31446/12, 8 February 2018.

[40] *Biancardi v Italy* App no 77419/16, 25 November 2021 (the first case in which the Court examined whether a journalist's civil liability for not de-indexing information published on the Internet had been compatible with Article 10 (freedom of expression) of the ECHR).

[41] *Magyar Kétfarkú Kutya Párt v Hungary* [GC] App no 201/17, 20 January 2020.

[42] See Guide to the Case-Law of the European Court of Human Rights Data Protection (31 August 2024).

[43] This is also the case in the area of private life and protection of personal data. See e.g. *Le Marrec v France* (dec.) App no 52319/22, 5 November 2024 with reference to the EU Data Protection Regulation. *Le Marrec v France* concerned the collection and use of applicant's IP address during his log-on to the website of the Family Allowances Office, disclosing his location. SSRN See Use of Council of Europe treaties in the case-law of the European Court of Human Rights (CoE, 2011); Erik Voeten, 'Why Cite External Legal Sources? Theory and Evidence from the European Court of Human Right' in Chiara Giorgetti and Mark Pollack (eds), *Beyond Fragmentation: Cross-Fertilization, Cooperation, and Competition among International Courts* (Cambridge University Press, 2022) 162.

[44] Data protection regulations as a legal response have been useful for specifying how data can be collected, processed and stored. These regulations also provide rights to the data subjects. See Lena Enqvist, 'Rule-based versus AI-driven Benefits Allocation: GDPR and AIA Act Implications and Challenges for Automation in Public Social Security Administration' (2024) 33(2) *Information and Communications Technology Law* 222.

data.[45] The regulation of data protection is important with its detailed rules for lawful data processing and the rights of data subjects.[46] For instance, 22(1) of the General Data Protection Regulation (GDPR) prohibits individual decisions from being taken in a fully automated manner.[47] However, experts agree that the challenges posed by AI go **beyond data protection**.[48] Also, data protection law has its own limits.[49] This study will not go into detail with data protection matters, in order to retain a more focused analysis.[50] Neither will the study address the liability regime that the EU is currently developing.[51]

[45] Recital (10), EU AI Act. See also Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 223). See Lina Jasmontaite-Zaniewicz and Julia Zomignani Barboza, 'Disproportionate Surveillance: Technology-Assisted and Automated Decisions in Asylum Applications in the EU' (2021) 33(1) *International Journal of Refugee Law* 89.

[46] McGregor, Murray and Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 *International and Comparative Law Quarterly* 314, 320, 325.

[47] There are many exceptions, however. See Regulation (EU) 2016/679 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, (General Data Protection Regulation (GDPR)). Regarding Article 22 of the GDPR see Case C-634/21 *OQ v Land Hessen* (Schufa) [2023]; see also Case C-203/22 *CK v Magistrat der Stadt Wien and Dun & Bradstreet Austria GmbH*, 27 February 2025.

[48] Elisabeth Renieris, *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse* (MIT Press 2023); Francesca Palmiotto, 'When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis' (2024) *German Law Journal* 1.

[49] Simona Demkova, 'The EU's Artificial Intelligence Laboratory and Fundamental Rights' in Melanie Fink (ed), *Redressing Fundamental Rights Violations by the EU: The Promise of the 'Complete System of Remedies* (Cambridge University Press, 2025*)* 391, 415: '[…] data protection law guarantees data subjects' rights with substantial number of exception and limitations, which is evident from the long list of exceptions to the general prohibition on processing special categories of personal data in Article 9(2) GDPR. Such a priori exceptions might not be subject to the same proportionality and necessity test as permissible limits to fundamental rights are under Article 52(1) CFR.'

[50] See C Novelli, F Casolari, P Hacker, G Spedicato and L Floridi, 'Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity' (2024) 55 *Computer Law and Security Review* 1.

[51] European Parliament Legislative Resolution of 12 March 2024 on the Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products, P9_TA (2024) 0132 (12 March 2024); see also Defective Products: Revamped Rules to Better Protect Consumers from Damage, European Parliament (2024); Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive), COM (2022) 496 final (Sept. 28, 2022).

AI has also been studied from the perspective of *human right law*, which in turn has incorporated data protection as an external framework to draw from when interpreting, for example, the right to private life.[52] Still, the full role of human rights law in regulating AI is not yet fully understood and needs further exploration.[53] In addition to human rights law and data protection, other approaches have been used to assess AI systems: *algorithmic accountability* and *ethics*.[54] More recently, both the **EU** and the **Council of Europe** have started to adopt specific legal frameworks to regulate AI directly.

It is therefore currently possible to identify four frameworks (data protection, human rights, ethics and AI regulation) for addressing harms related to AI systems.[55] They are indeed related. Yet, it also important to separate them.[56] In this study, the AI systems are reviewed from the perspective of two legal

[52] The GDPR sets out a comprehensive framework of principles regarding data protection, including the principles of transparency, purpose limitation and data minimisation (Article 5(1)(a) of the GDPR) to be applied when processing personal data. These principles can be important for the interpretation of Article 8(2) ECHR. See for example the judgment by the District Court of the Hague, SyRI , 6 March 2020, para 6.41.

[53] There are two lines of contributions: one arguing that human rights law is not helpful and another one arguing that human rights law is well-equipped to face harms that might be caused by AI systems. Examples of the first one is Sue Anne Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (2022) 40(1) *Nordic Journal of Human Rights* 216; in-Yan Liu, 'AI Challenges and the Inadequacy of Human Rights Protections' (2021) 40 Criminal Justice Ethics 2; Hin-Yan Liu, 'The Digital Disruption of Human Rights Foundations' in Mart Susi (ed), *Human Rights, Digital Society and the Law: A Research Companion* (Routledge 2019). For scholarship that fits within the second line (i.e. human rights law is useful), see Lorna McGregor, Daragh Murray and Vivian Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 *International & Comparative Law Quarterly* 309; Eileen Donahoe and Megan MacDuffee Metzger, 'Artificial Intelligence and Human Rights' (2019) 30 *Journal of Democracy* 115.

[54] Ethics has been proposed as a framework given the arguable failure of other regulatory frameworks. See Jacqui Ayling and Adriane Chapman, 'Putting AI Ethics to Work: Are the Tools Fit for Purpose?' (2021) AI and Ethics; Alessandro Mantelero, 'Regulating AI within the Human Rights Framework: A Roadmapping Methodology' (2020) European Yearbook on Human Rights 477. 'The algorithmic accountability approach focuses on how transparency, explainability and understandability in the design and implementation of algorithms enable individuals to exercise their rights' See Nicholas Diakopoulos, 'Algorithmic Accountability: Journalistic Investigation of Computational Power Structures' (2015) 3 *Digital Journalism* 398; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 *Harvard Journal of Law & Technology* 841.

[55] McGregor, Murray and Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 *International and Comparative Law Quarterly* 314; Yeung, Howes and Pogrebna, 'AI Governance by Human Rights-Centered Design, Deliberation, and Oversight: An End to Ethics Washing' in Dubber, Pasquale and Das (eds), *The Oxford Handbook of Ethics of AI* (2020) 77.

[56] When it comes to ethics and ethical guidelines, these might be characterised with different theoretical approaches, which can limit their utility. See Ad Hoc Committee on AI (CAHAI) Elaboration of the feasibility study, CAHAI(2020)08-fin, 15 June 2020.

frameworks: human rights law and the AI systems regulations adopted by the EU and the Council of Europe. Human rights law will be limited to the ECHR as interpreted by the Court. Different national systems intended to protect human rights are not covered. Neither is the European Charter of Fundamental Rights. Focusing on the ECHR allows for a more manageable legal analysis within the constraints of this study. While the relationship between the ECHR and the Charter remains an important topic, a comparative or dual-framework approach would require separate and extensive treatment beyond the scope of this report.

Even the engagement with the ECHR will be limited, since admissibility requirements will not be covered.[57] This is a serious limitation. The opacity of AI systems can pose a challenge to their judicial review, since specific individuals are not likely to have victim status and have their claims declared admissible.[58] The ECHR is not a system with an avenue to public interest litigation. For this reason, an individual who seeks to make a claim before the ECtHR that his or her human rights have been violated, needs to be prove that specifically he or she has been directly affected. If direct specifical harm not possible to demonstrate, the victim status requirement cannot be met. However, some national systems might have such an avenue and use the ECtHR's reasoning to review human rights claims.[59] For this reason any procedural admissibility restrictions, such as victim status, will be ignored in the context of this report.

The structure of the report is designed to reflect the limitations and challenges outlined above. Section 4 examines the EU's regulation of AI systems in the context of asylum, with particular attention to key provisions of the EU AI Act. Section 5 turns to the Council of Europe's AI Framework Convention. Section 6 places the European Convention on Human Rights (ECHR), as interpreted by

[57] Admissibility requirements and other preliminary procedural issues could be an object of a separate study.

[58] As explained here Mando (Adamantia) Rachovitsa and Niclas Johann 'The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch Syri Case (2022) *22 Human Rights Law Review* 10, the 'opacity of such algorithmic systems could pose a serious challenge to their judicial review, since individuals citizens will most likely to be unable to prove victim status and thus cannot raise admissible claims.' The authors also say 'But see from a data protection point of view the potential of Article 80 of the GDPR on representation of data subjects', which means that under the data protection regime, there are no issues with standing.

[59] The very same issue has arisen in the context of climate change litigation. See Supreme Court of the Netherlands (civil division), *The State of the Netherlands and Stichting Urgenda*, 20 December 2019, ECLI:NL:HR:2019:2007 as an example of public interest litigation allowed under the Dutch law, which can be also relevant for challenging AI systems. See also from a data protection point of view the potential of Article 80 of the GDPR on representation of data subjects.

the European Court of Human Rights (ECtHR), at the centre of the analysis. The final section offers concluding reflections and presents a set of policy recommendations.

It must be acknowledged that various rights protected under the European Convention on Human Rights (ECHR) can be negatively impacted by the use of AI systems. These include fundamental interests safeguarded by the prohibitions against inhuman or degrading treatment (Article 3 ECHR), the right to respect for private life (Article 8 ECHR), the right to non-discrimination (Article 14 ECHR), the presumption of innocence (Article 6 ECHR), the right to liberty (Article 5 ECHR), the right to effective remedy (Article 13 ECHR) and the freedom of movement (Protocol No. 4 ECHR). However, given this study's specific focus on asylum decision-making - and the need to maintain a manageable scope - Section 6 concentrates on Article 3 ECHR, which imposes obligations on States to protect asylum seekers from the risk of ill-treatment upon removal (i.e., the principle of *non-refoulement*), a core aim of the asylum determination process. Section 6 also addresses Article 8 ECHR, which, like Article 3, has been interpreted to include procedural obligations. These obligations may be engaged in the context of any national decision-making process, potentially influenced by AI systems, that might be characterised with procedural flaws. As a result of these necessary limitations in scope, other important rights, such as the prohibition of discrimination under Article 14 ECHR, are not addressed in this report.[60]

[60] Yet, see the analysis in Section 4.4.1.2. See more generally Discrimination, artificial intelligence, and algorithmic decision-making (Council of Europe, 2018). It should be also kept in mind that the role of non-discrimination in the context of asylum and migration, has been very limited. See Vladislava Stoyanova, 'Discrimination based on immigration status under the ECHR: Navigating between the factual versus the normative and the comparison-based versus the minimum-based treatment' (2025) 25(2) *International Journal of Discrimination and* Law 182.

4. Regulation of AI systems relevant to asylum by the EU AI Act

In 2024 the EU adopted its AI Act, a new law intended to enshrine 'a uniform legal framework […] for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems).'[61] One of the goals of the AI Act is consistency in AI regulation in the EU in order to use the benefits of the technology, while at the same time minimising risks.[62]

The Act has two main aims:

- First, high level of protection of fundamental rights and protection from the possible negative effects of AI
- Second, supporting and boosting innovation.

The Act refers to 'trustworthy and safe' AI systems that are 'developed and used in accordance with fundamental rights obligations.'[63] The Act acknowledges that AI systems 'may generate risks and harm to public interest and fundamental rights.'[64] However, the Act also notes that AI does provide 'competitive advantages', including in the areas of 'public services, security, justice'.[65] It refers to 'unlocking the potential of digital transformation'[66] and 'innovation'.[67] It is therefore important to understand the AI Act in light of its context. The Act is predominantly an internal market instrument based on Article 114 of the Treaty on the Functioning of the EU (TFEU). It treats AI systems

[61] Recital (1), Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonized rules on artificial intelligence (EU AI Act). The Act entered into force on 1 August 2024.

[62] White Paper on Artificial Intelligence – A European Approach to Excellence and Trust (European Commission 2020) COM (2020) 65.

[63] Recital (1), EU AI Act.

[64] Recital (5), EU AI Act.

[65] Recital (4), EU AI Act.

[66] Recital (8), EU AI Act.

[67] Recital (25), EU AI Act.

like products that must meet certain safety and quality standards.[68] That's why the law focuses on the companies and public bodies that develop **(providers)** and use **(deployers)** AI systems.[69] The idea is that if developers follow clear rules from the start, i.e. *ex-ante* requirements when they develop products, any risks from these products will be reduced. The aim isn't to remove all risks, but to minimize and manage them responsibly.

When it comes to asylum procedures, the AI Act has specific relevance. Section 4.1. explains the definition of AI systems. Section 4.2. shows that asylum authorities as public bodies, can be both providers and deployers of AI systems. Section 4.3. focuses on the classification of 'high risk' AI systems used in the area of asylum. It also clarifies the possibility for exceptions from the 'high risk' qualification, a possibility that is open for the use of systems in the area of asylum. Finally, Section. 4.4 focuses on the requirements imposed on providers and deployers of AI systems that fall within the 'high risk' qualification.

4.1 Definition of an AI system

Article 3(1) of the AI Act defines an AI system as

> a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

It follows that the AI Act does not apply to any automated system used by public authorities, since 'systems that are based on the rules defined solely by natural persons to automatically execute operations' are excluded.[70] The preamble to the AI Act clarifies that a key characteristic of AI systems is their

[68] Francesca Palmiotto 'The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation' *European Journal of Risk Regulation* (2025) 1, 8; Marco Almada and Nicolas Petit, 'The EU AI Act: A Medley of Product Safety and Fundamental Rights?' The EU AI act : a medley of product safety and fundamental rights?

[69] Article 3 of EU AI Act contains definitions of provider and deployer. Provider is defined as 'a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.' Deployer is defined as 'a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.'

[70] Recital 12, AI Act.

capability to infer. This capability to infer 'transcends basic data processing by enabling learning, reasoning and modelling.' Another key characteristic is that AI systems operate with some level of autonomy from humans. Despite these possible limitations, the definition has been considered relatively wide and meant to keep up with future technological development.[71]

4.2 The AI Act applies to private and public providers and deployers

It is important to clarify to whom the AI Act applies. The AI Act applies to *both* private and public providers and deployers inside the European Union. It also applies to providers and deployers in other countries that place AI systems on the EU market or that use a system that impacts parties in the EU.[72] The Act therefore applies to natural and legal persons, public authorities and agencies. All these entities are divided into two types: providers and deployers. Similarly to private entities and individuals, public authorities can be also providers when they develop AI systems and put them into service.[73] Similarly to private entities and individuals, public authorities can also be deployers of AI systems when they *use* the systems.[74]

According to the AI Act, if public authorities develop AI systems, these authorities are defined as providers and must comply with all the relevant requirements regarding the development of the systems. In addition, given the wide meaning of a provider, public authorities fall within the definition of a provider not only when they create and develop systems through their own IT departments. Public authorities are also providers when they commission external parties to develop AI systems, which could be against payment (e.g. conclusion of service contract) or for free (e.g. by university or research institutions). This means that 'the contracting authority [i.e. the public authorities, such as the asylum agencies], and not the contractor, is then responsible for the fulfilment of the many obligations imposed on providers.'[75]

[71] Kees Stuurman and Eric Lachaud, 'Regulating AI. A Label to Complete the Proposed Act on Artificial Intelligence' 44 (2022) *Computer Law and Security Review.* It is necessary to await the Commission's clarification of the concept of AI system in the guidelines on the practical implementation of the AI Act to be drawn up (see Article 96(1)(f) AI Act).

[72] Article 2(1), AI Act.

[73] Article 2(3), AI Act.

[74] Article 2(4), AI Act.

[75] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 9.

When public authorities simply use AI systems when performing their tasks by acquiring systems once available on the market (through for example supply contracts) or by using systems that exist under free and open-source licences,[76] then public authorities can fall within the definition of a deployer under the AI Act.[77] Once defined as deployers of an AI system, public authorities also have to comply with certain requirements (see Section 4.4.2 below).

As both possible providers and deployers, public authorities must comply with certain requirements. Even if the authorities only procure the systems, as deployers, they are an object of regulation. The regulation and the related regulatory requirements are, however, dependent on the classification of the AI system as a high-risk system.

4.3 The 'high-risk' AI system classification under the EU AI Act

The AI Act adopts a risk-based approach, imposing regulatory obligations in proportion to the level of risk associated with a given AI system.[78] This means that regulatory measures - including restrictions on development and deployment - are adjusted and tailored according to the perceived level of risk the system presents.

In accordance with this approach, the AI Act defines four levels of risks. The first level is AI systems with unacceptable risk. The second level is' high-risk' AI systems. These two types of systems will be explained below since they are an object of regulation. The third level of risk refers to AI systems that pose limited risk. Certain transparency obligations are imposed regarding these systems.[79] Finally, regarding AI systems that pose minimum risk, these can be

[76] Although Article 2(12) AI Act begins by stating that 'this Regulation does not apply to AI systems released under free and open-source licences,' it immediately clarifies that they are subject to the rules on prohibited and high-risk systems. These rules constitute the core of the AI Act regulation.

[77] In these cases, public authorities may also assume the role of provider when they substantially modify the purchased system or its purpose, or put their name or trademark on it, as provided for in Article 25 AI Act.

[78] Tobias Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU Artificial Intelligence Act Proposal' in L Colonna and S Greenstein (ed), *Law in the Era of Artificial Intelligence. Nordic Yearbook of Law and Informatics* 247.

[79] See Chapter IV of the AI Act that contains only Article 50. See also recital 132 of the AI Act about notification (natural persons should be notified that they are interacting with an AI system). The AI Act requires that users must be made aware that they are interacting with an AI system (e.g., chatbots or emotion recognition) (see Article 50(1) and 50(3)) or viewing outputs from (e.g., for GPAI or deepfakes) certain types of AI (see Article 50(2)).

developed and used subject to the existing legislation without additional legal obligations.[80] AI systems posing minimum or no risk are not required to comply with any of the requirements and standards in the AI Act.[81]

Table 1. AI Risk Levels and Corresponding Obligations under the EU AI Act

Risk Level	Examples	Obligations
Unacceptable risk	• Social scoring by public authorities • Subliminal manipulation • Harmful profiling	Prohibited or heavily restricted (generally banned unless modified)
High risk	• Biometric identification • AI in critical infrastructure • Medical devices • **Migration, asylum and border management**	Strict compliance (risk management, logs, human oversight, conformity)
Limited risk	• Chatbots requiring user disclosure	Transparency obligations (e.g. must inform users they are interacting with an AI system through labelling or other means)
Minimal risk	• AI in videogames • Spam filters • Photo filters	No special requirements under the Act (general safety provisions only, codes of conduct)

It is important to clarify the implication from the risk-based approach of the AI Act, before we examine the regulation surrounding the high and unacceptable risk categories. More specifically, the aim of the AI Act can generally be

[80] The first two levels are clear. However, there is some confusion as to the distinction between the last two levels of risk. See Tobias Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU Artificial Intelligence Act Proposal' in L Colonna and S Greenstein (ed), *Law in the Era of Artificial Intelligence*. Nordic Yearbook of Law and Informatics 247, 265.

[81] Providers and deployers of such systems can choose to voluntary comply with these requirements. See Article 95 AI Act. Articles 51 and 52 distinguish between general-purpose AI models and general-purpose AI models with systemic risks.

characterised as to *prevent regulation of AI systems that do not fall within the category of high-risk and unaccepted risk.* In this sense, the Act has been characterised as a 'regulatory shield'.[82] AI systems *not* classified as systems posing unacceptable risk or high risk, cannot be subjected to additional limitations by the Member States.[83] Member States may, however, impose additional obligations on public authorities using such systems.[84]

Another important initial clarification regarding the risk-based approach concerns the meaning of risk. Risk is not understood in a technical sense, as a risk for flaws or errors in the development of the systems. Risk refers to the *nature of the harm* that might be the consequences of such flaws and errors.[85]

Unacceptable risk – prohibition

According to Article 5 of the AI Act, certain systems pose an *unacceptable* risk and are therefore *prohibited*.[86] It is not necessary here to enumerate them by copying the text of Article 5 of the AI Act. What is important to point out is that Article 5 of the AI Act refers to 'prohibited AI practices' that can apply to any

[82] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 6.

[83] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 6: 'The most important consequence of subjecting these systems, which do not merit a high-risk rating, to the harmonised regulation of the AIA is precisely that they cannot be subject to additional restrictions by Member States. Member States may not subject their development to obligations restricting their free movement throughout the EU. Therefore, in relation to this huge range of AI systems that exist and, above all, will be developed in the future, it is more important what the AIA does not say than what it says: by subjecting them only to voluntary codes of conduct, the AIA grants complete freedom to their development. This "regulatory shield," designed not to hinder the development of AI in Europe, is arguably one of the most important effects of the AIA. It is paradoxical that the European legislator is accused of being over-regulatory when the AIA will serve, above all, to prevent the imposition of obligations on the providers of the vast majority of AI systems that will come onto the market.'

[84] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 6.

[85] 'The meaning of risk in the AI Act is therefore different from the meaning of risk applied in the technical world.' See H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation' available here Where Residual Risks Reside: A Comparative Approach to Art 9(4) of the European Union's Proposed AI Regulation by Henry L Fraser, Jose-Miguel Bello y Villarino :: SSRN

[86] See Recitals (29) (30)-(44), EU AI Act and Article 5 EU AI Act. There are exceptions listed in Article 5 of the AI Act.

area, including asylum.[87] An example of such a practice is using AI systems that classify persons based on their social behaviour with social scores. These have been referred to as social scoring.[88] It would be for example prohibited to use it to condition access to reception conditions. Another example is making risk assessments of persons to assess the risk of committing a criminal offence.[89] Such risk assessment will be prohibited, for example, for the application of exception to *refoulement*.[90] Other examples of 'prohibited AI practices' include creation of facial recognition databases, systems that infer emotions and 'real-time' remote biometric identification systems.[91]

What might be particularly pertinent to note in relationship to asylum is that biometric categorisation systems are listed in Article 5 of the AI Act as a prohibited AI practice. Such biometric categorisations systems 'categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation.' This prohibition could be pertinent if an asylum claim is, for example, based on sexual orientation or political opinion.

The prohibition, however, does not cover labelling and filtering of 'lawfully acquired biometric datasets, such as images, based on biometric data or categorising of biometric data in the area of law enforcement.' This exclusion is very illustrative of the overall regulatory approach of the AI Act, which

[87] Notably, there is a conceptual confusion in the AI Act since Article 5 refers to 'practices', while Article 6 uses the terms 'AI systems'. This raises the question as to the difference between 'practices' and 'systems'. Tobias Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU Artificial Intelligence Act Proposal' in L Colonna and S Greenstein (ed), *Law in the Era of Artificial Intelligence*. Nordic Yearbook of Law and Informatics 247, 266.

[88] The title of Article 5 of the AI that is 'prohibited AI practices' can be however misleading since social scoring is not prohibited in absolute terms under all circumstances. For example, according to Article 5(c)(i) of the AI Act social scoring is not prohibited if leads to treating persons in certain ways (even detrimental) in social contexts that are related 'to the contexts in which the data was originally generated or collected.' According to Article 5(c)(ii) of the AI Act detrimental treatment based on social scoring can be justified and proportionate and thus not prohibited. If the AI practice of social scoring falls within these exceptions, the system used to develop it might still be possible to define as 'high-risk', which will trigger other requirements under the AI Act.

[89] It would be misleading to define this practice as prohibited in absolute terms under all circumstances. The reason is that Article 5(1)(d) AI Act says that 'this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity.'

[90] For this exception, see Article 33(2) of the Convention relating to the Status of Refugees UNTS V 189, p. 137.

[91] See Articles 5(1)(h) and 5(2)-(5) AI Act as to the exceptions, i.e. when 'real time' remote biometric identification systems can be used.

makes the labelling of AI practices as prohibited misleading. Such a labelling can be deceptive since the prohibitions have exceptions whose scope can be ambiguous and context dependent.

'High-risk' – regulation

The second level of risk applies to AI systems classified as 'high-risk' systems.[92] The preamble of the AI Act clarifies that 'AI systems identified as high-risk should be limited to those than have a *significant* harmful impact on the health, safety and fundamental rights of persons in the Union.'[93] The severity of the harmful impact is therefore key. Recital (52) of the Preamble refers not only the severity of the harm, but also the probability of the occurrence of the harm: 'taking into account both the severity of the possible harm and its probability of occurrence.' The definition of risk is therefore made with reference to severity of harm and probability of the harm happening. Article 3(2) of the AI Act is also clear to the effect that risk is defined as 'the combination of the probability of an occurrence of harm and the severity of that harm'.

The AI Act does not define 'high-risk' in general terms. Article 6 with its reference to Annex III rather indicates areas where the use of AI systems can be assumed to be 'high-risk'.[94] AI systems used in migration, asylum and border control management are indicated as 'high-risk' systems in Annex III.[95] However, this does *not* mean that all AI systems used in any activities related to asylum are automatically and generally considered as posing 'high risk'. In this sense, AI systems used in the area of asylum and migration are 'high-risk' systems, *but only under certain conditions* they are to be regulated under the AI Act as systems posing 'high-risk'. This is because there are two key qualifiers for determining when systems are considered 'high-risk': one found in Annex III and the other in Article 6(3) of the AI Act. Each will be addressed in turn. The relationship between them can be illustrated in the following table:

[92] See Recitals (46), EU AI Act: 'To ensure consistency and avoid unnecessary administrative burdens and costs, providers of a product that contains one or more high-risk AI systems […], *should have flexibility with regard to operational decisions on how to ensure compliance* of a product [emphasis added].'

[93] See Recitals (46), EU AI Act and Article 6 EU AI Act.

[94] See also Article 6(1), EU AI Act with reference to Annex I that also identifies systems that are classified as high risk.

[95] See Recital (60), EU AI Act, Article 6(2) EU AI Act and Annex III. Other systems included in Annex III and also classified as high risk include AI systems that use biometrics, AI systems used for critical infrastructure, education and vocational training, employment and works' management, access to essential private and public services, law enforcement, administration of justice and democratic processes.

Table 2. Conditions Defining 'High-Risk' AI Use in Migration and Asylum Contexts

-	Cumulative conditions
Conditions to be met so that AI systems used in the area of migration, asylum and border management are regulated as 'high-risk' systems	The AI system is used for some specific purposes, such as examination of asylum applications. See Section 4.3.2.1.
	The AI system has to pass certain threshold for posing a risk for harm. See Section 4.3.2.2.

causation

Specification of the systems used in the area of migration, asylum and border control management in Annex III that are to be considered 'high-risk'

According to Annex III, 'high-risk' AI systems are systems used in the area of migration, asylum and border control management 'in so far as their use is permitted under relevant Union or national law.' The addition in the quotation can be understood as a legality requirement, i.e. the systems have to have a legal basis, including being permitted by the law.[96] Annex III then specifies four systems as related to migration, asylum and border control, that are to be considered 'high-risk'.

These four systems are: (1) AI systems intended to be used as polygraphs or similar tools; (2) AI systems intended to be used to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State; (3) AI systems intended to be used to assist competent public authorities for the examination of applications for asylum, visa or residence permits and for associated complaints with regard to the eligibility of the natural persons applying for a status, including related assessments of the reliability of evidence; (4) AI systems intended to be used for the purpose of detecting, recognising or identifying natural persons, with the exception of the verification of travel documents.

[96] The wording of Annex III in the AI Act is however very awkward since it can be also read to the effect that if the systems are not permitted under Union or national law, they cannot fall within the classification of 'high-risk' systems. On the other hand, if a system is not permitted under EU or national law, it can be considered as unlawful in any case.

The listing of these four specific systems suggests that other AI systems used in the field of asylum and migration – if they do not fall into one of these four categories – are not considered 'high-risk' under the AI Act.[97] However, the language used in Annex III is broad. For example, the third category refers to systems used for examining asylum applications and assessing the reliability of evidence. This means that AI systems assisting public authorities with both the legal evaluation and the factual assessment of asylum claims fall within its scope. As such, they are classified as 'high-risk' under the AI Act.

Risk threshold for the classification of a system as a 'high risk'

A system listed in Annex III is ***not*** automatically assumed to be high-risk. For example, a system used to examine asylum claims is ***not*** automatically *assumed to be a high-risk* by the mere fact that it is used in a sensitive area, such as asylum. In accordance with Article 6(3) of the AI Act, to be classified as a 'high risk' system an additional threshold needs to be passed.[98]

This provision stipulates that

> an AI system referred to in Annex III shall *not* be considered to be a high-risk where it does *not* pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.[99]

[97] The AI systems included in Annex III may increase or decrease, since new systems can be added or systems can be removed from the above-mentioned list. The addition of new examples of high-risk systems is meant to ensure flexibility as to how the technology is regulated given that it might develop very rapidly. Article 97 AI Act allows the Commission to adopt delegated acts under Article 290 TFEU to amend some of its provisions (the scrutiny by the Council and the European Parliament is still required). Annex III can be also an object of such amendments whose areas the Commission can extend (only in the framework of the eight existing areas) in accordance with the substantive criteria set out in Article 7 AIA. The Commission can also change the exception of Article 6(3) AIA.

[98] This threshold does not apply to AI systems that profile natural persons. See Article 6(3) third sentence that will be further analysed below in this subsection.

[99] The objective of this provision is to reduce the regulatory burden of the AI Act: 'the significance of the output of the AI system in relation to the decision or action taken by a human, as well as the immediacy of the effect, should also be taken into account when classifying AI systems as high risk.' See Second Presidency Compromise text (11124/22, 15.7.2022) at page 4, AIA-CZ-1st-Proposal-15-July.pdf For a more detailed analysis of Article 6 of the AI Act, see also Emilija Leinarte, 'The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act' (2024) 3 *Journal of AI Law and Regulation* 262, where the legislative process of Article 6(3) of the AI Act is described and it is explained how the co-legislators 'differed on both the threshold for risk and methodology for assessing it.'

A contrario, an AI system that assists public authorities in the examination of asylum claims is classified as a high-risk system *only* if it poses 'a significant risk of harm to the health, safety or fundamental rights of natural persons.' Here it is pertinent to clarify that it is the system *itself* that has to pose 'a significant risk', *not* the final decision taken by the public authority. The final decision (e.g., rejection of an asylum claim) might pose a risk of harm to the health, safety or fundamental rights. However, this is not what matters. What matters for an AI system to be classified as a 'high-risk' is that the system *itself* poses a risk to health, safety or fundamental rights when used to assist in the examination of the asylum claim.

This raises difficult questions of causation.[100] More specifically, it needs to be possible to demonstrate the *causal link* between the AI system and the risk of harm to health, safety and fundamental rights of persons. If the causal link between the AI system and the risk of harm is shown, the system can be classified as a 'high-risk' system. However, the relevant standard of causation is not addressed in the AI Act. Neither is the issue of how to measure the level of '*significant* risk of harm'. This paves the way for difficulties in the classification of the AI systems. Notably, Article 6(3) of the AI Act does not simply refer to risk, but to '*significant* risk of harm.'

The concept of risk is crucial to highlight here. Rather than demonstrating a causal link between the use of the AI system and actual harm, it is sufficient to show a causal link between the use of the AI system and the *risk* of harm. This simplifies the causal inquiry, as the risk of harm is hypothetical. Consequently, since demonstrating a causal link to risk is easier, classifying a system as 'high-risk' might be more straightforward. However, as previously mentioned, the risk must be 'significant'. The severity threshold implied by the standard of 'significant risk' may counterbalance the relative ease of demonstrating risk instead of actual harm.

Even if only *risk of harm* - as opposed to actual harm - needs to be demonstrated, we need to have some understanding what constitutes *harm*. This raises the question: Harm to what? Article 6(3) of the AI Act specifies 'harm to the health, safety or fundamental rights of natural persons'. It is challenging to envision scenarios where health and safety are compromised without simultaneously affecting fundamental rights. This makes the concurrent enumeration of these three concepts (i.e. health, safety or fundamental rights) difficult to comprehend.

[100] Sandy Steel, 'Legal Causation and AI' in E Lim and P Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press, 2024) 189.

The following clarification, however, might be helpful: harm to safety and health does not necessary amount to violation of fundamental rights, even if it interferes with the important interests protected by human rights law (e.g. life, private and family life).[101] For example, if the right to private life is interfered with, this interference can be *proportionate* and therefore not in violation of human rights law. This implies that a proportionality analysis must be included in the assessment of whether the AI system poses a risk of harm. Such proportionality analysis implies highly complex and context-dependent reasoning (see Section 6 for further elaboration). Therefore, the inclusion of fundamental rights in the threshold for classifying systems as high-risk, is *not* very helpful.

The inclusion of *significant risk to fundamental rights* can be interpreted to mean that the scope of possible harms is broader than just health and safety. Fundamental rights encompass more than just health and safety, extending to areas such as privacy, family life, religious beliefs, procedural rights, and more. This broader scope can be seen positively. However, as noted above, harm to interests such as privacy and family life can be *proportionate* and thus *not* in violation of human rights. Therefore, the expansion might not be very meaningful.

Article 6(3) of the AI Act stipulates that an AI system 'does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, *including by not materially influencing the outcome of decision making*'. The role of the addition reproduced in italics here has to be better understood. The addition could be understood as introducing the assumption that if the system does not materially influence the outcome (e.g. the decision to reject asylum), the system has to be assumed to be a system that does not pose a significant risk, and it is therefore not a 'high-risk' system.[102]

The words 'influence', 'significant' and 'outcome' must be elucidated for better understanding of this assumption. 'Influence' implies some causation between the use of the system and the outcome (e.g. the decision to reject asylum). To make things more confusing, this is not a causation between the use of the system and any harm to health, safety or fundamental rights. The assumption that an AI system is not a 'high-risk' system applies when the causal link between the use of the system and outcome of the decision taken by the public authorities, is weak. To put it differently, the assumption that an AI system is not a 'high-risk' system applies when this causal link is not material. The standard of not

[101] I use the terms 'fundamental rights' and 'human rights' interchangeably.

[102] Possible example could be AI systems used for translation for informative purposes or for the management of documents.

being material (i.e. 'not materially influencing') makes it easier to apply the assumption and therefore to classify the system as not being high-risk.

Meaning of 'not materially influencing the outcome of decision making'

When does an AI system 'materially' influence a decision (i.e. the outcome of the decision-making)? Article 6(3) of the AI Act introduces *assumptions* when this is *not* the case. Under the following four conditions therefore, an AI system is assumed *not* to materially influence the outcome and therefore *not* to be a high-risk system.

First, an AI system used in the area of asylum is not high-risk when is 'intended to perform a narrow procedural task'. This follows from Article 6(3)(a) of the AI Act.[103] Recital 53 of the AI Act clarifies that that '[t]hose tasks are of such narrow and limited nature that they pose only limited risks which are not increased through the use of an AI system in a context that is listed as a high-risk use in an annex to this Regulation.' The Recital provides examples: 'an AI system that transforms unstructured data into structured data, an AI system that classifies incoming documents into categories or an AI system that is used to detect duplicates among a large number of applications.' These examples suggest that the possibility for exclusion the classification of a system as a high risk should be interpreted narrowly.

Second, an AI system used in the area of asylum is not high-risk when is 'intended to improve the result of a previously completed human activity'. This follows from Article 6(3)(b) of the AI Act.[104] Recital 53 of the AI Act can be useful for better understanding this provision since it provides some examples. Improvement of the result from previously completed human activity will, for example happen when AI systems are intended to improve the language used in previously drafted documents.

Third, an AI system used in the area of asylum is not high-risk when is 'intended to detect decision-making patterns or deviations from prior decision-making patters and is not meant to replace or influence the previously completed human assessment, without proper human review.' This is stipulated in Article 6(3)(c) of the AI Act. Recital 53 of the AI Act explains that there is a low risk here since 'the use of the AI system follows a previously completed human assessment.

[103] 'An example could be an AI system that reads and classifies applications for asylum or a system that checks data received from a migrant against the data already existing in the database.' Emilija Leinarte, 'The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act' (2024) 3 *Journal of AI Law and Regulation* 262, 278. Further examples are provided in Recital 53 of the AI Act.

[104] An example could be language improvement tools.

Fourth, an AI system used in the area of asylum is not high-risk when is 'intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III."[105] Recital 53 of the AI Act clarifies that in this situation the possible impact of the output of the system very low. Examples included 'smart solutions for file handling, which include various functions from indexing, searching, text and speech processing or linking data to other data sources, or AI systems used for translation of initial documents.' All of these examples are very relevant in the asylum decision making process. As enumerated in Recital 53 of the preamble of the AI Act, these examples can suggest a narrow interpretation of the exclusions of systems from the 'high risk' classification.

These four conditions are presented as alternatives, meaning that any one of them may be sufficient on its own - or in combination with others - to exclude an AI system from being classified as 'high-risk'. For instance, if the system performs only a narrow procedural task, this alone may justify exclusion. Similarly, if an AI system is used solely to assist in the examination of asylum claims by performing a preparatory function, it would not fall under the 'high-risk' category.

[105] See also Article 6(6) AI Act that empowers the Commission to add new or to modify these four conditions upon evidence that AI systems listed in Annex III do not pose 'a significant risk of harm to health, safety or fundamental rights of natural persons.' This can be interpreted as a possibility of deregulation. See also Article 6(7) of the Act that can be viewed as a possibility for further regulation. It obliges the Commission to amend these four conditions by deleting any of them 'where there is concrete and reliable evidence that this is necessary to maintain the level of protection of health, safety and fundamental rights.'

By means of a table, this can be presented in the following way:

Table 3. Exemptions from High-Risk Classification under Article 6(3) of the AI Act

-	Four alternative conditions	AI Act
Under what conditions can an AI system used in the asylum process be regarded as having *no* meaningful impact on the outcome, and thus fall outside the 'high-risk' classification?	The system performs 'a narrow procedural task'.	Article 6(3)(a) of the AI Act
	The system is 'intended to improve the result of a previously completed human activity'.	Article 6(3)(b) of the AI Act
	The system is 'intended to detect decision-making patterns or deviations from prior decision-making patters and is not meant to replace or influence the previously completed human assessment, without proper human review.'	Article 6(3)(c) of the AI Act.
	The system is 'intended to perform a preparatory task.'	

At this point, it is important to consider how the general rule - that an AI system is not classified as 'high-risk' if it does not pose a significant risk to health, safety, or fundamental rights - relates to the four specific conditions that also lead to exclusion. As the table above showed, these conditions include: performing narrow procedural tasks, enhancing the outcome of a task already completed by a human, detecting patterns, or carrying out preparatory tasks.[106] As previously noted, once one of the four conditions applies - such as performing a narrow procedural task - the AI system is, by definition, not classified as 'high-risk' under the AI Act. But does this also mean that such a system cannot be seen as posing a significant risk to fundamental rights? Put differently: does the fact that an AI system in the asylum context performs only a narrow procedural function mean it can never threaten fundamental rights? According to the logic of the AI Act, this may be the case. However, from

[106] This is a reflection upon the logical relationship between the first and second paragraph of Article 6(3) of the AI Act.

the perspective of human rights law, the answer is different. As Chapter 6 will demonstrate, human rights analysis is context specific. This means that human rights law does not permit the automatic or definitional exclusion of AI systems from scrutiny simply because they meet one of the four conditions.

Article 6(3) of the AI Act, which sets out the conditions under which an AI system used in the asylum context is excluded from being classified as 'high-risk', raises another important question: If a system does not materially influence the outcome of a decision, is it automatically assumed to comply with human rights law? According to the logic of the AI Act, such a system may indeed be presumed not to pose a significant risk to fundamental rights. But does this mean it can never be incompatible with human rights? From the standpoint of human rights law, the answer is more complex. As Chapter 6 will show, an AI system might not materially influence the outcome and yet it might still be contrary to human rights.

Finally, it should be emphasized that the formulation of the four conditions leaves room for interpretation. In particular, as already mentioned above, the text of Article 6(3) of the AI Act specifies and defines when an AI system does *not* pose a significant risk to fundamental rights. The text of the AI Act also specifies and defines when an AI system does not materially influence a decision. *These specifications are however contestable.* Whether a system performs a *narrow* procedural task can be a contentious question that is open to interpretation. More specifically, the meaning of 'narrow procedural task' is ambiguous. The same applies to the actual meaning of improving the result, not influencing and having only a preparatory task. This interpretational ambiguity can be addressed in the future via guidelines by the Commission.[107] The point here is that the conclusion that a system, for example, performs narrow procedural task, might not be that straightforward and might depend on the context.[108]

In conclusion, classifying AI systems used in migration, asylum, and border control as 'high-risk' entails that these systems are subject to stricter regulatory requirements. However, this section has shown that the AI Act

[107] See Article 6(5) AI Act that stipulates that the Commission no later than 2 February 2026 has to issue guidance including practical examples of AI systems that are high-risk.

[108] See also Francesca Palmiotto, 'When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis' *German Law Journal* (2024) 1, 11, where it is noted 'Worryingly, the new version of the AI Act seems to suggest that fundamental rights are unaffected when AI systems do not have a prevalent role in decision-making.'

introduces several ambiguous exceptions to that classification.[109] In addition, Article 2(3) of the AI Act contains a broader exemption: it excludes from the Act's scope any AI systems used solely for military, defence, or national security purposes. Crucially, because migration, asylum and border management are often linked to national security concerns, AI systems used in these areas could potentially fall under this exception – thereby escaping regulation altogether.[110]

Profiling

As already clarified above, a system listed in Annex III is *not* automatically assumed to be high-risk. A system therefore used to examine asylum claims is *not* automatically *assumed to be a high-risk* by the mere fact that it is used in a sensitive area, such as asylum. The reason is that Article 6(3) of the AI Act imposes threshold conditions that need to be fulfilled for a system to be classified as a 'high risk'. These thresholds, however, do not apply to AI systems that profile natural persons. Article 6(3) third sentence stipulates that 'an AI system referred to in Annex III shall *always* be considered to be high-risk where the AI system performs profiling of natural persons.' It then follows that if an AI performs profiling, it is assumed to pose significant risks of harm to the health, safety or fundamental rights.

The definition of profiling is important here. Recital 53 of the AI Act indicates that profiling is to be defined in accordance with Article 4 (4) of Regulation (EU) 2016/679 or Article 3(4) of Directive (EU) 2016/680 or Article 3(5) of Regulation (EU) 2018/1725. These provisions define profiling as

> any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

If an AI system used in the asylum process includes automated processing of personal data to analyse or predict, for example, behaviour or movements, such a system has to be assumed to be a 'high risk' and be an object of regulation as envisioned by the AI Act.

[109] The application of the exceptions is not subject to administrative authorisation, as initially proposed by Parliament. It is rather left to the assessment of the system provider.

[110] On this point, see also Mario Pasquale Amoroso, 'Intelligent Borders: Exploring the Suitability of Artificial Intelligence Systems in Refugee Status Determination under International Law' (2024) *Refugee Survey Quarterly* 12.

Procedural safeguard for the classification

While the interpretational ambiguities that might exclude the classification of AI systems as 'high-risk' might be difficult to resolve, it is important to note that Article 6(4) of the AI Act introduces a procedural safeguard. As a result, providers of an AI system used in migration, asylum and border control management *who consider* that the system does not pose a high-risk 'shall document its assessment.'[111] Such documentation should happen before the system is put into service. Article 6(4) of the AI Act also adds that '[u]pon request of national competent authorities, the provider shall provide the documentation of the assessment.' Providers have to also register the system.[112]

This means that excluding an AI system from the 'high-risk' category does not require prior administrative approval. Instead, it is up to the provider to assess whether one of the four exclusion conditions applies – for example, if the system performs only a narrow procedural task.[113] Private companies can act as providers of AI systems, which may later be used by public authorities, including in the processing of asylum claims. If a provider determines that its system does not fall under the 'high-risk' classification, it is required to document this assessment. In practice, this means the provider must explain why it believes the system does not pose a high level of risk. This documentation must be made available upon request by national competent authorities. This requirement serves as an important safeguard by introducing a degree of transparency into the process.

Crucially, there is no independent review of a provider's decision to classify an AI system as not 'high-risk'. This has significant implications for systems used in migration, asylum, and border control. While their inclusion in Annex III of the AI Act might suggest they are automatically treated as high-risk, this assumption is undermined by Articles 6(3) and 6(4), which allow for exceptions. As a result, these systems are not necessarily presumed to be high-risk from the outset, and the procedural safeguards surrounding their classification are weak.

[111] The provider is also subject to the registration obligation set out in Article 49(2) AI Act. For the documentation and registration requirement that providers have as part of their self-assessment, see also Recital 53. It requires providers who consider that their AI systems are not high-risk to draw up documentation of the assessment which must be provided to national authorities upon request, as well as register such systems in the EU database. See also Recital 131.

[112] See Article 49(2) AI Act. The specificities of the obligation to register are not entirely clear: Does it simply imply registration of the system or also registration of the reasoning as to why the system is not a high-risk?

[113] Providers have to perform self-assessment whether the system poses a risk to fundamental rights.

Under the current framework, it is the provider who determines whether the system falls under the high-risk category. This determination is based on the provider's own specification of the system's intended use and internal assessment. There is no requirement for an objective or external review before the system is deployed. Once the provider concludes that the system does not pose a significant risk to health, safety, or fundamental rights, it is exempt from the obligations set out in Chapter III of the AI Act.[114] Notably, this self-assessment does not need to be verified before the system is put into use - deployment can occur immediately after the provider's classification decision.

Because the conditions for when an AI system is not considered 'high-risk' are difficult to interpret, assessing whether a system falls into this category is complex. As noted above, this kind of assessment involves addressing challenging questions - such as whether the system has a causal impact on outcomes, whether its use is proportionate, how human rights are understood, and what qualifies as a procedural task. Importantly, private companies - the ones that might provide the systems and that have to make these assessments - are not responsible for evaluating compliance with human rights law. All of this contributes to weak procedural safeguards under the current framework.

4.4 The requirements for AI systems if classified as 'high-risk'

Requirements for providers of AI systems

High-risk AI systems are allowed; however, they must comply with certain *ex-ante* rules. This means that before being used, providers must comply with specific requirements outlined in Chapter III of the AI Act. These requirements include (1) establishing risk management system; (2) ensuring data governance; (3) keeping technical documentation and records; and (4) maintaining transparency, human oversight, accuracy, cybersecurity, and robustness.[115] These four requirements will be explained below.

High-risk systems must also be registered in a public database. However, if the AI system is used for law enforcement and migration, only the supervisory authority will have access to the database.[116] This means that AI systems used

[114] S Wachter, 'Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the EU, the United States, and Beyond' (2024) 26(3) *Yale Journal of Law and Technology* 671, 685.

[115] See Articles 9-15, AI Act.

[116] Article 49(4), AI Act.

for migration-related purposes do not need to be included in a public database.[117]

Table 4. Ex-ante requirements for AI systems

Ex-ante requirements for AI systems used in the area of asylum if classified as high-risk
Establishing risk management system
Ensuring data governance
Keeping technical documentation and records
Maintaining transparency, human oversight, accuracy, cybersecurity, and robustness

Risk management system

Article 9 of the AI Act requires the establishment, implementation, documentation and maintenance of risk management system throughout the lifecycle of the high-risk AI system. The key question here concerns the *scope* of risks and the *nature* of the risks for which risk management system is required. This question is examined in detail below. To that end, the following four sub-sections provide a systematic analysis, each anchored in specific quotations from Article 9 that illustrate the scope and nature of the risks the risk management system is intended to cover.

'known and the reasonably foreseeable risks' that can be 'reasonably mitigated'

The AI Act does not specify *how* providers should identify and estimate risks. As part of their risk management system, providers are obliged to identify and analyse 'known' risks that 'high-risk' systems can pose to health, safety or fundamental rights. How could the term 'known' risks be understood? A risk could be defined as known if 'the harm has occurred in the past or is certain to occur in the future.[118]

[117] See Article 71 AI Act. These systems will be contained in a non-public section of the database. See Ludivine Sarah Stewart, 'The Regulation of AI-based Migration Technologies under the EU Act: (Still) Operating in the Shadows' (2024) 30 *European Law Journal* 122, 135: 'The exclusion of the areas of migration, asylum and border management regarding the obligation to register in the public section of the EU database removes an important tool for accessing information.' This weakens the prospect of individuals raising legal challenges.

[118] J Schuett, 'Risk Management in the Artificial Intelligence Act' (2024) 15 *European Journal of Risk Regulation* 367, 376. Schuett argues that 'to avoid circumventions, "known" refers to what an organisation could know with reasonable effort, not what they actually know. For example, a risk should be considered known if there is a relevant entry in one of the incident databases or if a public incident report has received significant media attention.'

Providers are required to identify and assess 'reasonably foreseeable risks'. However, the concept of such risks is open to debate. It relates to harms that have not yet occurred but could be predicted. Predicting them, however, requires time, effort, and resources. This raises a key question: how much effort is a provider expected to invest in identifying these risks? The AI Act does not offer a clear answer. A further complication is that the more thoroughly a provider investigates potential risks, the more likely it becomes – after the fact – that those risks will be seen as having been reasonably foreseeable. This creates uncertainty about whether it is in the provider's interest to actively search for risks at all.

As a result, the overall framework offers little in the way of legal clarity or predictability. As Schuett has observed, '[o]n the one hand, providers need legal certainty. In particular, they need to know when they are allowed to stop looking for new risks. On the other hand, the AI Act should prevent situations where providers cause significant harm but are able to exculpate themselves by arguing that the risk was not foreseeable. If this were possible, the AI Act would fail to protect health, safety and fundamental rights.'[119] Schuett continues to suggest that 'a possible way to resolve this trade-off is the following rule of thumb: the greater the potential impact of the risk, the more effort an organisation needs to put into foreseeing it.'[120]

The severity of the potential harm is thus suggested as a metric for deciding how much efforts a provider should invest in identifying risks. While this approach may provide some guidance, it remains limited. The concept of harm, particularly in the contexts where AI systems influence human decision-making, is complex and still subject to debate. The impact of AI driven decisions on individuals or groups can be multifaceted, making it difficult to fully anticipate.

Further complicating matters, Article 9(3) of the AI Act introduces additional uncertainty related to the definition of risks that 'high-risk' systems may pose. It stipulates that the risks for which providers are required to have risk management system 'shall concern only those which may be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or the provision of adequate technical information.' It then follows that risk identification and analysis is limited not only to 'the known and the foreseeable risks', but also to risks that can be reasonably mitigated or

[119] J Schuett, 'Risk Management in the Artificial Intelligence Act' (2024) 15 *European Journal of Risk Regulation* 367, 376.

[120] J Schuett, 'Risk Management in the Artificial Intelligence Act' (2024) 15 *European Journal of Risk Regulation* 367, 376: 'For example, it should be extremely difficult for a provider to credibly assure that a catastrophic risk was unforeseeable.'

eliminated. This means that if there is a risk that is known, but it *cannot* be reasonably mitigated in the design of the AI system, the provider is *not* required to include this type of risk in its risk management system.

'intended purpose'

According to Article 9(2)(b) of the AI Act, if the system is not used as intended or is misused in an unforeseeable way, any ensuing risks do *not* have to be identified. This guarantees that providers are under the obligation to identify and analyse risks that *they can control*. This ensures better legal certainty for providers. Providers are required to identify and estimate risks that may emerge when the high-risk system 'is used in accordance with its intended purpose, and under conditions of reasonably foreseeable misuse.'[121] To comply with this requirement, providers must identify potential users, the purposes for which the AI system can be used and possible misuses of the system.

The regulatory approach described above becomes problematic when an AI system is used for purposes different from the originally intended one - especially if risk assessments were conducted solely in relation to that intended purpose. An additional unresolved question is whether the potential for a system to be repurposed for multiple uses should itself be treated as a factor that heightens the overall risks.

'other risks' identified after the system is put into service

Since the risk management system is expected to be implemented throughout the entire lifecycle of the high-risk AI systems, risks must be identified and evaluated even *after* the system is developed or brought to the market. As a result, providers are required to establish a post-market monitoring system.[122] The establishment and the documentation of such a post-market monitoring system must be done 'in a manner that is proportionate to the nature of the AI technologies and the risks of the high-risk AI system.' The reference to proportionality suggests that providers have flexibility in determining the specifics of the post-market monitoring system, depending on the context.

To ensure effective post-market monitoring, providers need data regarding the operation of the AI systems. But how can they get the data? Article 72(2) of the AI Act addresses this by stating that relevant data '*may* be provided by deployers or […] *may* be collected through other sources.' The use of 'may' is somewhat concerning, as it raises uncertainty about whether providers will have access to the data to monitor the system.

[121] The terms 'intended purpose' and 'reasonably foreseeable misuse' are defined in Article 3(12) and (13) of the AI Act.

[122] See Article 9(2)(c) and Article 72(1) AI Act.

'targeted risk management measures' and 'acceptable residual risks'

So far, the risk management process has been described. In this section, the focus shifts to what needs to be done if risks are identified. Providers are obliged to take 'appropriate and targeted risk management measures' to address the risks. Importantly, these measures are not required for all risks, but only for those that are reasonably foreseeable and can be reasonably mitigated, as clarified earlier.

These measures must strike a balance between two things: the burden placed on providers when regulating AI systems, and the benefits of reducing potential risks.[123] This balancing requirement effectively limits how far risk management measures can go. In other words, if a measure creates too much burden for providers, that burden must be weighed when deciding whether the measure is truly necessary. This leads to an important question: who gets to decide whether the balance between protecting human rights and limiting burdens is appropriate? The answer is - the providers themselves.

Risk management measures must also identify 'residual risks'. The AI Act does not provide a specific definition of this term. However, 'residual risks' is a commonly used term in other sectors, such as environment protection, cyber security, medicine, disaster management, and food safety.[124] It can be generally defined as whatever risk remains after risk controls have been put in place to mitigate inherent risks of an activity.[125] According to this definition, such residual risk are considered acceptable.

The AI Act does not define what makes residual risks 'acceptable'.[126] Stepping back, this reflects a broader position within the Act: even high-risk AI systems may be allowed to operate despite posing some level of residual risk. As long as those risks are deemed acceptable, the system is not prohibited under the Act. Instead, providers are required to implement risk management measures to mitigate them. These measures must be understood in light of the AI Act's overarching principle that regulatory burdens – including the costs of risk

[123] This is my interpretation of Article 9(4) AI Act, a provision that is extremely convoluted and close to incomprehensible.

[124] See Andrew Gorecki, C*yber Breach Response That Actually Works Organizational Approach to Managing Residual Risk* (Wiley, 2020).

[125] H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation' available here Where Residual Risks Reside: A Comparative Approach to Art 9(4) of the European Union's Proposed AI Regulation by Henry L Fraser, Jose-Miguel Bello y Villarino :: SSRN

[126] Article 9(5) AI Act.

management -must be proportionate to the risks involved.[127] This framework suggests that certain residual risks to fundamental rights can be tolerated under the Act. However, whether such tolerance is compatible with human rights law is a separate question. Individuals may still suffer harm from these risks, which could amount to a violation of human rights - a concern I will return to in Section 6.

The regulatory approach in the AI Act therefore acknowledges that it is not always possible to eliminate risks entirely.[128] Consequently, residual risks may be deemed acceptable if they are proportionate to the benefits of a 'high-risk' AI system.[129]

As mentioned above, the AI Act does not define 'residual risk'. Neither does it define acceptable 'residual risks'. However, it does regulate what providers must do with these risks.[130] In particular, the second sentence of Article 9(5) AI Act indicates that risk management measures shall be undertaken *so that* any residual risk can be considered as acceptable. These risk management measures are (a) measures for elimination and reduction of risks '*as far as technically feasible* through adequate design and development of the high-risk AI system'; (b) '*where appropriate*, implementation of adequate mitigation and control measures addressing risks that cannot be eliminated'; (c) providing information[131] and *where appropriate* training to deployers.

Several observations can be made regarding these risk management measures. First, Article 9(5) AI Act does not specify measures for risk management. It then follows that there is no regulation as to the actual measures. Second, the AI Act's approach to risk management - framing residual risk as potentially 'acceptable' - does not appear to impose meaningful constraints on providers. For instance, the qualification 'as far as technically feasible' grants providers wide discretion, limiting the enforceability of obligations. Third, vague terms like 'where appropriate' further increase provider discretion, allowing them to

[127] See the EU Commission Explanatory Memorandum, COM(2021) 206 final, 21 April 2021, para 3.5.

[128] See e.g. Cindy Jardine et al, 'Risk Management Frameworks for Human Health and Environmental Risks' (2003) 6(6) *Journal of Toxicology and Environmental Health*, Part B 569, 572, 633.

[129] This evaluation 'depends partly on the cost and effectiveness of precautions taken relative to alternatives.' H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[130] It is difficult for me to understand the logical connection between the first and the second sentence of Article 9(5) of the AI Act. If I were to apply contextual interpretation, I would argue that the second sentence was meant to assist in the interpretation of acceptable residual risks.

[131] As to the provision of information, see also Article 13 of the AI Act.

decide both what risk management measures to adopt and whether those measures are appropriate. Fourth, it is questionable whether simply providing information or training to deployers should count as an adequate risk management measure. One may ask whether the mere act of informing others meaningfully reduces or mitigates the risk of harm. Fifth, the terms 'appropriate' and 'acceptable', as used in the AI Act, are inherently vague. For example, 'acceptable' can be interpreted through a proportionality lens - linking the risk management measure to its associated cost. This interpretation aligns with the AI Act's broader goal of balancing the benefits and risks of AI system development.

One relevant approach is to assess the acceptability of residual risks by referring to established standards.[132] Here it is relevant to observe that Article 40(1) of the AI Act establishes the following presumption: if 'high-risk' systems that are in conformity with harmonized standards, the systems shall be presumed to be in conformity with Section 2 (i.e. risk management system, data governance etc.) of the Act. The advantage of standards in general is that they are concrete in this way increasing clarity and certainty about what kinds of measures can be evaluated as decreasing residual risks "as far as possible" or to an "acceptable" level. Article 40(1) of the AI Act, however, adds 'to the extent those standards cover those requirements or obligations.' This means that it is still unclear to what extent risk management is covered by these standards.

The reference to standards might create the impression that assessing acceptable residual risk is solely a technical matter. However, 'judgments about acceptability of residual risks from high-risk AI systems are normative judgments."[133] Why? even if the provider undertakes risk management measures and predicts risks, they must still make a *normative* judgment about whether the remaining risk are acceptable. Acceptability is also context dependent. For example, an AI system that inaccurately assesses whether a refugee meets the income requirements for family reunification might be seen as posing acceptable risks. In contrast, an AI system that inaccurately assesses a refugee's country of origin, leading to a decision to reject an asylum claim (and possibly deport the individual), might be deemed to pose unacceptable risks. While risk management measures can help quantify risks and facilitate comparisons, they cannot definitively tell providers whether a specific risk, in a specific context, is acceptable.[134]

[132] H. Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[133] H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[134] H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

As Fraser and Bello y Villarion argue, a key question posed by the AI Act's risk management requirements (Article 9) is: 'how should the costs and benefits of precaution factor in the judgment of acceptability of risks?' In their view, the assessment of risk acceptability hinges on a cost-benefit analysis. Risks must be weighed against their benefits, and precautionary measures should not be so costly that they outweigh the benefits[135] In Section 6, I will revisit this position from the perspective of human rights law.

A critical issue is determining who bears the costs and who get benefits. Various stakeholders, such as AI system providers, state authorities, individuals affected by the systems, and society as a whole, might experience different effects. As Fraser and Bello y Villarion note, '[h]arms and benefits are not necessary, mensurable, and they are not always experienced by the same person.'[136] For example, an AI system that helps in the assessment whether an asylum seeker should be considered as vulnerable might yield cost-saving benefits for the receiving State (an economic benefit that might enable the authorities to provide better reception conditions for more individuals), but it may also pose risks of discrimination for certain groups. This implies that the nature of the harm or risk is both pecuniary (financial) and non-pecuniary. Furthermore, the harm or risk is also distributed among a potentially large number of individuals. An AI system that helps in the assessment of asylum claims might save costs for the government while still posing risks, in contrast to systems used by private entities (e.g., banks). The difference lies in the fact that any cost-saving benefits for public administrations ultimately benefit taxpayers, potentially improving well-being and human rights for various groups within society.

Data governance

As previously mentioned, 'high-risk' AI systems are allowed; however, they must comply with certain ex-ante requirements. In the previous section, I explained that providers must establish risk management systems prior to the use of the AI and throughout its lifecycle. Another requirement imposed by the AI Act upon providers of 'high-risk' systems is ensuring data governance (Article 10 AI Act).[137] Data governance aims to achieve high quality data. Data quality is central, since incorrect training data normally results in incorrect outputs by the AI system.[138]

[135] H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[136] H Fraser and J Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[137] If the data used by the AI system qualifies as a personal data in accordance with Article 4(1) of the GDPR, then the requirements imposed by the latter are also relevant (one of these is accuracy of the data). The key requirement for the application of the GDPR is whether a natural person is directly or indirectly identifiable.

[138] Frank Pasquale, 'Data-Informed Duties in AI Development' (2019) 119 *Columbia Law Review* 1917.

Article 10 of the AI Act imposes requirements concerning the quality of the data for the training of the system, for the validation of the system and the testing of the system.[139] The training data, the validation data and the testing data shall be subjected to 'management practices *appropriate* for the intended purpose of the high-risk AI system.' The use of the adjective 'appropriate' is a serious qualification regarding the 'practices' that providers need implement to ensure the data quality. This adjective suggests a context-dependent assessment. But who makes this assessment? In other words, who decides what is 'appropriate' for ensuring that the system, for example, is trained with high-quality data? The answer is: the providers.

The text of Article 10(3) of the AI Act continues to use qualifier that are similar to the above-mentioned one (i.e. 'appropriate'). The use of such qualifiers ultimately reduces the regulatory impact of the Act. Specifically, Article 10(3) states that '[t]raining, validation and testing data sets shall be relevant, sufficiently representative, and *to the best extent possible*, free of errors and complete in view of the intended purpose.' The addition of the phrase "to the best extent possible" acknowledges that perfect and complete data is an ideal that is hard to achieve.

Achieving completely unbiased or neutral data remains a significant challenge.[140] The AI Act does not assume that the data used for the training for 'high-risk' systems is free from bias. Instead, the AI Act imposes a requirement upon providers to manage biases. In particular, providers are required to have practices that concern

> *examination in view of* possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations.[141]

[139] These are referred to as 'training data', 'validation data' and 'testing data' that are defined in the AI Act. On Article 10 of the AI Act, see also Marvin van Bekkum and Frederik Zuiderveen Borgesius, 'Using Sensitive Data to Prevent Discrimination by Artificial Intelligence: Does the GDPR Need a New Exception?' 48 (2023) *Computer Law and Security Review* 1; Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257.

[140] S Wachter, 'Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the EU, the United States, and Beyond' (2024) 26(3) *Yale Journal of Law and Technology* 671, 689: 'Neutral data is a fantasy'.

[141] Article 10(2)(f), AI Act.

This can be understood as a procedural requirement upon the providers. They have to examine the data to discover possible biases. The requirement to take 'appropriate measures' to detect biases is also a procedural one.[142]

Providers are further required to take 'appropriate measures' to prevent and mitigate possible biases.[143] Article 10 of the AI Act does specify certain measures for preventing bias,[144] such as using relevant and sufficiently representative data with 'appropriate statistical properties'.

However, providers overall have a wide discretion in how they achieve high quality data to prevent bias.[145]

A clarification is also needed regarding the fact that bias is not defined in the AI Act.[146] A relevant question to ask is what standards should be used to assess whether the data is biased. Another relevant question concerns the relationship between bias and non-discrimination.[147] There might be distortions and imbalances in,[148] for example, the training data.[149] However, such distortions and imbalances in and of themselves might not constitute a legally relevant disadvantage from the perspective of anti-discrimination law.[150] This is because, in anti-discrimination law, what matters is the output of the model,

[142] Article 10(2)(g), AI Act.

[143] Article 10(2)(g), AI Act.

[144] Article (10)(3-6), AI Act.

[145] See Sandra Wachter et al., 'Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law' (2021) 123 *West Virginia Law Review* 735, 744, where it is noted that it is difficult to understand how providers can or should achieve high quality data.

[146] Sandra Wachter et al., 'Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI' (2011) 41 *Computer Law and Security Review*.

[147] If non-discrimination is approached from the perspective of EU law, it should be kept in mind that the AI system should be deployed in an area covered by the EU anti-discrimination directives, so that EU anti-discrimination law can be relevant.

[148] Imbalance is understood as over- or under-representation of certain groups in the data set, which can lead to 'a deterioration in the prediction quality for protected groups and ultimately to systematically negative distortion.' Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257, 286.

[149] The question as to access to the training data so that possible imbalances and distortions can be identified is a separate one. This question pertains to proof and evidence. It is relevant to note that in *Kelly* Case C-104/10, para 30 and *Meister* C-415/10, para 36, the CJEU held that the mere presumption of discrimination does not entitle to access to the data.

[150] Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257, 272.

not how it was trained.[151] This leads us back to the complex question of causation between harm and the AI system.

The difficult question of causation can be resolved by *reframing* the nature of the harm and *reframing* the meaning of quality in the assessment of data quality. As to the first, it is possible to accept that the harm is not understood as the negative outcome (e.g. denial of protection, denial of family reunification or denial of certain reception conditions, actual distinctions between different groups of asylum-seekers), but as the *risk* of such harm. In other words, the reframing of the nature of the harm as risk of harm, makes the establishment of causation easier. See also Section 6.3 below, where this is further addressed. As to the second, since the actual quality of the data cannot be assessed and, in any case, there is no neutral/objective standard for such an assessment, the problem can be reframed as risks for the quality of the data. Article 10 of the AI Act reframes this as 'data governance.' The reframing of the problem as not the quality of the data *per se*, but as the risks for the quality of the data, also facilitates causation.

In light of this facilitation, the two risks (i.e. risk of harm and risks to the quality of the data) are not treated any more as separate. This is important for the regulation of AI. As Hacker notes '[…] the starting point for the treatment of quality and discrimination risks is that these two risks are often so closely interwoven that they should not be considered separately and subjected to desperate regulations, but should be treated by a single piece of regulation concerning training data."[152] In relationship to AI systems used in the context of refugee status determination, however, the reframing of the harm as risk of harm is perplexing and creates a conceptual difficulty. The reason is that in light of this reframing, what then would need to be examined is whether there is risk for creating risk for *non-refoulement*. So, the reason for this perplexity is that the harm itself is risk of harm (i.e. risk of persecution). For more a more detailed discussion, see Section 6.4.2.

[151] A further clarification is due to the following effect – what matters ultimately is not even the outcome from the AI model, but the harm suffered by the individual. If the outcome from the model did not have a decisive influence on the harm, then the causal link between the model (including how it has been trained and its data quality) and the harm might be insignificant. Such a decisive influence might not exist, if the model outcome is merely a recommendation to be taken into account.

[152] Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257, 283.

Similarly, the relationship between bias and human rights law is another open question. Risks of discrimination have been understood as being linked to the quality of the data.[153] However, as Hacker has noted, such a causal link cannot and should not be assumed since risks of discrimination can exist independently of any qualities of the training data.[154] Decisions made by humans can also be biased, whether consciously or unconsciously. This has been widely acknowledged in the area of asylum.[155] The difference is that AI systems are *designed*, and as such, biases may be explicitly and directly *regulated*.[156] The AI Act does exactly this: it regulates. However, the regulation in the AI Act is framed with reference to the concept of 'appropriate', which is vague with unclear relationship with human rights and anti-discrimination law. For example, one might ask whether inappropriate should be equated with discriminatory and thus contrary to human rights law.[157]

To regulate bias more precisely, Article 10(3) of the AI Act stipulates that training, validation and testing data 'shall have the *appropriate* statistical properties, including *where applicable*, as regards the persons or groups of persons in relation to which the high-risk AI system is intended to be used.' The basis on which groups might be distinguishable is not mentioned. However, it can be assumed that distinctions based on, for example, race or nationality are covered. The standard of 'appropriate' is, however quite vague and it can be therefore questioned whether it constitutes any real regulation. The challenge is how to regulate in a predictable and consistent manner: how can we

[153] For example, if the training data contains imbalances by including only certain groups, the output of the model might disadvantage these very groups or other groups. This disadvantage can be understood as creating distinctions between different groups. Yet, even if distinctions between groups are created, these might be justifiable. In this sense, distinctions are not necessary contrary to the right to non-discrimination.
See A. Tischbirek, 'Artificial Intelligence and Discrimination' in T. Wischmeyer and T. Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020) 103, 114.

[154] Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257, 261.

[155] Andrew Schoenholtz, Jaya Ramji-Nogales, Philip Schrag, *Refugee Roulette: Disparities in Asylum Adjudication* (NYU Press 2009); Katerina Glyniadaki, 'Deciding on Asylum Dilemmas: A Conflict between Role and Person Identities for Asylum Judges' (2024) 50(12) *Journal of Ethnic and Migration Studies* 2879.

[156] Philipp Hacker, 'A Legal Framework for AI Training Data – from First Principles to the Artificial Intelligence Act' (2021) *Law, Innovation and Technology* 257, 261.

[157] To be more specific, if no 'appropriate measures to detect, prevent and mitigate possible biases' in the training, validation and testing data have been undertaken, does this directly imply violation of the right to non-discrimination? The answer is not likely to be affirmative.

establish legal criteria and standards that can be operationalized, so the data can be assessed as being of sufficient quality?[158]

Asylum and immigration are defined as high-risk and thus Article 10 AI Act is relevant. However, within the area of asylum, there may be a wide range of high-risk systems. Some are used to determine whether a person is eligible for a residence permit based on refugee status, while others may be used for different purposes. The potential harms or risks of harm, including harm to fundamental interests protected by human rights law, vary depending on the specific application. A differentiated, risk-based interpretation of Article 10 of the AI Act appears necessary depending on the area of application of the system. In other words, the intended purpose of the system matters in the assessment of what is appropriate.

This is further reflected in Article 10(4) of the AI Act, which aims to ensure the representativeness of the data 'to the extend required by the intended purpose'. Representativeness is an important criterion for assessing data quality. In particular, Article 10(4) of the AI Act stipulates that data sets should take into account 'the characterises or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used.' This criterion is especially relevant in the context of asylum and migration, where systems are used to facilitate decisions about individuals from diverse geographical regions and cultural backgrounds.

Notably, the criterion of temporal relevance is not explicitly mentioned in Article 10 of the AI Act. However, this criterion is critical in the context of asylum, as decisions need to be made in light of the present situation (i.e., whether there is a risk of ill-treatment at the time of the decision). The criterion of temporal relevance can, however, be subsumed under the requirement that the data should be relevant and contextually appropriate.

To recap, the formulation of Article 10 of the AI Act can be seen as an acknowledgement that the data can never be ideal. This explains the use of vague terms, such as 'appropriate' and 'to the best extent possible.' Article 10 of the AI Act also implies differentiations among the 'high-risk systems' with reference to the risk. In other words, within the 'high-risk' category, there are

[158] An important aspect concerns the cost of regulation, i.e. the cost of imposing qualitative standards, which might make the development of models more expensive and more difficult. It is difficult to balance this cost with harms to important interests protected by human rights law and non-discrimination. This balancing is even more difficult given the uncertain causation between the model (including the data used by the model and the output of the model) and any harm. The balancing is further complicated given that the operation of the model *as a whole* might be actually beneficial.

varying levels of risk. The regulation is designed to tolerate different margins of error depending on the type and probability of the risk involved..[159]

Technical documentation

In addition to risk management and data governance, the AI Act imposes requirements upon the providers to draw up technical documentation (Article 11), to ensure traceability of the functioning of the AI system through record-keeping (Article 12) and to ensure transparency of the systems (Article 13).

As to the technical documentation, Article 11 of the AI Act requires documentation of the system before it is put into service. The objective of such documentation is demonstrating that the AI system complies with the requirements demanded for high-risk systems and the competent authorities have information to assess compliance. These are important procedural guarantees that, as Section 6 will explain, can have a key role in assessing compliance with human rights law.

Record-keeping

Article 12 of the AI Act demands automatic recording as to how the high-risk system has been used. Logs have to be automatically recorded over the lifetime of the system. The system therefore needs to have capabilities that enable the recording of the events (logs). The relevant question here is what events (logs) are required to be recorded. No general requirement for all possible events is imposed.[160] Rather Article 12(2) of the AI Act requires that the following events are recorded. First, events relevant for 'identifying situations that may result in high-risk AI system presenting a risk' to the health or safety, or to fundamental rights, of persons. Second, events relevant for facilitating the post-market monitoring of the AI system by the providers. Third, events relevant for monitoring the operation of the AI system by the deployers. The overall objective here is tracing the functioning of the system. However, as Enqvist has observed the AI Act 'is not very detailed on the specific content of these logs, which limits the extent to which the Regulation itself provides some detail on the specific objects of the human oversight."[161]

It is notable indeed that Article 12(2) of the AI Act *does not specify exactly what needs to be recoded*. It indicates three general objectives of the recording. Yet, how the objective of 'identifying situations that may result in high-risk AI systems

[159] Luciano Floridi, 'The European Legislation on AI: A Brief Analysis of its Philosophical Approach' (2021) 34 *Philosophy and Technology* 215, 238.

[160] It might not be also feasibly possible to record everything. It can be also questioned whether a fully comprehensive record-keeping that creates a maze of logs, is helpful for better oversight.

[161] Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 523.

presenting a risk' to fundamental rights, is to be achieved specifically via record-keeping is not specified. There is no answer to the question what needs to be recorded so that it can be traced how the system has been used so that it might be possible to assess whether the system presents risks to fundamental rights.

Notably, such a specification as to what actually needs to be recorded is ensured in the AI Act *only* in relationship to remote biometric identification systems. Pursuant to Article 12(3) of the AI Act, remote biometric identification systems shall have the following capabilities so that their usage can be traced: recording of the period of each use of the system; the reference database against which input data has been checked by the system; the input data for which the search has led a match; the identification of the natural persons involved in the verification of the results.

All of the above means that Article 12 of the AI Act does *not* necessary require the providers to ensure that *all* high-risk systems have capabilities to record the input data or to record the reference database against which the input data has been checked by the system.[162] The question then that arises is the following: how could it be possible to identify systems used in the area of asylum (i.e. 'high-risks' system) as posing a risk to fundamental rights, when no records are available? Identification of possible risks to fundamental rights might remain impossible because the system has not recorded, for example, the input data.

The possibility of having records is very important from a human rights law perspective. The *ex-post* determination of a human rights law violation presupposes some causation between the AI system and the harm. The determination of causation will be difficult unless the AI system is developed in such a way so that it preserves and allows access to evidence and information that can help us to establish causation (i.e. causal information).[163] If the output of the AI system is the sole basis for a harmful outcome, the causation might be clear. An example would be a system that filters all asylum-seekers from a specific country to specific reception centres. However, if the AI system *only facilitates* decisions that then in turn might lead to harm, then the causal inquiry becomes uncertain and complicated. These will be for example systems that provide information (the system's output) that might be

[162] See also Article 19 of the AI Act that requires from providers of high-risk AI systems to keep the logs referred to in Article 12(1), automatically generated by their high-risk AI systems, *to the extent such logs are under their control.*

[163] Sandy Steel, 'Legal Causation and AI' in E. Lim and P. Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press, 2024) 189, 191.

then used by a public official in his or her decision-making (e.g. granting protection or not). The issue that arises is whether the system might have relied on flawed data and on flawed algorithms designs.[164] Unless the basis on which the system delivered its output can be reconstructed, it might not be possible to establish that the AI system had a causal impact upon the final decision (e.g. granting or rejection of asylum).

Such a reconstruction seems to depend on the possibility to know the inputs made into the AI system. If the inputs are not recorded, it is not possible to know what would have happened if a different input was made.[165] Nor might it be possible to know what would have happened (i.e., what decision would have been taken regarding for example the asylum application), if the AI system was not used at all. This record-keeping depends on how the AI system was designed. Although, as clarified above, Article 12 of the AI Act does not explicitly require ensuring capability of recording of the inputs for all 'high-risk' systems,[166] such a requirement might be indirectly imposed by human rights law.

More specification, the above-described situation reveals a case where a person's ability to successfully claim a possible human rights law violation is dependent upon the evidence-recording *by the State*. If a public institution is the provider and the deployer of the AI system, then the omission to develop a system that records the inputs (or omits to perform some other relevant record-keeping operations), might create presumption that the State is responsible for the harm. This might lead to shifting the burden of justification to the State to demonstrate that the system did not cause harm.[167]

A public institution might be, however, only a deployer of a system. Should such a presumption exist in this case? Should the burden of justification be shifted? What might be possible here is invoking a positive obligation upon the State to use systems that have capabilities of record keeping. Such a positive human rights obligation can be justified by the expectation that the State uses systems with record keeping capabilities. If such capabilities are not present,

[164] It is even difficult to define here what flawed might mean in this context to being with.

[165] If it possible to demonstrate that a different input would have led to a different output, then it might be the state official who used the system that made a mistake. In this case, it is the state official to blame, not the system and its developer. However, if the developer did not provide sufficient training instructions, then it might be still possible to blame the developer/provider.

[166] Such a requirement is explicitly imposed only in relationship to remote biometric identification systems. See Article 12(3) AI Act.

[167] For such a presumption and such a shift more generally in human rights law, see V Stoyanova*, Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (Oxford University Press 2023)168.

the burden of justification can be shifted to the State. This could be an independent procedural positive obligation, which would imply a violation based on the mere omissions of using systems without data gathering and retention features. It could be also an obligation that facilitates the establishment of facts necessary for the application of human rights in relation to AI systems.[168] Section 6 will further explain the role of procedural positive obligations.

There has been a proposal for 'counterfactual explanations' as a means of explaining opaque outputs of decisional AI.[169] However, in the context if asylum this is not a relevant proposal since in general if a person is denied protection, the negative decision does not provide a counterfactual explanation: if your claim contained x, y and z features, then the decision would have been positive.

Despite the record-keeping and data gathering requirements that might be extracted directly from Article 12 of the AI Act or indirectly via innovative and creative interpretations of human rights law, there will always be certain level of evidential uncertainty in relation to different factual questions. It might simply not be possible to demonstrate (on the balance of probability) that the system or certain features of the system caused harm. We might have a situation of systemic inability to demonstrate causation. From a human rights law perspective, two relevant questions can be asked concerning such a systemic difficulty in coming forward with facts and evidence. First, should human rights law develop causal assumptions? In particular, if an AI system has been used and harm has materialized, a presumption of responsibility might be established.[170] In this situation, one does not need to know via record keeping how the system specifically reached a decision or facilitated a decision that led to the harm. The second question is as follows: if such assumption is not desirable, should record keeping be imposed as an independent procedural obligation? The justification for such an independent positive obligation is that the reasons as to why a harmful decision was made might be central to the harm itself.

[168] Sandy Steel, 'Legal Causation and AI' in E. Lim and P. Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press, 2024) 189, 195: 'A positive duty to create products with data-gathering and retention features does not itself seem to further the safety of the product (except indirectly, in so far as it restores the threat of tort liability for unsafe AI products). Rather, it serves to facilitate the ascertainment of facts necessary for the enforcement of legal rights in relation to the product.'

[169] Sandy Steel, 'Legal Causation and AI' in E. Lim and P. Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press, 2024) 189, 192.

[170] Such a presumption of responsibility might be achieved by reversing the burden of proof or by accepting that a material increase in risk is a sufficient proof of causation.

Transparency, human oversight and accuracy

Article 13(1) of the AI Act provides that the high-risk systems shall be developed in such a way that they can 'enable deployers to interpret a system's output and use it appropriately.' The systems shall have instructions for use that shall include *inter alia* information about the limitations of the system, expected level of accuracy or information that is relevant to explain the output.[171]

These requirements for transparency are intended to ensure human oversight. The latter is another requirement imposed upon the providers[172] when they develop high-risk systems.[173] The human oversight requirement is specified in Article 14 of the AI Act. In particular, the high-risk systems shall be developed in a way so that they can be 'effectively overseen by natural persons.'[174] The requirement for human oversight is however formulated with various qualifications. For example, Article 14(3) of the AI Act refers to oversight measures 'when technically feasible' and that are 'appropriate to be implemented by the deployer'. Article 14(4) of the AI Act contains similarly vague term: 'appropriate and proportionate', which implies that natural persons shall be enabled to perform the task of oversight when 'appropriate and proportionate'. Article 14(4) of the AI Act specifies in what ways in particular natural persons can be enabled to oversee the systems. However, the general qualifier of 'as appropriate and proportionate' opens the different specified ways of oversight for various interpretations and uncertainties.

According to Article 14(4) of the AI Act, human oversight is achieved by enabling natural persons to *understand* the capacities and the limitations of the systems, to be *aware* of the risk of automation bias and to correctly interpret the output. Article 14(4) of the AI Act adds that the natural person shall be enabled to 'decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the

[171] Stefan Larsson and Fredrik Heintz, 'Transparency in Artificial Intelligence' (2020) 9 *Internet Policy Review* 1, 3.

[172] The assumption underpinning the regulation is that the providers are best equipped to know as to what information needs to be transferred to the overseers. In this sense, providers have a large scope of discretion. Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 521.

[173] See Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 518, where the interrelationship between the requirement for human oversight and the other requirements is explained. In particular, Enqvist notes '[w]ithout a sufficient level of system transparency […], the human overseers would have nothing to review.'

[174] Daria Onitiu, 'The Limits of Explainability and Human Oversight in the EU Commission's Proposal for the Regulation on AI- a Critical Approach Focusing on Medical Diagnostic Systems' (2022) 32 *Information and Communications Technology* Law 170, 181.

high-risk AI system'. The natural person tasked with the oversight shall be also enabled to 'intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.'

The AI Act imposes additional requirements regarding remote biometric identification systems. In particular, the measures for enabling the oversight of remote biometric identification systems 'shall be such as to ensure that, in addition, no action or decision is taken by the deployer on the basis of the identification resulting from the system unless that identification has been *separately verified and confirmed by at least two natural persons* with the necessary competence, training and authority.' Article 14(5) of the AI Act, however, adds that '[t]he requirement for a separate verification by at least two natural persons shall not apply to high-risk AI systems used for the purposes of law enforcement, migration, border control or asylum, where Union or national law considers the application of this requirement to be disproportionate.' It then follows that remote biometric identification systems used for migration control and asylum are *not* subjected to the requirement of separate verification by at least two natural persons. A consideration about proportionality can be an important safeguard here. However, this seems to be more like a legality requirement given the reference to Union and national law. The reason is that if EU law or national law considers the requirement for a separate verification as proportionate, it is proportionate. *A contrario*, if EU law or national law considers the requirement for a separate verification as not proportionate, it is not proportionate under the terms of the AI Act. No proportionality assessment independent from legality assessment done with reference to EU law or national law, seems to be envisioned by Article 14(5) of the AI Act.

The final requirements imposed upon providers of high-risk AI system is accuracy. In particular, Article 15(1) of the AI Act stipulates that these systems 'shall be designed and developed in such a way that they achieve *an appropriate level of accuracy*, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle.' The ambiguity of the requirement for appropriateness is notable. The Commission is tasked to 'encourage, as appropriate, the development of benchmarks' how to measure the appropriate levels of accuracy.[175]

[175] Article 15(1) AI Act.

It is also notable how the language of Article 15 of the AI Act shifts from accuracy to resilience, possibly under the understanding that accuracy as an objective is not achievable. It is rather resilience to 'possible errors, faults or inconsistencies' that can be attained.[176]

Article 15(4) of the AI Act specifically regulates high-risk AI systems that continue to learn. The provision stipulates these systems 'shall be developed in such a way as to eliminate *or reduce as far as possible the risk* of possibly biased outputs influencing input for future operations (feedback loops), and as to ensure that any such feedback loops are duly addressed with *appropriate mitigation measures*.' As already mentioned before, the meaning of 'biased output' and, in particular, its relationship with anti-discrimination law, is contentious. Risks of such bias outputs have to be reduced 'as far as possible'. Feedback loops, understood as cases 'where data outputs influence inputs for future operations'[177] are accepted as posing a risk that needs to be appropriately mitigated.

Requirements for deployers of AI systems

While providers are an object of wide-ranging regulations under the AI Act, the requirements for deployers are more limited. These can be summarised as human oversight, recordkeeping, monitoring duties, performance of fundamental rights impact assessment and registration duties.[178] The focus here will be on clarifying these requirements as imposed on public authorities when they use 'high-risk' systems.[179]

Prior to clarifying these requirements, it is relevant to highlight that the obligations set out in Articles 26 and 27 of the AI Act apply to administrative authorities as deployers whenever they use a high-risk system. The obligations therefore apply irrespective of whether the system is used in the context of a formal administrative procedure that leads to formal legal decisions (e.g. granting or refusing an asylum, visa or residence permit application),[180] or in the context of a factual administrative assessment. The decision or the action of the public authority that deploys a high-risk system might therefore have a

[176] 'Resilience' is distinguished from robustness in the AI Act. See Article 15(4) that stipulates that '[t]he robustness of high-risk AI systems may be achieved through technical redundancy solutions, which may include backup or fail-safe plans.'

[177] See Recital (67) AI Act.

[178] See Articles 26-27 and 49(3) AI Act.

[179] Some of the requirements are only or almost exclusively imposed on public authorities. See Article 26(8) and (10) and Article 27 AI Act.

[180] See point 7(c) in Annex III of the AI Act.

legal or factual nature. Articles 26 and 27 of the AI Act apply to both. The type of administrative action in question does not therefore matter.[181]

Outputs by AI systems can be considered *factual* when the system is used to discover, demonstrate, or prove a fact that is relevant to the decision.[182] Examples can include identification of the country of origin of a person, specification of the person's identity, specification of identity based on language or accent, or fraud detection of documents. Such factual outputs can be used to assess the credibility asylum seekers' claims.[183] Outputs by AI systems can be considered *legal* when the systems are used, for instance, for formulation of legal reasoning, for identification of relevant legal precedents (i.e. other relevant judgments) that might help the decision-maker to reason in the specific case, or for identification of legal provisions that might be relevant to the resolution of the specific case.[184]

The requirements upon the deployers as indicated in Articles 26 and 27 of the AI Act must be complied with both when the administrative action is fully automated and when it is only partially automated. In light of Article 6(3) of the AI Act what matters is that the AI system materially influences the decisions referred to in Annex III where the high-risk systems are enumerated.

The requirements for human oversight, recordkeeping, monitoring duties and fundamental rights impact assessment and registration duties upon the deployers, can be better explained in light of their temporal relevance. Some of them are relevant *prior* to deployment of the system; while others *during* the deployment of the system.

[181] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 10.

[182] This has been also labelled as 'automated evidence'. See Francesca Palmiotto, 'When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis' (2024) *German Law Journal* 1, 20.

[183] Niamh Kinchin, 'Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination' (2021) 37(3) *Law in Context* 45, 55.

[184] Systems are created to replicate and replace labour-intensive discovery tasks, i.e. discovering relevant documents. The system identifies such documents for review and usage in litigation. See Niamh Kinchin and Davoud Mougouei, 'What can AI do for Refugee Status Determination? A Proposal for Removing Subjective Fear' (2022) 34 *International Journal of Refugee Law* 373, 386.

Table 5. Responsibilities of Deployers Before and During Use of High-Risk AI Systems

Requirements for deployers of AI systems *prior* to the deployment of the systems	Requirements for deployers of AI systems *during* the deployment of the systems
human rights impact assessment	use of the systems in accordance with the instructions for use
	human oversight
	ensure that the input data is 'relevant and sufficiently representative'
	suspend the systems if the deployer has reasons to consider that they constitute a risk
	keep record of the logs
	inform natural persons that they are subject to the use of a high-risk systems

Requirements for deployers prior to the deployment of AI systems

Human rights impact assessment

Pursuant to Article 27 of the AI Act prior to using the system, public authorities shall perform human rights impact assessment.[185] The provision further specifies what this assessment should consist of. It is not necessary to repeat all these specifications; one of them, however, can be highlighted. In particular, the assessment of the impact must include identification of the *specific risks of harm* likely to have an impact on the categories of natural persons or group of persons. This means that if a system is to be deployed in the asylum context, the public authorities have a procedural obligation to identify the risks for the different groups of asylum-seekers. The public authorities also have to identify the measures that will be taken, if these identified risks materialize. To identify risks in the impact assessment, the deployers have to take into account the information provided by the provider regarding the design and the limitations of the AI system. This is an illustration of how the requirements upon the deployers are intertwined with those imposed upon the providers.

Registration of the system

Prior to the first use of the Annex III systems (i.e., high-risk systems where asylum fits), public authorities (only public authorities) are obliged to register

[185] Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 *Computer Law & Security Review* 106020.

such use in the European database.[186] The publicity of certain systems is, however, restricted. Migration, asylum and border control management fall within these restrictions, which means that the publicity of these systems is restricted.[187]

Requirements for deployers during the deployment of AI systems

Human oversight

As to the requirements upon deployers *during* the deployment of the system, the deployers must use the system in accordance with the instructions for use. The requirement for human oversight is important,[188] as also related to requirement for ensuring AI literacy.[189] The obligation of human oversight, however, seems to be qualified: this requirement is without prejudice to 'the deployer's freedom to organize its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider."[190] The 'without prejudice' provision is an acknowledgement that ensuring human oversight and training of personal demands resources and incurs costs. While the deployer has discretion how to spend resources, the requirement for human oversight remains intact. In this sense, Article 26(3) of the AI Act does not allow balancing between the possibly competing interests of cost/resource management and human oversight; such balancing might result in reducing oversight for the sake of saving resources. This does not seem to be allowed under Article 26(3) of the AI Act.[191]

Why is human oversight important? The rationale behind it is the understanding that 'humans are better able to make complex deductions relation to the social dimensions of law and other norms."[192] From a legal perspective, human oversight is a procedural guarantee.[193] The specificities of the oversight in terms of exactly what procedural steps need to be taken, are however left indetermined in the AI Act. So, while the answer to the why question is clear, the answer to the how question (i.e. How is oversight to be achieved?) is difficult to ensure.

[186] Article 26(8) EU AI Act.

[187] See Niovi Vavoula, *Regulating AI at Europe's Borders* Verfassungsblog (2024).

[188] Article 26(2) AI Act.

[189] See Article 4 AI Act.

[190] Article 26(3) AI Act.

[191] Article 26(2) AI Act, however, also says 'without prejudice to other deployer obligations under Union or national law', which might imply such a balancing.

[192] Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 511.

[193] Riikka Koulu, 'Proceduralising Control and Discretion: Human Oversight in Artificial Intelligence Policy' (2020) 27 *Maastricht Journal of European and Comparative Law* 720.

One possible procedural choice that can ensure oversight is that the output from the AI system (e.g. the suggestion that an asylum-seeker comes from a specific country of origin) is not directly accepted but needs to be validated by a human decision-maker. Yet, as Lena Enqvist has observed, there is no consensus on what the human overseers should exactly do in terms of procedural steps so that human oversight be achieved.[194] As a procedural requirement, human oversight cannot guarantee a result (e.g. correct decision). It can only be used as an indication that the system operates in a way likely to lead to correct decisions.[195] For this reason, concerns have been raised that the existence of this procedural safeguard framed as 'human oversight' creates a false sense of security when governments adopt algorithms in their decision-making.[196]

Relevance and representativeness of the input data

The deployer is also required to ensure that the input data is 'relevant and sufficiently representative', a requirement that is qualified with the addition that this is to be done 'to the extent the deployer exercises control over the input data."[197]

Suspension of the system

The use of the system needs to be suspended if the deployer has reasons to consider that it constitutes a risk.[198]

Record-keeping

The deployer also needs to record the logs that are automatically generated by the system.[199] This obligation is, however, qualified in two ways.
First, Article 26(6) of the AI Act requires the deployers to keep a record of the logs 'to the extent such logs are under their control'. Second, logs should be kept for at least six months, unless provided otherwise in EU or national law.

Providing information

The obligation to *inform* natural persons that they are subject to the use of high-risk systems is very important.[200] Individuals should be informed when AI systems are used to make or assist in decisions that affect them. A possible

[194] Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 514.

[195] This is generally the case with any procedural guarantees. See Section 6.4 below.

[196] Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 *Computer Law and Security Review* 1, 18.

[197] Article 26(4) AI Act.

[198] Article 26(5) AI Act. Risk within the meaning of Article 79(5).

[199] Article 26(6) AI Act.

[200] Article 26(11) AI Act.

weakness of the obligation to inform is that the AI Act does not specify what information should be communicated. Recital 93 of the Act, however, clarifies that such information should include the intended purpose of the system and the type of decisions it takes or assists in taking, as well as the right to obtain a more detailed explanation.[201]

The obligation of the deployers to inform persons about the use of AI systems has to be read in light of the *right to explanation*.[202] The right to explanation applies to high-risk systems, which include systems used in the area of migration, asylum and border management. The right to explanation applies to any affected person subject to a decision taken on the basis of the output from a high-risk AI system. The decision has to produce legal effects or similarly significant effects on the person 'in a way that they consider to have an adverse impact on their health, safety or fundamental right.' What kind of information is the person entitled to obtain? The affected person has to obtain from the deployer 'clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.'[203] It needs to be also noted that Article 86 of the AI Act (Right to explanation of individual decision-making) refers to 'the right to obtain', not simply the right to request.[204]

The right to be informed and the right to explanation are closely related to the obligation upon public authorities to provide reasons for their decisions, which is an important procedural safeguard.[205] All of this implies that when public authorities take decisions in the areas indicated in Annex III of the AI Act and these decisions are taken fully or particularly by using AI systems, the

[201] Article 26(11) AI Act refers to Article 13 of Directive (EU) 2016/680 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. This allows Member States to limit, delay or omit such information in order to protect criminal investigations and for the other purposes indicated in Article 13(3) of Directive (EU) 2016/680.

[202] Article 86 AI Act. See Melanie Fink and Michèle Finck, 'Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration' 47 (2022) *European Law Review* 376.

[203] See however the possibilities for limitations allowed by Article 86(2) and (3) of the AI Act. See also Case C-203/22 *CK v Magistrat der Stadt Wien and Dun & Bradstreet Austria GmbH*, 27 February 2025 that can be used as a source of inspiration for interpreting the requirement for providing meaningful explanation.

[204] This 'implies that Article 86 AIA obliges to provide the corresponding explanations "ex officio," and not only when requested by the person exposed to the AI system.' Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 12.

[205] See the right to good administration in Article 41(2)(c) of the Charter of Fundamental Rights of the EU. See Section 6.4 below.

decisions will have to include reasoning and, crucially, this reasoning will have to include clear and meaningful explanation as to the role of the AI systems for the decision.[206] This is the case irrespective of whether the person affected by the decision requests an explanation about the use of the AI system.[207] Overall then, an important implication from the right to be informed and the right to explanation is that AI systems used in the area of asylum and migration by public authorities must be explainable.[208] This excludes the use of systems whose functionalities are not understood.[209]

4.5 Interim conclusion

As a starting point, the AI Act classifies AI systems used in the area of asylum as 'high-risk' system. This means that AI systems that assist public authorities in both the legal aspects as well as the evidence/factual aspects of the asylum claims can be covered by the 'high-risk' classification. This classification is intended to trigger a whole gamut of requirements that such systems need to comply with.

Yet, this starting point (i.e. the classification of AI systems used in the area of asylum as 'high-risk') is extremely uncertain. The reason is that an AI system assisting competent authorities in the examination of asylum claims is classified as a high-risk system under the AI Act *only* if it poses 'a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.' (Article 6(1) AI Act). This can be understood as an exclusion clause since it introduces conditions that exclude the system from the high-risk classification.

[206] A parallel can be drawn here with the CJEU's reasoning in *Ligue des Droits Humains* C-817/19, 27 January 2022, para 210: 'In particular, the competent authorities must ensure that the person concerned – without necessarily allowing that person, during the administrative procedure, to become aware of the pre-determined assessment criteria and programs applying those criteria – *is able to understand how those criteria and those programs work*, so that it is possible for that person to decide with full knowledge of the relevant facts whether or not to exercise his or her right to the judicial redress guaranteed in Article 13(1) of the PNR Directive [Directive 2016/681 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime], in order to call in question, as the case may be, the unlawful and, inter alia, discriminatory nature of the said criteria [emphasis added] [references omitted].'

[207] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 12.

[208] See also Article 13 AI Act – in particular subparagraphs 3(b)(iv) and (vii).

[209] One relevant question here is whether the system can be explainable without access to the source code. This access is regulated by Articles 74(13) and 78 (confidentiality) of the AI Act that limit access to the source code even vis-à-vis market surveillance authorities.

There are three reasons as to why this exclusion is cryptic and thus highly confusing. First, the reference to 'fundamental rights of natural persons' is noteworthy. 'Fundamental rights' are not a source of some clear standards, as Section 6 of this study will elaborate upon. They rather require context-dependent analysis. Second, the threshold of 'significant risk' to fundamental rights introduces further ambiguity. In particular, it raises the question what severity threshold of 'significance' has to be met so that the 'high-risk' classification is applied.

Third, to make things worse, the exclusion clause enshrined in Article 6(1) of the AI Act suggests that there is no significant risk to fundamental rights when the AI system used in the asylum process has *not* materially influenced the outcome of the decision making. The AI Act introduces clarifications as to the conditions under which it can be assumed that the system has not influenced the outcome. One of them is, for example, when the system performs a 'narrow procedural task', which in itself is an ambiguous standard open to various interpretations.

To align things, the assumptions underpinning the AI Act steer the regulatory regime away from regulation and restrictions to the deployment of such systems. The starting point in the AI Act (i.e. the 'high-risk' classification of AI systems used in the area of asylum) is undermined. It would be rather fairer to say that there is very little regulatory restriction originating from the Act given the above-mentioned exclusions clause. This is not counterbalanced with some robust procedural safeguards since the exclusion of an AI system used in the area of asylum from the high-risk classification, is not subjected to a prior administrative authorisation (see Section 4.3.3. above).

What conclusions can be drawn regarding the regulatory requirements imposed by the AI Act if an AI system used in the area of asylum passes the threshold of the 'high-risk' classification?

The system has to be registered in a database; however, since the AI system is used for migration-related purpose, the database is not public. This is a serious limitation.

The high-risk classification triggers certain requirements (1) establishing risk management system; (2) ensuring data governance; (3) keeping technical documentation and records; and (4) maintaining transparency, human oversight, accuracy, cybersecurity, and robustness.[210]

[210] See Articles 9-15, AI Act.

As to the risk management, this study explained that a high-risk system used in the area of asylum can pose *acceptable residual risks*, which does not make it prohibited. Rather such risks need to be only managed, and the AI Act suggests how. Given the use of vague terms such as 'appropriate', the management of such risks is not an object of clear and robust requirements under the terms of the AI Act. The latter openly accepts that management of risk can be costly. In this sense, the assessment of whether risks are acceptable depends on cost-benefit analysis. This type of analysis is ultimately continent on normative (value-related) judgments. It would be problematic if such an analysis would be left to private actors, i.e. the developers of systems.

More importantly, risks might be assessed as acceptable under the terms of the AI Act. This, however, does not necessarily make them acceptable under human rights law, a point elaborated upon in Section 6 of this study.

As to the data governance, this study explained that high-risk systems used in the area of asylum are required to be developed with high-quality data as a matter of principle. In this sense, the training data, the validation data and the testing data have to meet certain qualitative standards. The regulatory bite of the AI Act is however limited give the use of qualifiers such as 'appropriate' and 'to be best extent possible.'

Keeping of technical documentation and records is important for the maintaining of transparency and for ensuring human oversight, accuracy, cybersecurity, and robustness of the high-risk systems. The requirement for record-keeping is however far from robust, which might create obstacles for identification of risks or harms and thus for challenging the use of the AI system, including from a human rights law perspective. The requirement for human oversight is similarly of questionable robustness, given that is it formulated with reference to qualifiers such as 'appropriate and proportionate.'

Finally, it should be kept in mind that the the AI Act does not prevent Member States from imposing additional requirements upon public authorities when the latter use AI.[211] For example, Article 26(3) AI Act confirms this regarding the obligation of human oversight. In fact, such additional more specific requirements might be demanded if specific deployments of AI systems are tested against human rights law standards. This will be the object of enquiry of Section 6. Prior to engaging with human rights law, however, the CoE AI Framework Convention will be explained.

[211] Oriol Mir, 'The Impact of the AI Act on Public Authorities and on Administrative Procedures' (2023) 4 CERIDAP 238, 246.

5. Regulation of AI systems relevant to asylum by the Council of Europe AI Framework Convention

Similarly to the EU, the CoE has very recently adopted a Framework Convention on AI.[212] Similarly to the AI Act, the Framework Convention has been inspired by, one the one hand, 'the unprecedented opportunities' that AI might offer,[213] including by improving the efficiency of administrative procedures,[214] and by the risks that systems might pose to human rights, on the other. In contrast to the AI Act, the CoE AI Framework Convention is not a product safety regime. The CoE AI Convention must be understood in light of the broader mandate of the forum within which it has been accepted. This forum is namely CoE whose mandate is focused on protection of human rights, democracy and rule of law.

Relatedly, the regulatory approach adopted by the CoE AI Convention is different since the level of specificity of its provisions is much lower in comparison with the EU AI Act.[215] The CoE AI Framework Convention is intended to leave 'a considerable margin of discretion for States as to how they implement the broader principles and objectives.'[216]

Similarly to the path followed in Section 4 where the EU AI Act was addressed, this section will explain the obligations imposed by the CoE AI Convention by taking the following steps. Section 5.1 presents the definition of AI systems adopted by the CoE AI Convention. Section 5.2. explains how the CoE AI

[212] At the time of writing the Convention has not entered into force. See Full list - Treaty Office

[213] See the Preamble.

[214] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 20.

[215] This also explains why it is a framework convention and not a convention. The difference between the two is that 'a convention regulates a specific matter or area in a more concrete way, typically by creating rights and obligations, whereas a framework convention rather sets out broader principles and areas for action which have been agreed between States Parties.' CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 134.

[216] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 135.

Convention applies to public authorities, including authorities engaged in asylum procedures. Section 5.3. clarifies the risk-based approach adopted by the treaty which shapes the content of the obligations imposed upon States. The content of these obligations is finally explained in more detail in Section 5.4.

5.1 Definition of an AI system

CoE AI Framework Convention defines AI systems in the following way:

> machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment.

Despite the different formulation, this definition closely mirrors the definition in the EU AI Act.

5.2 The CoE AI Convention applies to public authorities

As other CoE conventions, the CoE AI Framework Convention is meant to regulate the conduct of States, not of private actors. The CoE AI Convention therefore applies to its State Parties, i.e. to public authorities. It also applies to private actors acting on the behalf of public authorities. Public authorities might delegate their responsibilities to private actors or direct them to act. This can happen when private actors operate pursuant to a contract with a public authority or when private actors provide public services.[217]

Yet, private actors are regulated indirectly 'by virtue of the rights granted to, and obligations assumed by states' under the CoE AI Convention.[218] The latter instrument therefore can indirectly address private actors, which is very similar to how the ECHR can indirectly regulate private actors. This indirect regulation is achieved in the following way: 'Each Party shall address risks and impacts arising from activities within the lifecycle of AI systems by private actors […].'[219] Risks are to be addressed by the adoption of legislative, administrative or other

[217] Explanatory Report, para 28.

[218] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 91.

[219] Article 3(1)(b) Framework Convention.

regulatory measures.[220] Overall then it can be concluded that the CoE AI Convention imposes positive obligations upon States to regulate private actors. As clarified in Section 6, such positive obligations in turn can be a source for inspiration for the development of positive obligations under the ECHR, an instrument equipped with an adjudicative body (i.e. the ECtHR) to review individual complaints. In contrast, the CoE AI Convention is not equipped with such a dispute settlement judicial mechanism.

5.3 The risk-based approach

Article 1(2) of the CoE AI Convention stipulates that measures meant to regulate AI systems 'shall be graduated and differentiated as may be necessary in view of the severity and probability of the occurrence of adverse impacts on human rights, democracy and the rule of law throughout the lifecycle of artificial intelligence systems.' Therefore, similarly to the EU AI Act, the CoE AI Convention has a risk-based approach: the measures 'need to be tailored to the level of risk posed by an artificial intelligence system within the specific spheres, activities and contexts.'[221] The State Parties have discretion how to balance any competing interests in each sphere.[222] The Explanatory Report does note that the public sector sphere, including border control and asylum, might have some specificities that need to be taken into account.[223] It also highlights that in certain spheres, such as asylum and immigration procedures, there are 'power asymmetries' that need to be taken into consideration.[224]

The Explanatory Report to the treaty is useful in better understanding what is meant by the 'lifecycle' of the system. In particular, this lifecycle includes (1) planning and design, (2) data collection and processing, (3) development of artificial intelligence systems, including model building and/or fine-tuning existing models for specific tasks, (4) testing, verification and validation, (5) supply/making the systems available for use, (6) deployment, (7) operation and monitoring, and (8) retirement.[225]

Contrary to the EU AI Act, the CoE AI Convention does not create different levels of risks. It does not therefore differentiate AI systems based on the risk that they might pose. This also implies that the obligations imposed are not tailored with reference to the level of risk a system might be accepted to pose.

[220] Explanatory Report, para 29.

[221] Explanatory Report, para 17.

[222] Explanatory Report, para 17.

[223] Explanatory Report, para 17.

[224] Explanatory Report, para 18.

[225] Explanatory Report, para 15.

Neither is the regulation by the treaty tailored to specific sectors (such as asylum, immigration and border control). Rather Chapter II of the Convention imposes general obligations that apply to all AI systems that fit within the Article 2 definition.[226] As already clarified, the addressees of these obligations are States.

5.4 The obligations imposed upon States

Starting point and gaps meant to be addressed

To better understand the obligations imposed upon States by the CoE AI Convention, it is pertinent to underscore its starting point. It does not aim to establish new human rights law obligations. This is made very clear in its Explanatory Report:

> no provision of this Framework Convention is intended to *create new human rights or human rights obligations* or undermine the scope and content of the existing applicable protections, but rather, by setting out various legally binding obligations contained in its Chapters II to VI, *to facilitate* the effective implementation of the applicable human rights obligations of each Party in the context of the new challenges raised by artificial intelligence.[227]

The text of Article 4 of the CoE AI Convention reflects this starting point. This provision imposes an obligation upon the State Parties to 'adopt and maintain measures to ensure that the activities within the lifecycle of AI systems are consistent with obligations to protect human rights, as enshrined in application international law and its domestic law.' This text does not add much specificity. It does not add much to the uncertainty regarding the answer to the question how human rights law limits and regulates AI systems.

To better understand the role of the CoE AI Convention and its possible contribution, it is useful to go back to the Feasibility Study prepared prior to the adoption of the Convention. The Study highlights four 'legal gaps': (1) specification/concretisation of obligations; (2) identification of concrete principles specifically relevant to the AI systems; (3) responding to systemic harms; and (4) responding to cross-border/transboundary nature of the harm.

[226] See the exclusion of AI systems related to the protection of States' national security interests and national defence, as stipulated in Article 3(2) and (4) CoE AI Convention. 'National security' is distinguished from 'public security'. See Explanatory report, para 32.

[227] Explanatory report, para 13 (emphasis added).

Table 6. Legal gaps meant to be addressed by the CoE AI Framework Convention

Legal gaps meant to be addressed by the CoE AI Framework Convention		
1. concretisation of obligations	-	Article 4 of the Convention – concretisation not achieved
2. identification of concrete principles specifically relevant to the AI systems	human control and oversight technical robustness transparency, explainability and traceability competence	Article 7 (human dignity and individual autonomy) Article 8 (transparency and oversight) Article 9 (accountability and responsibility) Article 10 (equality and non-discrimination) Article 11 (privacy and personal data protection) Article 12 (reliability) Article 13 (safe innovation)
3. responding to systemic harms	protection of electoral processes and democratic institutions	See Article 5 of the Convention.
4. responding to cross-border/transboundary nature of the harm	-	See Article 25 (international cooperation)

As to the *first* legal gap, the Feasibility Study notes that 'the rights and obligations formulated in existing legal instruments tend to be articulated broadly or generally, which is not problematic as such, yet can in some instances raise interpretation difficulties in the context of AI.'[228] Concretisation of the obligations corresponding to the rights is therefore proposed. The *second* 'legal gap' is a manifestation of the first one since it is about explicitly legally enshrining certain principles that are specifically relevant to the operation of AI systems. These include human control and oversight, technical robustness, transparency, and explainability.[229] Traceability via record-keeping

[228] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 84.

[229] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 85.

and documentation of logs is also added,[230] which can enable the identification any causal links between harms and the operation of the AI systems. This in turn can enable the possibility of challenging the operation of the AI system under human rights law and the assessment whether there is any human rights law violation. Another principle that the Feasibility Study indicates as being specifically relevant to the operation of AI systems is ensuing that developers and deployers of AI systems have the necessary competence. The Feasibility Study observes that if these principles are not specifically regulated, this leads to uncertainty for both developers and deploys, which might hamper innovation.

The *third* legal gap identified in the Feasibility Study and meant to be addressed by the CoE AI Convention, concerns the wider social impact from the deployment of AI systems.[231] This impact transcends human rights law that is focused on the individual and on more specific harms. This impact relates to broader questions about the operation of liberal democracies, including for example how AI systems might interfere with electoral processes and democratic institutions. The *fourth* legal gap identified in the Feasibility Study concerns the 'lack of common norms at international level,' which is an obstacle for the trade of AI systems and for mutual trust.[232]

Protection of human rights

It is the *first* and the *second* gap that will be in focus. As to the first one that concerns concretisation of the obligations corresponding to the rights, it was already suggested in the beginning of this section that the CoE AI Framework Convention does not really achieve such a concretisation. Article 4 of the treaty simply refers to the obligation upon States to protect human rights.

Instead of clarifying, the Explanatory Report to the Convention actually confuses as to the nature of Article 4 of the Convention. It states that

> Parties are free to choose the ways and means of implementing their international legal obligations, provided that the result is in conformity with those obligations. *This is an obligation of result and not an obligation of means.* In this respect, the principle of subsidiarity is essential, putting upon the Parties the primary responsibility to ensure respect for human rights and to provide redress for violations of human rights.[233]

[230] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 85.

[231] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 86.

[232] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 88.

[233] Explanatory Report, para 38.

If it is prevention of harm or risk of harm that is the result meant to be achieved, the obligations cannot be ones of result. Harm and risk of harm materialize all the time in relationship to various activities, and this harm and risk by themselves are never the *sole* basis for finding breach of obligations under human rights law.[234] If it is regulation of activities that is the result meant to be achieved, then indeed the obligation to regulate as such can be conceptualized as an obligation of result. Overall then, not only does Article 4 fail to concretize, but the clarifications added to it by the Explanatory Report are confusing.

Principles related specifically to the activities of AI systems

The second gap meant to be addressed by the CoE AI Convention is explicit legal enshrinement of certain principles that are specifically relevant to the operation of AI systems. This is indeed achieved by Chapter III of the treaty that enumerates principles 'related to activities within the lifecycle of artificial intelligence systems'. These are, however, 'purposefully drafted at a high level of generality' and are meant to have 'very broad application to a diverse range of circumstances.'[235] As the Explanatory Report of the CoE AI Convention suggest, the drafters of the treaty assumed that the 'detailed legal regime of human rights protection with its own set of rules'[236] is sufficient to further tailor the application of these general principles.

As Section 6 in this study will show, it is questionable whether the human rights regime can provide any detail; this regime is rather yet to develop so that it can address the new challenges and related harms to important interests posed by the use of AI systems. As Section 6 will also demonstrate, the human rights regime does have the potential to respond to the challenges, but it needs to be developed. In this development, the principles originating from the regulatory framework established specifically for AI systems is an important source of guidance.

Being a source of guidance, the regulatory principles enumerated in Chapter III of the Convention need to be understood. This is the aim pursued in the following sections that consequently address the following principles: human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination and data protection, reliability and finally, risk management.

[234] V. Stoyanova, 'Framing Positive Obligations under the European Convention on Human Rights Law: Mediating between the Abstract and the Concrete' (2023) *Human Rights Law Review*.

[235] Explanatory Report, para 49.

[236] Explanatory Report, para 51.

Human dignity and individual autonomy

Article 7 of the CoE AI Convention enshrines the principle of respect for human dignity and individual autonomy. The Feasibility Study clarifies that '[t]o safeguard human dignity, it is essential that human beings are aware of the fact that they are interacting with an AI system and not misled in this regard.' The Study adds that 'they should in principle be able to choose not to interact with it, and to not be subject to a decision informed or made by an AI system whenever this can significantly impact their lives, especially when this can violate rights related to their human dignity.'[237]

The Explanatory Report, however, does not go that far. It clarifies among other things that respect to human dignity implies that individuals are not reduced to 'mere data points'.[238] According to the Explanatory Report therefore, individual autonomy does imply an ability to make choices and decisions and that individuals have 'control over the use and impact of artificial intelligence technologies in the live.'[239] The possibility to refuse to be subjected to a decision-making process by a public authority that might use an AI system, is however *not* mentioned as part of the conceptualisation of 'individual autonomy' in the Explanatory Report to the treaty.

Transparency and oversight

The second principle enshrined in the CoE AI Convention is transparency and oversight. In particular, Article 8 of the treaty stipulates that

> Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems.

[237] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 99.

[238] Explanatory Report, para 53.

[239] Explanatory Report, para 55.

It is notable that explainability is not mentioned in the text of the treaty.[240] The Explanatory Report suggest that explainability is part of the principle of transparency. It clarifies that explainability

> [] refers to the capacity to provide, subject to technical feasibility and taking into account the generally acknowledged state of the art, sufficiently understandable explanations about why an artificial intelligence system provides information, produces predictions, content, recommendations or decisions, which is particularly crucial in sensitive domains such as healthcare, finance, *immigration, border services* and criminal justice, where understanding the reasoning behind decisions produced or assisted by an artificial intelligence system is essential. In such cases transparency could, for instance, take the form of a list of factors which the artificial intelligence system takes into consideration when informing or making a decision.[241]

Accountability/responsibility and remedies

Transparency can enable the identification of harm and the identification of actors that might have caused this harm, which is turn is indispensable for ensuring responsibility.[242] The CoE AI Convention thus specifically addresses accountability and responsibility. In particular, Article 9 of the treaty is generally framed in the following way:

> Each Party shall adopt or maintain measures to ensure accountability and responsibility for adverse impacts on human rights, democracy and the rule of law resulting from activities within the lifecycle of artificial intelligence systems.

[240] Yet, see para 108 from the CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020: 'they [those affected by the decision] should receive an explanation of how decisions that impact them are reached. While an explanation as to why a system has generated a particular output is not always possible, in such a case, the system's auditability should be ensured. While business secrets and intellectual property rights must be respected, they must be balanced against other legitimate interests. Public authorities must be able to audit AI systems when there is sound indication of non-compliance to verify compliance with existing legislation. Technical burdens of transparency and explainability must not unreasonably restrict market opportunities, especially where risks to human rights, democracy and rule of law are less prominent. A risk-based approach should hence be taken, and an appropriate balance should be found to prevent or minimise the risk of entrenching the biggest market players and / or crowding out and, in so doing, decreasing innovative socially beneficial research and product development.'

[241] Explanatory Report, para 60 (emphasis added).

[242] Explanatory Report, para 69.

This might demand the adoption of new legal frameworks at national level or adaptation of the existing judicial, administrative, civil or other national liability regimes.[243] Such responsibility regimes are inherently linked with the right to effective remedy enshrined in Article 14 of the CoE AI Convention. Similarly to Article 8 of the treaty (transparency and oversight), Article 14 requires measures for documenting relevant information regarding AI system.[244]

There are qualifiers, however: such documentation is required when the system has 'the potential to *significantly* affect human rights.'[245] One can wonder what 'significantly' means in this context and who decides whether the harm is significant or not. Addition qualifier comes with Article 14(2)(b) of the CoE AI Convention: States have to adopt measures to ensure that the information about the AI system is '*sufficient* for the affected persons to contest the decision(s) made or *substantially* informed by the use of the system, *and where relevant and appropriate,* the use of the system itself.' The following questions are left open: Who decides when the information is *sufficient*? and Who decides whether the decision is *substantially* informed by the use of the system? Finally, no clear and straightforward obligation is imposed to inform about the use of the system itself. The commentary on Article 14 offered by the Explanatory Report seem to justify all these qualifications in the following way:

> It is also important to recall that exceptions, limitations or derogations from such transparency obligations are possible in the interest of public order, security and other important public interests as provided for by applicable international human rights instruments and, where necessary, to meet these objectives.[246]

Article 15 of the treaty strengthens the right to effective remedies by adding procedural safeguards. In particular, Article 15(1) of the treaty stipulates that 'where an artificial intelligence system *significantly impacts upon the enjoyment of human rights,* effective procedural guarantees, safeguards and rights, in accordance with the applicable international and domestic law' shall be ensured.

Article 15(2) of the Convention has an additional weakness. Its texts says that 'Each Party shall *seek to ensure* that, *as appropriate for the context*, persons interacting with artificial intelligence systems are notified that they are interacting with such systems rather than with a human.' The requirement to

[243] Explanatory Report, para 66.

[244] Article 14(2), CoE AI Convention.

[245] Article 14(2) (a), CoE AI Convention.

[246] Explanatory Report, para 99.

'seek to ensure' is weaker than 'to ensure'. The robustness of the requirements for notification is further undermined by the qualification 'as appropriate for the context'. It opens the possibility for an argument that notification in the context of, for example, refugee status determination procedure, is not appropriate. Without notification, the affected person cannot know that an AI system has been used, which can undermine the possibility for subjecting the system to any security including via legal channels for establishing responsibility.

Non-discrimination and data protection

Article 10 of the CoE AI Convention imposes an obligation of adopting measures for ensuring non-discrimination, 'as provided under applicable international and domestic law.' Article 10 should be read together with Article 17 of the treaty that stipulates that '[t]he implementation of the provisions of this Convention by the Parties shall be secured without discrimination on any ground, in accordance with their international human rights obligations.' These non-discrimination provisions do not add any specificities.[247] Article 11 of the treaty that is about privacy and personal data protection can be characterized in the same way.

Reliability

Article 12 of the CoE AI Convention imposes an obligation upon States to take measures 'as appropriate' to promote reliability.[248] These measures are meant to ensure 'adequate quality and security' of the AI systems. The Explanatory Report suggests that these measures include adoption of standards, technical specifications, assurance techniques and compliance schemes. The Explanatory Report adds that Article 12 of the treaty is 'based on the assumption that the Parties are best placed to make relevant regulatory choices.'[249]

Risk management

Article 16 of the CoE AI Convention stipulates that the State Parties shall adopt measures for the 'identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and

[247] It is hard to understand the implications from para 77 of the Explanatory Report that claims that Article 10 of the CoE AI Convention demands 'overcoming structural and historical inequalities.'

[248] It is notable that the text of Article 12 of the CoE AI Convention does not say that the State Parties take measures to *ensure* reliability, but only to *promote* reliability. This implies that the conduct by States will be assessed in light of their efforts, not with reference to the actual result (i.e. actual reliability).

[249] Explanatory report, para 84-89.

potential impacts to human rights, democracy and the rule of law.'[250] Such measures shall be 'graduated and differentiated'. This means that States have a wide scope of flexibility as to what measures to take. States also flexibility as to what measures should be taken depending on the areas where AI systems are deployed.[251] Article 16(2) of the treaty adds some details by indicating what such measures should take account of; it does not however undermine the above-mentioned flexibility. The State Parties also have wide flexibility as to how to respond to the risks that might be identified.

5.5 Interim conclusion

The provisions of the Council of Europe AI Convention are marked by a high degree of flexibility and abstraction, granting significant discretion to its State Parties. Notably, the Convention must be applied without prejudice to existing human rights law,[252] which – as Section 6 will suggest – may evolve to articulate and strengthen standards in future concrete cases. The treaty does not create new human rights or obligations; rather, its stated aim is to support the effective implementation of existing human rights commitments. As a result, the Convention establishes a relatively weak regulatory framework. It is also worth noting that it makes no explicit reference to AI systems used in the context of asylum.

In contrast, the EU AI Act explicitly regulates AI systems used in the context of asylum as 'high risk' systems. The high level of abstraction for formulating the provisions of the CoE AI Convention can be also linked to the choice of not incorporating concrete risk levels and tailoring regulations to the levels. In this respect, the Convention differs from the EU AI Act. Yet, given the ambiguities in the formulations of the risk levels and the possibilities for exclusions in the AI Act, the more general approach in the Convention does not necessary imply less robust regulation. Neither does it necessarily imply that the more abstractly formulated provisions in the Convention might not be possible to invoke for regulating concrete technologies of AI supported decision-making used in the context of asylum.

[250] Democracy and rule of law are not addressed here in this study.

[251] Explanatory report, para 106.

[252] See Article 3(1)(b) of the CoE AI Convention: 'When implementing the obligation under this subparagraph, a Party may not derogate from or limit the application of its international obligations undertaken to protect human rights, democracy and the rule of law'.

6. Human rights law

The study so far has explained the regulation of AI systems introduced by two regulatory frameworks (the EU AI Act and the CoE AI Convention). This final section turns to human rights law to directly respond to the research questions posed in the beginning of the study. These questions were framed as an inquiry about compliance with human rights law, and more specifically compliance with the European Convention on Human Rights.

The address the compliance question, it is first important to reflect upon the different regulatory approaches adopted by human rights law and by the AI systems regulations described in the previous two sections. Section 6.1. does this. Section 6.2 then explains the interaction of the two regulatory frameworks at a more structural level. This interaction goes in both directions. First, human rights law is directly and explicitly referred to in the AI regulations. Second, at the same time, if human rights law were to be invoked for challenging the application of AI systems and imposing any limits (i.e. regulatory limits) on such applications, the AI regulations adopted at the EU and CoE levels would be key for this challenge.

While the AI regulatory frameworks analysed in Sections 4 and 5 contain multiple references to human rights, harm to such rights, or risks to such rights, Section 6.3 underscores that if human rights law were to be invoked for challenging the application of AI systems *the specific harm needs to be identified*. General references to human rights are not useful. Section 6.3 formulates an argument that the use of AI systems might be possible to conceptualize as harming procedural interests protected by human rights law. In this sense, the harm can be perceived as *a procedural harm.* This conceptualization of the harm is of key importance for the framing of any obligations that States might have under human right law. More specifically, the State is under procedural positive obligations to regulate. These obligations are explained in Section 6.4.

The right not to be subjected to refoulement can be also understood as triggering negative obligations under human rights law, since the State might be expected to refrain from removing the person. Any risks posed by the use of AI systems in the asylum determination procedure, can be therefore conceptualised as limitations upon important interests protected by Article 3 ECHR. Such usages can be also conceptualized as limitations upon the right to private life protected by Article 8 ECHR, if AI systems use personal data.

Article 3 ECHR is not a qualified right and therefore not subject to a proportionality/balancing assessment: the Court does not engage with an analysis whether the risk of harm upon removal is proportionate to the State legitimate interests (i.e. immigration control, saving of public funds via efficient procedures).[253] Yet, the case law of the ECtHR is clear to the effect that compliance with the negative obligation not to *refoule*, while not subject to the above mentioned balancing, it is still qualified. The reason is that compliance is assessed in light of certain procedural standards, which makes the analysis offered in Section 6.4 relevant.

The right to private life, however, is a qualified right under the terms of Article 8(2) ECHR. Any measures of interference with private life via the involvement of an AI system in the asylum procedure, have to meet certain requirements to be deemed compatible with human rights law. These requirements are legality, suitability and proportionality, which are examined in Section 6.5.

6.1 Different regulatory approaches

It is important to begin by emphasising the distinct regulatory approach of human rights law, particularly when compared to the AI regulations discussed in Sections 4 and 5. Human rights treaties, such as the European Convention on Human Rights (ECHR), articulate abstract rights grounded in fundamental human interests – such as the right to life, respect for private and family life, and protection from torture and other forms of ill-treatment. These treaties reflect values whose significance is rarely contested. In this sense, there is broad consensus on the importance of protecting human life, safeguarding individuals from inhuman or degrading treatment, and ensuring a private life that encompasses personal data protection.

At the same time, there is broad agreement that these fundamental interests are not absolute and must often be balanced against competing interests,[254]

[253] V Stoyanova, 'How Exceptional Must 'Very Exceptional' Be? Non-Refoulement, Socio-Economic Deprivation and Paposhvili v. Belgium' 29(4) *International Journal of Refugee Law* (2017) 580.

[254] Even the right not to be subjected to *refoulement*, is difficult to perceive as absolute if its corresponding obligations are conceptualized as positive obligations upon the State to allow a person to remain or as a positive obligation to conduct risk assessment. See V. Stoyanova 'How Exceptional Must 'Very Exceptional' Be? Non-Refoulement, Socio-Economic Deprivation and Paposhvili v. Belgium' 29(4) *International Journal of Refugee Law* (2017) 580.

whether public or individual.[255] In the context of migration, this balancing process has particular features. A core premise in human rights reasoning is that States have the sovereign authority to control their borders and to decide who may enter and remain within their territory.[256] It is within this balancing exercise – typically conducted through a proportionality assessment – that the specific obligations of States, corresponding to individual rights, are defined and clarified.

Notably, human rights treaties do not contain and do not specify obligations. For instance, Article 1 of the ECHR stipulates that 'The High Contracting Parties shall secure to everyone within their jurisdiction the rights and freedoms defined in Section I of this Convention.' The obligation to secure the rights and freedoms is formulated in very abstract terms. It does not tell us anything about how to balance fundamental individual interests and general interests; accordingly, the ECHR does not formulate obligations at a more concrete level. The measures that the State is prohibited to undertake (i.e. the State's negative obligations) since they disproportionately harm important interests, are not specified. Nor are the measures that the State is required to take under its positive obligations - measures aimed at preventing or addressing harm.

At this juncture, *the first key distinctive feature of human rights law emerges: its high level of abstractness.* This puts it in sharp contrast to the regulatory approach of the AI Act, where the focus is on the formulation of some more specific requirements. As explained in detailed in Section 4 above, the AI Act has a regulatory approach featuring technical specific detailed regulations. In the AI Act standards play a key role. When the actors that are regulated (providers and deployers of AI systems) comply with the regulatory standards and when they follow the requirements indicated in the AI Act, conformity is assumed.[257] Put is differently, there is a presumption of conformity, if the specific requirements are followed.

As also explained in Section 4 above, the regulations as formulated in the AI Act are framed with reference to certain adjectives such as 'appropriate', which creates uncertainty and, relatedly, increases the level of abstraction of

[255] See, for example, the text of Article 8(2) ECHR that formulates the following interests that complete with the interests of protecting private life: 'national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.'

[256] It is a 'well-established principle of public international law, recognised by the Court, according to which the States Parties, subject to their treaty obligations, including the Convention, have the right to control the entry, residence and expulsion of aliens.' *M.N. and Others v Belgium* [GC] App no 3599/18, decision 5 March 2020, para 124.

[257] See Articles 40 and 41(3) AI Act.

the requirements. Yet, the flexibility created with these references is incomparable with the high level of abstraction that generally characterises human rights law.

Although human rights law is formulated at a high level of abstraction, it must still provide tools for regulating conduct in order to be meaningful. The starting point is that it regulates the conduct of the State – though, as will be discussed below, this can be nuanced. For now, the focus is on the regulatory approach itself. The key question is: how can human rights law regulate conduct when the rights it protects – grounded in fundamental interests – are not accompanied by clearly defined obligations in the text of the European Convention on Human Rights (ECHR)? Crucially, it is only through identifying and specifying the corresponding obligations that we can understand what human rights require and evaluate whether new technologies comply with those requirements. Put simply, we need a clear sense of the obligations attached to rights before we can assess compliance or identify potential violations.

The obligations corresponding to the rights enshrined in the ECHR have been generally conceptual as positive obligations and negative obligations. Despite the limitations of this distinction, the ECtHR refers to it and uses it to adjudicate cases.[258] Negative obligations demand from the State to reframe from certain conduct. In this sense, they are *prohibitions*. The review of whether a State breached a negative obligation follows a particular structure that includes *inter alia* an assessment of whether the measure that interfered with the individual interest protected by the human right, pursues legitimate aims. These aims reflect general interests that have been formulated in the text of the Convention as national security, public safety, the economic well-being of the country, prevention of disorder or crime, protection of health or morals.[259] Protection of national borders is a legitimate interest. Fostering innovation, competitiveness and effectiveness via the development and deployment of AI systems, can be subsumed within the legitimate objective of economic well-being of the country.

[258] L. Lavrysen, *Human Rights in a Positive State* (Intersentia 2016); A Mowbray, *The Development of Positive Obligations under the ECHR* (Hart Publishing 2004); V. Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (OUP 2023).

[259] Article 8(2), ECHR. In the text below, the terms 'general interest', 'legitimate aim', 'general legitimate aim' are used interchangeably.

If the interference measure does pursue one of these general aims, this is a manifestation of how an important individual interest as protected by the ECHR can be in tension with general legitimate interests.[260] Even if the measure pursues legitimate aims, an assessment needs to be made whether the measure is suitable and necessary in a democratic society. The suitability test implies asking the question whether the measure contributes to the achievement of the legitimate general aims. This is normally easily accepted in the human rights law review.[261] In contrast, the necessity test introduces a highly complex analysis in the review. The necessity test can be understood as an entry point for a proportionality assessment, which will be better explained in Section 6.5 below.

What is most relevant at this juncture is that the proportionality assessment brings us to one of the most important regulatory features of human rights law: human rights law demands *context dependent assessment* whether interference measures that harm fundamental interests *are disproportionate and thus prohibited*. Any measure of interference assessed as disproportionate would be in violation of human rights law and thus prohibited under human rights law. The proportionality assessment is context dependent since the assessment is centred on the specific individual circumstances of the particular right holder.

In addition to being context dependent, a feature that as mentioned before necessarily follows from the proportionality review, the regulatory approach of human rights law is underpinned by another key feature. Namely, since it is the individual at the centre of the review, the assessment whether human rights law prohibits certain measures (vie the imposition of negative obligations) or demands regulation (via the imposition of positive obligations) is *highly*

[260] In addition to these general interests, as the text of Article 8 of the Convention suggests, 'the protection of the rights and freedoms of others' can be also a legitimate aim for infringing rights. In this situation, tension might arise between individual interests. If both of these individual interests ground rights enshrined in the ECHR, there might be a tension between human rights. For the sake of simplicity, I ignore this scenario.

[261] See A. McHarg, 'Reconciling Human Rights and the Public Interest: Conceptual problems and Doctrinal Uncertainty in the Jurisprudence of the European Court of Human Rights' (1999) 62 *Modern Law Review* 671. See also See B. Cali, 'Balancing Human Rights? Methodological Problems with Weights, Scales and Proportions' (2007) 29 *Human Rights Quarterly* 251, 265.

individualized.[262] To put it plainly, what might be proportionate in some contexts in relationship to some individuals, might be disproportionate in other contexts. As I will further clarify below, this indeterminacy, individual-centrality and context-dependency of human rights law is both its weakness and its strength.

The same individualized and context-dependent regulatory approach characterises positive human rights obligations. In fact, *positive obligations are context-dependent to a higher degree*.[263] This is very important since it is exactly positive obligations that can be invoked as obligations upon the State to regulate the use and the application of AI systems. This so if it is private actors that develop and deploy the systems. This is also likely to be the case even if state authorities, such as immigration authorities, develop and deploy the systems or only deploy systems developed by private actors. The reason is that there is no general prohibition on the use of the systems. This statement needs to be qualified since if AI systems fall within the scope of Article 5 of the AI Act, they are prohibited, and their use will most likely not comply with the legality requirement under human rights law. This might make them prohibited under human rights law also as a matter of negative obligations.[264]

In the absence of a general prohibition, the central question from a human rights law perspective is whether – and how – AI systems should be regulated as a matter of positive obligations. These obligations may, in fact, require States to intervene through regulatory measures. Such regulation must address not only private actors involved in the development or deployment of AI systems, but also public authorities engaged in similar activities. Viewed in this light,

[262] It would then follow that human rights law is difficult to invoke for addressing collective harm. Scholars have made the case for expanding the notion of harm so that it is not limited to individualistic notion. See J. Niklas, 'Human Rights-Based Approach to AI and Algorithms Concerning Welfare Technologies' in W. Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (2021) 517, 524; B Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 *Philosophy and Technology* 478; L. Taylor, L. Floridi and B. van der Sloot, 'Introduction: A New Perspective on Privacy' in L. Taylor, L. Floridi and B van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (2017) 4. Yet, this undermines some basic foundations of human rights law. From a procedural perspective, claims cannot be adjudicated if concrete individuals do not have a victim status by not being directly affected. Very similar issue has arisen in the efforts to litigate climate change by using human rights law. See G. Letsas, 'The European Court's Legitimacy after *KlimaSeniorinnen'* (2024) 5(4) European Convention on *Human Rights Law Review* 444.

[263] V Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (OUP 2023).

[264] Given the wording of Article 5 of the AI Act, the conclusion that an AI system is prohibited is far from straightforward. Legislators might struggle to regulate rapid technological developments. See *S. and Marper v the United Kingdom* App no 30562/04 and 30566/04, 4 December 2008, for the legality requirement.

both the AI Act and the Council of Europe's AI Framework Convention can be understood as efforts by States to fulfil their positive obligations under human rights law.

Any positive obligation under human rights law to regulate is, however, very abstract. This obligation needs to be specified. As the practice and the reasoning of the ECtHR shows, this specification is achieved in light of the context of the specific case and with reference to the standards of knowledge, causation and reasonableness.[265] To break it down, for a State to have a positive human right obligation to regulate, the following needs to be demonstrated. First, it needs to be shown that the State knew or ought to have known about harm or risk of harm to the fundamental interests. Second, it needs to be shown that the omission to regulate caused the harm or the risk of harm. Third, even if the State had actual or putative knowledge about the harm or the risk of harm and even if the regulation would have prevented the harm, any regulatory interventions by the State are requires to be only reasonable.

The key point here is that the standard of reasonableness implies that regulatory interventions by the State as measures commanded by positive obligations in human rights law, are context specific and context dependent. In this way, the assessment of breach of positive obligations as contingent on the standard of reasonableness is similar to the assessment of breach of negative obligations as contingent on the proportionality test.

This results in ambiguity regarding the precise commands imposed by positive human rights obligations. This ambiguity, along with their context-dependent nature, is notably greater when compared to negative obligations. The reason is that States have discretion how to comply with positive obligations. These obligations *do not impose concrete commands* on States how to achieve better protection of fundamental interests. States can make *regulatory choices*. Positive obligations in human rights law fully accommodate this regulatory discretion as long as the choice made is effective.[266] However, the standard of effectiveness is very abstract, which again brings us back to the fundamental regulatory feature of human rights law – its high level of abstraction.

The different regulatory approaches can be also explained in the following way: while human rights law has a public/administrative/constitutional law approach, the AI Act has a product safety approach. As mentioned in Section 4, the AI Act regulates AI systems from the perspective of product safety.

[265] V. Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (OUP 2023).

[266] V. Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (OUP 2023).

Irrespective of who develops or uses the systems (private or public actors), the systems are viewed as products that pose risks, including risks to fundamental interests protected by human rights.[267] This approach explains why the AI Act follows the principles of self-regulation, presumption of conformity, self-control, *ex-post* supervision for compliance with the requirements by national market surveillance authorities.[268] The product safety approach also explains why the AI Act is predominately directed at the providers.[269]

Public authorities can be also providers and, in this sense, are also regulated. However, the AI Act does not have specific section dedicated to public authorities. It does not directly address certain problematic issues that can be identified from a public/administrative/constitutional law perspective.[270] Such issues include inter alia (i) the legal basis when public authorities use AI systems, (ii) how the usage of the systems relates to administrative discretion that public authorities have, (iii) the challenges to provide reasoned decisions when AI systems are used, or access to the codes by the public in general.[271] In addition to these three, another issue that arises when public authorities use systems concerns public trust and legitimacy.[272]

[267] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 3. Concerns have been expressed as to how AI-related risks, which concern ethical and fundamental rights issues, are addressed through harmonisation techniques which were developed to address health and safety concern. See Sybe de Vries et al, 'Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation' in Gavin Robinson (et al, eds) 'Future-proof Regulation and Enforcement for the Digitalised Age' (2023) 8 *European Papers, Journal of Law and Integration* 3.

[268] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk* Regulation 1, 3.

[269] Lena Enqvist, "Human Oversight' in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15(2) *Law, Innovation and Technology* 508, 523.

[270] Francesca Palmiotto 'The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation' *European Journal of Risk Regulation* (2025) 1, 8: 'When AI is used in decision-making processes, individuals are not consumers of AI products but are subject to them. AI is not a dishwasher but a technology that poses risks to fundamental rights and democratic values, which transcend market regulation and consumer law.'

[271] Oriol Mir, 'The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?' (2024) *European Journal of Risk Regulation* 1, 2. See Simona Demková, *Automated Decision-Making and Effective Remedies* (Elgar 2023); Herwig C H Hofmann and Felix Pflücke (eds), *Governance of Automated Decision Making and EU Law* (OUP 2024).

[272] Johan Laus, Sandra Watcher and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the EU AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 *Regulation and Governance* 3.

In summary, the distinctive features of human rights law can be visually represented in the following table:

Table 7. Distinctive features of human rights law

Distinctive features of human rights law
High level of abstractness
Context-dependent assessment shaped by the principles of proportionality and reasonableness
Focus on the individual
No specification of obligations (positive or negative obligations)

The key differences between the regulatory approaches of human rights law and the AI legal frameworks discussed in Sections 4 and 5 are illustrated in the table below:

Table 8. Comparative Overview – Human Rights Law vs. AI Regulation Approaches

Key differences between the regulatory approaches of human rights law and the regulation of AI system	
Human rights law	Regulation of AI system
Regulation of State conduct	Regulation of both private and public actors
High level of abstraction	More concrete standards and regulation inspired by product safety approach
Post-factum assessment	Ex-ante guarantees that create presumption of compliance
Human rights law	Regulation of AI system

6.2 Intertwinement of the regulatory frameworks

Despite their differing regulatory approaches, human rights law and the EU and Council of Europe frameworks for AI regulation are closely interconnected. This interrelationship unfolds along two related axes. First, human rights law is explicitly referenced and relied upon in the AI regulatory instruments adopted by both the EU and the CoE (Section 6.2.1). Second, these AI regulations can, in turn, inform the development and interpretation of human rights law (Section 6.2.2).

Human rights law invoked in the AI regulations

First, human rights law is invoked multiple times in both the AI Act and the Council of Europe's AI Convention. However, two important clarifications are necessary. As discussed in Section 6.1 above, it would be incorrect to assume that these references automatically translate into concrete regulatory standards.

Instead, they should be understood as acknowledgments of the fundamental individual interests that may be at risk of harm. What specific measures are required to prevent such harm - or mitigate the associated risks - is a question to which human rights law does not currently offer clear or immediate answers. In fact, it may not yet offer answers at all.[273] Human rights law is still in the process of evolving into a framework capable of addressing these emerging challenges.

Second, compliance with the requirements of the AI Act does not imply compliance with human rights law. Compliance with the requirements might indeed shape the human rights law reasoning (including the assessment of proportionality and reasonableness) by way of creating certain starting points (i.e. presumptions) in the reasoning. Yet, given its individual-centred nature and context-dependency, human rights law at all times might emerge as a source of 'disruption' and challenge with a finding that rights have been violated. In other words, the legal framework of the AI Act does not mark the conclusion of the legal debate regarding the nature and scope of, for example, the requirement of human oversight. This relates to the flexibility, abstractness and context-dependency of human rights law. This also reveals the potential of human rights law and the positive aspect of its flexibility, abstractness and context-dependency.

The nature of the harm that human rights law is meant to protect can be also malleable, which can help in using it as a source in reaching the conclusion that the AI regulations are not sufficiently stringent. An example to this effect will be the argument that non-transparent machine learning models that are difficult to interpret (i.e. difficult to understand why they generate a specific output) might be contrary to the right to effective remedy. The reframing of the harm via the procedural right to effective remedy, can facilitate the argument for more regulation/limitations upon AI systems, as a matter of human rights law. Section 6.4 below will further explain the role and meaning of procedural harm.

AI regulations invoked in the context of human rights law

The second axis of the interrelationship between human rights law and AI regulation moves in the opposite direction: the AI Act can serve as an external reference point for human rights law. Specifically, it may provide a useful framework for assessing whether a State has fulfilled its positive or negative obligations under human rights law. This would not be unique since it has

[273] Sue Anne Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (2022) 40(1) *Nordic Journal of Human Rights* 216.

happened in other areas, such as climate change or secret surveillance,[274] where detailed technical requirements and standards have been developed. These requirements and standards are then invoked in the human rights law reasoning for determining the content of obligations. For instance, the risk-based approach in the AI Act that implies prohibition or regulation of certain systems depending on their areas of application (such as asylum and migration), might not necessary be followed in human rights law. In other words, prohibitions as the content of negative obligations and further regulation as the content of positive obligations, might be demanded under human rights law *irrespective* of the area. Yet, the assumption that AI systems used in the area of asylum are high-risk and that these need to be better regulated, is an important starting point that can be utilized in the human rights law reasoning.

6.3 The understanding of harm in human rights law

Having explained at a more structural level the interaction of the two regulatory frameworks, here I will focus on human rights law. It might appear basic, but we need to start with the following fundamental question: What is a human right? While the 'human' element is clear and in the context of this report, the focus is on asylum-seekers, the 'right' element is more controversial. Without necessary entering into philosophical debates, there is an agreement that human rights are based on fundamental/important interests. When these interests are harmed, human rights law can be invoked.

What is the harm, however? Do we conceptualize the harm in terms of some specific outcomes. Example might be the rejection of an asylum application that leads to a deportation order. Another example of an outcome-based harm is death of asylum-seekers trying to reach safe shores. Another possible harm is deprivation of liberty for the purposes of immigration control,[275] which in turn can have negative repercussions for the asylum claim. In alternative, do we conceptualize the harm in terms a procedural harm? An example to this effect would be flawed procedure leading to a rejection of the asylum claim.[276]

[274] See *Verein KlimaSeniorinnen Schweiz and Others v Switzerland* [GC] App no 53600/20, 9 April 2024; Centrum för Rättvisa v Sweden [GC] App no 35252/08, 25 May 2021.

[275] See Article 5(1)(f) ECHR.

[276] A relevant question here pertains to the standards that should be used as a reference for assessing incorrectness. Is the procedure incorrect because of what data has been used for the development of the system or because of the data used for testing the system? This would imply technical (computer science related) standards for assessing incorrectness. In alternative (or in conjunction) is the procedure flawed because of how state officials use the system in their decision-making process?

Another possibility is that the administrative decision to detain an asylum-seeker was taken on the basis of a flawed procedure. Do we conceptualize the harm as a *risk* of some negative outcomes for the individual? Do we conceptualize the harm on a more systemic level? For example, forecasting tools for assessing risks for future movement and displacement towards Europe might be used to prevent people from leaving countries, which in turn might prompt them to use more dangerous migration routes. If such a systemic approach is taken, how do we causally link this systemic harm to harm to specific individuals? Another question that might arise is whether it is possible to articulate the harm to important interests when the role and the operation of the AI systems is not clear. How can the extent and the seriousness of any harm be articulated if the causal role of the AI system is not clear?[277]

It is important to recognize that in the context of asylum and migration control, certain harms - though ethically troubling - may fall outside the scope of what human rights law is designed to address. Human rights law has its limits: it does not necessarily engage with the broader structural conditions that permit such harms to arise in the first place. Specifically, it does not challenge the foundational premise that nation States have wide discretion to control who may enter and remain within their territory. When AI systems are deployed within this framework, human rights law may have limited relevance. This is not a shortcoming unique to the use of AI; rather, it reflects the broader limitations of human rights law in addressing the structural dynamics of state sovereignty and border control.

If we place this limitation aside, the tendency of human rights law has been to conceptualize harm widely, including by *reconceptualising harm as risk of harm*.[278] This would mean that the mere exposure to risk is perceived as harm. This is quite prominent in the area of asylum, where the harm is precisely reconceptualized as a risk of harm (e.g. risk of persecution, or risk of torture, inhuman or degrading treatment). The question at the heart of the inquiry (i.e., whether the State might breach the right not to be subjected to torture, inhuman or degrading treatment) is whether there are risks of such severe

[277] District Court of The Hague, SyRI judgment (2020) ECLI:NL:RBDHA:2020:1878, para 6.44: the extent and seriousness of the interference with the right to respect for private life is 'colored by the answer to the question what precisely SyRI is.' The question was about how to legally interpret a risk report submitted by the system. The Dutch court concluded that although 'the use of SyRI in and of itself is not aimed at having legal effect – whether in private law, administrative law or criminal law – a risk report does have a similarly significant effect on the private life of the person to which the risk report pertains' (para 6.59).

[278] Vladislava Stoyanova, 'Fault, Knowledge and Risk within the Framework of Positive Obligations under the European Convention on Human Rights' 2020 *Leiden Journal of International Law* 601.

ill-treatments in the place where the person might be deported. This in turn shifts the attention towards the process of assessing this risk. The attention is directed towards the decision-making process performed by the national asylum authorities. If this process includes an AI system, can any risks (inherent, residual, acceptable to use the language from the AI Act) be reconceptualized as the harm *itself*? In other words, are the risks that the use of AI system poses to the risk assessment, harms? If yes, this seems to imply a tenuous causal link between the AI system (or some specific features of the system) and harm understood in a more traditional sense as actual negative effects upon fundamental interests.

If the focus of the analysis is on the interests protected by the right to *non-refoulement*, the answer should be negative. Put it differently, any risks that the use of AI system might pose to the risk assessment (i.e. the assessment whether there is risk of ill-treatment upon removal), might not really be understood as harms. The reason is that what matters is the risk of persecution or other forms of ill-treatment. Yet, since human rights law has reconceptualized harm as risk of harm in this way shifting the attention to the process of decision-making, risks that inhere in this process are relevant. In this sense, risks that inhere in the decision-making process are considered. In sum, the use of AI systems might be possible to conceptualize as harming procedural interests protected by human rights law. In this sense, the harm can be perceived as *a procedural harm.* This reconceptualization of the harm is of key importance for the framing of any obligations that State might have.

6.4 Procedural positive obligations

Rights are the logical connection between important interests and obligations.[279] It is the State that is the holder of these obligations and in this sense, human rights law regulates conduct of States. The conduct of private actors can be assessed in the human rights law review. This assessment is performed by asking how the State has regulated private actors, whether such regulations are sufficient, or how the State has responded to the conduct of private actors. It then follows that although private actors might be involved in harms, it is the conduct of the State and whether this conduct complies with obligations as might be demanded by human rights law that matters. As already mentioned above, in terms of their content, these obligations are negative and positive.

[279] J Raz, 'On the Nature of Rights' XCIII Mind (1984) 194, 200. Rights are 'intermediary conclusions between statements of the right-holder's interest and another's duty.' J Raz, 'Legal Rights' in Raz, *Ethics in the Public Domain: Essays in the Morality of Law and Politics* (Oxford Clarendon Press, 1994) 254, 259.

The obligation corresponding to the right to *non-refoulement* has been framed as a negative obligation: as a *prohibition* upon the State not to remove a migrant if there is a risk of harm. This is however a very simplistic way of understanding States' obligations. The assessment of breach of this obligation in fact includes a review of what procedure has been put in place by the State and what procedural safeguards have been followed.[280] The key question for assessing compliance with this obligation is therefore *how* the risk assessment has been performed by the relevant national authorities. In this sense, the content of the obligation is measures of *procedural safeguards*. This aligns with the conceptualisation of the harm, as explained above. Namely, it is a procedural harm that is at the centre of the inquiry whether human rights law has been breached.

At this point, a better understanding of the procedural obligations in inhuman rights law is pertinent. First a more general overview is offered, which enables an understanding of how human rights law has impacted national administrative procedures (Section 6.4.1). Thereafter, the analysis will zoom on more specifically on procedural obligations developed in the context of asylum (Section 6.4.2). The procedural obligations are explained in light of the case law of the ECtHR.

Positive human rights obligations and national administrative procedures

The positive obligations under the ECHR have been developed to the effect that they demand that States have proper national procedures with sufficient procedural guarantees. The assumption underpinning this development is that procedures contribute to the effective application of the substantive guarantees. In other words, fair procedures are arguably more likely to lead to fairer results/outcomes.[281] Put it differently, fair procedures are more likely to prevent harm (i.e. harmful outcomes).

The ECHR incorporates some explicit procedural guarantees.[282] Some of them are triggered *post factum*, i.e. after the harm has materialised (e.g. the State has the positive obligation to investigate for the sake of fact-finding/evidence collection). In contrast, the procedural guarantees discussed here are meant

[280] Spyridoula Katsoni, 'Is the Obligation Not to *Refoule* a Positive Obligation? An Intermediate Approach Toward the Classification Dilemma' (2025) 25(1) *Human Rights Law Review*.

[281] E. Brems, 'The "Logics" of Procedural-Type Review by the European Court of Human Rights' in J. Gerards and E. Brems (eds) *Procedural Review in European Fundamental Rights Cases* (CUP, 2017) 17, 18.

[282] Articles 5, 6, 13 of the ECHR and Articles 1, 2, 3 and 4 of Protocol 7 ECHR.

to apply *ex ante*, since their rationale is to prevent harm or the risk of harm (e.g. to prevent an arguably harmful decision by the national authorities or to prevent a decision that might pose a risk of harm).[283]

A clarification is due from the outset as to how the ECtHR has developed procedural safeguards as measures that form the content of positive obligations in human rights law. In particular, these procedural safeguards do not have an autonomous, independent and self-standing role. As Brems has observed, 'procedural shortcoming is not necessarily conclusive for the finding of a violation; in many cases it is one among several factors that contribute to such a finding.'[284] Arnardóttir has also explained how 'procedural elements' are invoked 'among the balance of reasons when the Court pronounces on the substantive merits and assesses the proportionality or reasonableness of a measure.'[285] Similarly, Gerards has noted how the Court 'has woven procedural elements into its substantive reasonableness review.'[286] Gerards has also added that '[p]rocedural arguments are supportive, i.e. as part of the overall set of arguments to be taken into account in building a "narrative" leading up to' a judgment.[287]

Procedural guarantees as supportive elements in the overall determination of the content of the positive obligations

To clarify the dependent nature of the procedural guarantees under the ECHR, it is necessary to examine the specific state conduct being challenged by the applicant. This conduct stems from *decisions* made at the national level, which may originate from the legislature, administrative authorities, or judicial bodies.

[283] To avoid confusion, the following clarification is due. In Section 6.1 it was stated that human rights law implies *post factum* review. The assessment whether a State has complied with its procedural positive obligation is therefore *post factum*. It is a *post factum* assessment whether and how the procedural guarantees had been applied at the relevant point in time in the past. The assumption is that at this relevant point, if the procedural guarantees had been followed, it would have been less likely that harm would materialize. In this sense, the procedural guarantees themselves can be perceived as being *ex ante*.

[284] E. Brems, 'Procedural Protection. An Examination of Procedural Safeguards Read into Substantive Convention Rights' in E Brems and J. Gerards (eds) *Shaping Rights in the ECHR* (CUP, 2014) 137, 158.

[285] O. Arnardóttir, 'The "Procedural Turn" under the European Convention on Human Rights and Presumptions of Convention Compliance' (2017) I CON 9, 14.

[286] J. Gerards, 'Procedural Review by the ECtHR: A Typology' in J. Gerards and E. Brems (eds) *Procedural Review in European Fundamental Rights Cases* (CUP, 2017) 127, 129.

[287] J. Gerards, 'Procedural Review by the ECtHR: A Typology' in J. Gerards and E. Brems (eds) *Procedural Review in European Fundamental Rights Cases* (CUP, 2017) 127, 149.

When the decision is made by the legislature or the government,[288] it is likely to be more abstract, reflecting its general applicability. Conversely, decisions by administrative or judicial bodies are often more specific and tailored to the particular circumstances of specific individuals. It is the latter situation that is of interest here since the focus is when an administrative/public body issues an asylum decision.

A distinction can be made between the outcome of decisions and the process through which those decisions are made. The outcome pertains to the content or substance of the decisions at the national level, reflecting how different interests have been balanced in a specific manner. The process, on the other hand, refers to the steps undertaken at the national level to arrive at these decisions, including the method used to balance competing interests.[289]

The Court has established specific *procedural standards/safeguards* in its reasoning for evaluating national processes. These standards may include access to the procedure, its overall quality, timeliness, effectiveness, independence, the involvement of affected individuals, and the clarity of the reasoning behind decisions.[290] The role these procedural standards play in the Court's reasoning can vary in significance and in their relationship to the outcome (i.e., the reasonableness of the decision). This variability makes it challenging to treat these standards – either individually or collectively – as distinct, self-standing positive obligations. Instead, they serve as integral components of the Court's reasoning. At a broader level, however, it can be asserted that States have a positive obligation to establish effective national procedures to safeguard rights. As much importantly, the reasoning behind the national decisions has to be clear and possible to understand.

These procedural guarantees can be viewed as undermined, if it is not possible to assess the nature of the AI system used since, for example its model or any features used in the model have not been disclosed. Non-disclosure of the data that is processed by AI systems or the absence of notifications to individuals, raises also doubts as to whether the procedure is of sufficient

[288] E.g. *M.A. v Denmark* [GC] App no 6697/18, 9 July 2021, para147-50. See also M Saul, 'The European Court of Human Rights' Margin of Appreciation and the Process of National Parliaments' (2015) 15 *Human Rights Law Review* 745.

[289] L. Huijbers, *Process-based Fundamental Rights Review* (Intersentia, 2019) 113.

[290] E.g. *Hattan and Others v the United Kingdom* [GC] App no 36022/97, 8 July 2003, para 104; *Roche v The United Kingdom* [GC] App no 32555/96, 19 October 2005 para162-7; *A., B. and C v Ireland* [GC] App no 25579/05, 16 December 2010 para 267; *Gaskin v the United Kingdom* App no 10454/83, 7 July 1989 para 49; *Uzbyakov v Russia* App no 71160/13, 5 May 2020, para106; *P. and S. v. Poland* App no 57375/08, 30 October 2012, para 111; *Tanda-Muzinga v France* App no 2260/10, 10 July 2014 para 82.

quality.[291] Disclosure and notification only after specific requests by individuals that might be affected, can be also challenged as not offering sufficient procedural guarantee. The procedural safeguard that decisions affecting individuals must be reasoned, can be also undermined when these decisions are based on AI systems. This might be the case if there is no manual review for verification of correctness, relevance, and completeness of the information generated by the AI system.[292] Automation and confirmation bias should be also considered in this context.[293]

As already suggested above, the procedural guarantees developed in human rights law are not absolute. It would then follow that the absence of transparency in the decision-making process due to the involvement of AI systems or the non-disclosure of certain features (e.g. the risk model used, the risk indicators, the training data), cannot automatically lead to the conclusion that human rights law has been breached. The use of systems or certain feature of the systems might be deliberately kept secret,[294] which might be justifiable.[295] It then follows that there might be public interests at stake as to why public authorities might choose *not* to disclose details about the operation of the system. On this point, in the balancing reasoning, the State might have to justify the absence of transparency.

As much important, even if overall the procedural guarantees might be found insufficient in the context of procedures where an AI system has been used, this in itself does not mean that the State would have the concrete obligation under human rights law to disclose all the inner workings of the system (assuming that such inner workings are knowable and thus possible to disclose)

[291] See District Court of The Hague, SyRI judgment (2020) ECLI:NL:RBDHA:2020:1878, para 6.49.

[292] For similar conclusions, however, from the perspective of EU fundamental rights law, see Simona Demkova, 'The EU's Artificial Intelligence Laboratory and Fundamental Rights' in Melanie Fink (ed), *Redressing Fundamental Rights Violations by the EU: The Promise of the 'Complete System of Remedies* (Cambridge University Press, 2025*)* 391, 408.

[293] Saar Alon-Barkat and Madalina Busuioc, 'Human-AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice' (2023) 33 *Journal of Public Administration Research and Theory* 153.

[294] J Niklas, 'Human Rights-Based Approach to AI and Algorithms Concerning Welfare Technologies' in W. Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (2021) 517, 521–2.

[295] An argument might be that individuals should not be enabled to 'game the system. 'See A. Rachovitsa and N. Johann, 'The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch *SyRI* Case' (2022) 22 *Human Rights Law Review* 1, 11.

or even some concrete features of the system.[296] The reason is that it is the *procedural* guarantees assessed *in their entirety* that are at the core of the human rights law review and the assessment of breach. The implication is that the concrete content of the obligation is not necessary disclosure of concrete features. Non-disclosure of some features might be counterbalanced by other procedural guarantees, which reveals the flexibility that characteries human rights law. Another implication is that since some procedural failures might be only one element in the reasoning, correcting them might not necessary be required (i.e. correcting these procedural failures might not necessary be the content of the obligation).[297] It all depends on the reasoning. In some situations, the ECtHR might reason in a way by which a specific procedural failure is given a very prominent role in the reasoning. In this case, it is easier to argue that this precise failure needs to be removed.

Any procedural guarantees that might be demanded by human right law, have another characteristic, which might reduce their robustness as a tool for challenging the use of AI systems. When the ECtHR invokes procedural guarantees, it frequently evaluates not only the guarantees but also the outcome of procedure (i.e. the decision eventually taken by the national authorities). It then follows that the reasonableness of *both* the process and the outcome (i.e., the substance of the decision impacting the applicant), are evaluated.[298] As a result, the Court's analysis may not always clearly differentiate between whether the national decision (the outcome) was reasonable or whether the process leading to that decision met sufficient qualitative standards. In some cases, the absence of procedural safeguards leads the Court to view the outcome as more questionable. Conversely, adherence to procedural guarantees at the national level increases the likelihood that the Court will accept the outcome as reasonable.

[296] See A. Rachovitsa and N. Johann, 'The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch *SyRI* Case' (2022) 22 *Human Rights Law Review* 1: 'It needs to be clarified that the Court did not find that the SyRI legislation's unlawfulness, insofar as the use of SyRI was concerned, entailed an obligation on the state to disclose the inner workings of the risk model. (para 6.115)'.

[297] On the difficulties in determining the precise content of the obligations corresponding to the rights protected by the ECHR, see Vladislava Stoyanova, 'Correlativity Between Human Rights and Positive Obligations and Its Role for the Execution of Judgments Delivered by the European Court of Human Rights' (2024) 5 European Convention on *Human Rights Law Review* 455.

[298] J Gerards, 'Procedural Review by the ECtHR: A Typology' in J. Gerards and E Brems (eds) *Procedural Review in European Fundamental Rights Cases* (CUP, 2017) 127, 155; J. Gerards 'Procedural Review by the European Court of Human Rights – a Typology of Functions for the Court's Reasoning' in D. Harvey and X. Groussot, *Process Federalism in EU Law* (EULawLive 2024).

It follows from the above that there is a scope of flexibility as to the relationship between the outcome and the process in the human rights law review. Yet, the development of procedural guarantees under the ECHR has implied a shift of focus in favour of the process. The focus on procedural guarantees in the Court's reasoning does not only imply that the reasonableness of the decisions/the outcome as such might either not be addressed or be addressed only indirectly, but it also implies that the causation between the harm claimed by the applicant and the alleged omission by the State is of less important.

This is of key significance, since it implies that the human rights law review does not focus on omissions *as causally linked to harm*. Rather, it centres on procedures and procedural safeguards, based on the normative understanding that sound procedures are both intrinsically valuable and more likely to lead to fair and lawful outcomes. As a consequence, the limited possibilities to factually reveal the inner workings/the design of AI systems and the limited expertise in even understanding them, including how they cause harm, are less important. The procedural safeguards might operate in tandem with primary issues relating to rights, since, for example, the Court might still invoke the severity of the harm in its reasoning. Yet, when the Court's review is about the national process, the actual harm and its causes are in the background.

To recap, procedural guarantees developed under human rights law may serve as a tool for challenging the use of AI systems and certain features of their design or application. However, the inherent flexibility and context-specific nature of these guarantees make it difficult to define precise procedural requirements. In some areas of its case law, the European Court of Human Rights has articulated standards that national decision-makers must consider – one such area being asylum decisions, as the next section will explore. Still, as will also be noted, these procedural standards remain abstract, and their concrete implications depend heavily on the specific circumstances of each case.

Procedural guarantees in the context of asylum

AI systems might be used in various ways in the refugee status determination procedure. They might be used for assisting in, for example, the identification of priority cases or in carrying out any repetitive tasks in the decision-making process. They might be used to help in the credibility assessment. They might help in the assessment of evidence: '[t]he provision of evidence by AI systems such as biometrics, automated interviewing and digital analysis may provide further material facts and help the decision-maker corroborate evidence.'[299]

[299] Niamh Kinchin, 'Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination' (2021) 37(3) *Law in Context* 45, 59.

Fully automated AI supported decision-making with algorithms without human intervention, would be an extreme example.[300] At the time of writing, the ECtHR has not adjudicated a case where an AI system has been used in the asylum procedure. It is possible, however, by analogy to reflect upon how the standards already developed in the case law might be relevant if such a case were to arise.

The Court has adjudicated many cases in which it had to review whether substantial grounds had been shown for believing that the person in question, if removed, would face a real risk of being subjected to ill-treatment in the destination country.[301] To make this review, the Court has developed certain procedural standards, which can be also based on the combined application of Article 3 and Article 13 of the ECHR. To this effect, where the individual has an 'arguable complaint' that his removal would expose him to treatment contrary to Article 2 or 3 of the Convention, the individual must have access to an effective remedy at national level. This remedy must be effective in law and in practice. This necessary requires inter alia, *independent and rigorous scrutiny* of any claim that there exist substantial grounds for fearing a real risk of treatment contrary to Articles 2 or 3.[302] The national authorities have *to assess claims in an adequate and sufficient way* supported by relevant materials originating from reliable and objective sources.[303] This would mean that the quality of the sources used, including the quality and consistency of the methods for collecting and forwarding information, is an important procedural guarantee.[304]

[300] Niamh Kinchin, 'Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination' (2021) 37(3) *Law in Context* 45, 59.

[301] See. e.g. *F.G. v Sweden* [GC] App no 43611/11, 23 March 2016, para 110-127; *J.K. and Others v Sweden* [GC] App no 59166/12, 23 August 2016, para 77-105.

[302] M.*S.S. v Belgium and Greece* [GC] App no 30696/09, 21 January 2011, para 293; M.*K. and Others v Poland* App no 40503/17, 42902/17 and 43643/17, 23 July 2020, para 142-148 and 212-220; *D.A. and Others v. Poland* App no 51246/17, 8 July 2021, para 89-90; *Hirsi Jamaa and Others v. Italy* [GC] App no 27765/09, 23 February 2012, para 201-207; *Sharifi and Others v Italy and Greece* App no 16643/09, 21 October 2014, para 240-243. *Abdolkhani and Karimnia v Turkey* App no 30471/08, 22 September 2009, para 107-117; *Gebremedhin [Gaberamadhien] v France* App no 25389/05, 26 April 2007, para 53-67; I.*M. v France* App no 9152/09, 2 February 2012; *Chahal v. the United Kingdom* [GC] App no 22414/93, 15 November 1996, para 147-154.

[303] *Khasanov and Rakhmanov v Russia* [GC] App no 28492/15 and 49975/15, 29 April 2022, para 103; F.*G. v Sweden* [GC] App no 43611/11, 23 March 2016, para 117.

[304] In assessing the weight to be attached to country material, the Court has found that consideration must be given to the source of such material, in particular its reliability and objectivity. In respect of reports, the authority and reputation of the author, the seriousness of the investigations by means of which they were compiled, the consistency of their conclusions and their corroboration by other sources are all relevant considerations. See *Khasanov and Rakhmanov v Russia* [GC], 2022, para 114; *J.K. and Others v Sweden* [GC], 2016, para 88.

The Court has also added that the assessment as to whether there is a real risk must necessarily be a *rigorous* one.[305] The standard of rigorousness as a procedural guarantee should be distinguished from accuracy and correctness that pertain to the outcome (i.e. was the person actually exposed to the risk of ill-treatment). When it comes to accuracy and correctness, it is very difficult (even impossible) to assess them in the context of asylum. The reason is that one needs to engage in a counterfactual assessment as to what would have happened in case of a negative decision of rejection and removal to the country of origin. In alternative, one needs to assess and study what had actually happened to the person after deportation (i.e. was he or she actually exposed to real risk). I will revert to this point below. I will revert to these specificities of the process below.

The Court has held that asylum-seekers need to have a *dequate information about the asylum procedure* to be followed and their entitlements in a language they understand, and have access to a reliable communication system with the authorities. In addition to the availability of interpreters, training of staff and access to legal aid, the Court has also added the requirement that asylum-seekers be given the *reasons for the decisions* taken by the national authorities.[306] The opacity of asylum decisions, which implies opacity of the underlying reasoning, has been a generally recognized problem. In this sense, the use of AI systems might not lead to a change. Yet, the use of AI systems to support decision-making 'shifts the risk to processes that are less, or not at all, accessible to public, democratic, political and judicial scrutiny'.[307]

There is at least one other *procedural shift* that happens when AI systems are used and that can be challenged as being incompatible with procedural safeguards. AI supported decision-making can restrict the scope of discretion exercised by *individual* state agents (i.e. individual decision-makers tasked to take decisions within the refugee status determination procedure). This leads to more standardised procedures. On the other hand, as Westerling has noted 'standardised and collective decision-making is nothing new' since '[m]anuals and country of origin-specific policies have long been directing and steering asylum officials towards consistent decision-making.'[308] Such procedural

[305] F.*G. v Sweden* [GC] App no 43611/11, 23 March 2016, para113; *Khasanov and Rakhmanov v Russia* [GC] App no 28492/15 and 49975/15, 29 April 2022, para 109.

[306] *M.S.S. v Belgium and Greece* [GC] para 300-302, 304, and 306-310; *Hirsi Jamaa and Others v Italy* [GC] para 204; *D v. Bulgaria* App no 29447/17, 20 July 2021, para 120-137.

[307] Mathilda Arvidsson and Gregor Noll, 'Decision Making in Asylum Law and Machine Learning' (2023) 92 *Nordic Journal of international Law* 56.

[308] Frida Alizadeh Westerling, 'Technology-Related Risks to the Right to Asylum: Epistemic Vulnerability Production in Automated Credibility Assessment' (2022) 13 *European Journal of Law and Technology.*

standardisations can prevent subjectivity and arbitrary discretion. Yet, Westerling also clarifies that '[a]utomation strengthens the collective use of discretion', which 'also changes the nature of collective discretion.' Westerling adds that '[b]ehind the technological interfaces are systems analysts and software designers creating ways to analyse data (evidence), which in some respects broadens but also anonymises the group(s) of people practising collective discretion. Hence, it transfers and distributes discretion to a larger group of people.'[309] Noll and Arvidsson have identified the same problem. They have noted how *discretion is shifted* form the discretion of the decision-makers within the administrative procedure to the discretion of the designers of AI systems who might have 'unchecked discretion in data wrangling',[310] in how the initial training data is collected, selected, and curated.[311]
The complexity of datafication therefore needs to be highlighted. This in turn implies difficulties in questioning the training data. The difficulties for questioning the outputs of the systems needs to be also added. Asylum seekers would face obstacles 'to prove that the data is flawed, misleading, low in probative value, or inapplicable to the case at hand'.[312] All of this places the person in a vulnerable position.

A key argument in favour of using AI supported decision making has been effectiveness and efficiency.[313] Indeed, effectiveness in terms of speed of the procedure is important.[314] On the other hand, the speedy processing of asylum claims cannot take priority over the *effectiveness* of the essential procedural guarantees to protect from risk of ill-treatment. This principle of effectiveness

[309] Frida Alizadeh Westerling, 'Technology-Related Risks to the Right to Asylum: Epistemic Vulnerability Production in Automated Credibility Assessment' (2022) 13 *European Journal of Law and Technology*.

[310] This means 'strategies for selecting and managing large, aggregated dataset to produce a model and story.' S Jiang and J Khan, 'Data wrangling practices and collaborative interactions with aggregated data' (2020) *International Journal of Computer-Supported Collaborative Learning* 257; D Lehr and P Ohm, 'Playing with the data: What legal scholars should learn about machine learning' (2017) *U.C. Davis Law Review* 653.

[311] Mathilda Arvidsson and Gregor Noll, 'Decision Making in Asylum Law and Machine Learning' (2023) 92 *Nordic Journal of international Law* 56, 90.

[312] Frida Alizadeh Westerling, 'Technology-Related Risks to the Right to Asylum: Epistemic Vulnerability Production in Automated Credibility Assessment' (2022) 13 *European Journal of Law and Technology*.

[313] Ludivine Sarah Stewart, 'Fair and Efficient Asylum Procedures and Artificial Intelligence: Qua Vadis Due Process?' (2024) 55 *Computer Law and Security Review* 1, 4.

[314] *M.S.S. v. Belgium and Greece* [GC] para 292; *E.H. v. France* App no 39126/18, 22 July 2021, para 195; *B.A.C. v Greece* App no 11981/15, 13 October 2016, para 36-46.

has been used for challenging short-time period for lodging appeals.[315] It has not been used for challenging the use of AI systems yet. When considering efficiency and what can be sacrificed for the sake of efficiency, at least two distinctiveness of the use of AI in the asylum procedure, should be noted: first, the high-stakes nature of the decisions (*refoulement*) and second, the power imbalance between the decision-makers and the affected individuals.[316]

Some form of participation of the asylum-seeker has been also an important procedural guarantee. For example, the person should be allowed the possibility to respond to findings that might negatively affect the application.[317] It would then follow that the person should be allowed access to the output of AI supported systems, i.e. the report. However, Francesca Palmiotto has argued 'accessing only the final output of the system may not be sufficient to challenge automated evidence.'[318] The person should be offered the possibility to access the process that led to this output. To what extent such access should be provided as a matter of human rights law procedural guarantee is, however, yet to be tested.

Challenge to the mere application of the relevant legal standards

So far Section 6.4.2. examined how the procedural guarantees as developed in human rights law, might be compromised with the use of AI systems in the asylum decisions-making process. The objective of this final part in the Section is similar since it also formulates a procedural argument (i.e. undermining of procedural guarantees); however, this argument goes further than mere a procedural problem. In particular, the argument advanced here is that the use of AI in the asylum determination procedure might be difficult to square with the mere *nature* of the procedure.

In particular, this study will explain how *AI supported decision-making presents distinctive problems for applying the legal standards in the procedure about assessment of protection needs.* To explain this, it is first relevant to highlight

[315] An unreasonably short time-limit to submit a claim, such as in the context of accelerated asylum procedures or to appeal a subsequent removal decision can render a remedy practically ineffective, contrary to the requirements of Article 13 taken together with Article 3 of the Convention. See e.g. *I.M. v France* App no 9152/09, 2 February 2012; *R.D. v France* App no 34648/14, 16 June 2016, para 55-64; *E.H. v. France* App no 39126/18, 22 July 2021, para 180-207.

[316] Ludivine Sarah Stewart, 'Fair and Efficient Asylum Procedures and Artificial Intelligence: Qua Vadis Due Process?' (2024) 55 *Computer Law and Security Review* 1, 5.

[317] Gregor Noll, 'Evidentiary Assessment in Refugee Status Determination and the EU Qualification Directive' (2006) 12 *European Public Law* 303–304.

[318] Francesca Palmiotto 'Procedural Fairness in Automated Asylum Procedures: Fundamental Rights for Fundamental Challenges' (2024) 55 *Computer Law and Security Review* 1, 6.

that the assessment of risk of ill-treatment under Article 3 ECHR is forward looking. It is about risks that might materialise in the future. There is no clear standard as to how near or far this future should be.[319] For example, should the risk materialize immediately after removal and arrival in the country of origin? In alternative, should the risk assessment also consider risks that might materialize soon and not that remotely after arrival? If the latter, the timeline for assessing remoteness is far from clear.[320]

AI systems might access volumes of data and be capable to quickly process it to predict risk. The systems are trained on data (i.e. training data), which can be conceptualised as the 'ground truth'.[321] This could imply, for example, that the credibility will have to be tested against this 'ground truth'. As Niamh Kinchin has, however, asked what is the 'ground truth' when the relevant legal standard is based on risk in the future? She has also clarified that '[e]ven where the algorithm is designed to learn and reason through repeated experiences, the environment must be sufficiently regular to be predictable as 'learning can only properly happen with feedback on which decisions were correct, which incorrect, and on what grounds'.[322] Decision-makers in the area of asylum, however, do not normally have the opportunity to verify whether their decisions were correct. This implies that there is no test data against which the AI system might be tested (during the design phase and after being placed in operation). The availability of such data would imply the existence of some empirical studies indicating the correctness of the decisions taken by the national decision-making authorities in the area of asylum. Such studies would have to reveal what happened to the persons once deported to the country of origin. Such studies would have to reveal whether indeed the risks that they claimed in the context of the specific procedure actually materialized. These studies are difficult to find and to perform. In other words, we face the problem of absence of test data.

[319] How far or near in the future from the point in time when the national authorities have to take a decision that the applicant has protection needs since there is a risk of ill-treatment upon removal.

[320] For a detailed discussion, see Michelle Foster, Hannah Gordon, Helene Lambert and Jane McAdam, "Time' in Refugee Status Determination in Australia and the United Kingdom: A Clear and Present Danger from Armed Conflict?' (2022) 34 *International Journal of Refugee Law* 163.

[321] Niamh Kinchin, 'Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination' (2021) 37(3) Law in Context 45, 65; see also Niamh Kinchin, 'The Human in the Feedback Loop: Predictive Analytics in Refugee Status Determination' (2024) 6(3) *Law, Technology and Humans 23*.

[322] Niamh Kinchin, 'Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination' (2021) 37(3) *Law in Context* 45, 65.

As much important, even if we had such empirical studies that might be useful for the training of potential AI systems, the outcomes of such studies *cannot be legally relevant*. Refugee status determination procedure is about risk assessment. The core question is whether there is risk of ill-treatment at the time when the decision-making body makes the assessment. The assessment might be negative, which implies that the outcome of the assessment is the decision that there is no risk at the relevant time (i.e. when the refugee status determination decision is taken). Upon removal, the person might be subjected to ill-treatment (i.e. in reality, the risk does materialize). However, this actual materialisation of the risk does not render the initial risk assessment decision made by the national authorities necessary incorrect.

The main point here is that there is no data to verify the correctness of decisions. Niamh Kinchin and Davoud Mougouei have formulated the problem in the following way: the AI system predicts risk of future events based on past events, 'rather than the 'risk of risk".[323] AI cannot produce outputs without some relationship to prior knowledge (i.e. the training data). In this sense the outputs are linked to the past. In contrast, risk assessment in the context of asylum is forward-looking - it concerns not just future events, but future risks, which adds an additional layer of complexity.

Niamh Kinchin and Davoud Mougouei also add the following '[a]lgorithms learn by utilising training data that are historical and that can only be labelled according to evidence of past persecution in other places, based on the outcomes of other people's cases. Algorithms may prove useful for making predictions based on generalized risk, or even the previous outcomes of asylum adjudications, but will be less effective in relation to the claimant's own, inevitably subjective, experiences of persecution and risk, and the experiences of those close to them.'[324] Kinchin and Mougouei here point out the *individualized* nature of the assessment as to whether a person has protection needs.

6.5 Negative obligations

In contrast to Section 6.4 that outlined the procedural positive obligations under human rights law that can be relevant for regulating the use of AI systems in the asylum determination procedures, the objective of this Section to assess the role of negative obligations. As already mentioned in Section 6.2 above, the

[323] Niamh Kinchin and Davoud Mougouei, 'What can AI do for Refugee Status Determination? A Proposal for Removing Subjective Fear' (2022) 34 *International Journal of Refugee Law* 373, 393.

[324] Niamh Kinchin and Davoud Mougouei, 'What can AI do for Refugee Status Determination? A Proposal for Removing Subjective Fear' (2022) 34 *International Journal of Refugee Law* 373, 396.

right to *non-refoulement* can be also understood as triggering negative obligations, since the State is expected to refrain from removing the person. An argument could be therefore formulated that any risks posed by the use of AI systems in the asylum determination procedures, could be conceptualised as limitations upon the important interests protected by Article 3 ECHR. Given the tenuous causal links between such limitations and the actual risk of harm (i.e. risk of ill-treatment upon removal), such a conceptualisation is very contestable. Rather, as already explained above, such 'limitations' are likely to be conceptualised as possible procedural failures in the assessment of risk.[325]

It should be also considered that if an AI system is used to assess protection claims, such a system might use personal data, which might also make Article 8 of the ECHR a relevant provision. Article 8 of the ECHR protects the right to private and family life. Any measures of interference with private life within the asylum determination procedure must meet certain tests so that they can be deemed compatible with human rights law. These tests are the legality (Section 6.5.1), suitability (Section 6.5.2) and proportionality (6.5.3).[326]

This Section reviews these three tests by asking the question whether they can be complied with if national procedures include the involvement of AI systems, which could be conceptualised as an interference with the interests protected by the right to private life. If the interference does not satisfy any of the three tests, it would amount to violation of negative obligations under Article 8 of the ECHR.

Legality

Interferences with private life that the application of AI systems might imply, must be in accordance with the law. According to the case law of the ECtHR, this does not need to be a law adopted by the national parliament: this test can also be met with any generally binding regulation or even legal standards established by national courts. The relevant question to ask is whether the interference had 'some basis in domestic law.'[327]

Not only does Article 8 ECHR demands a legal basis, but this basis on which the interference is grounded must be sufficiently accessible and foreseeable. This means that the legal basis must be sufficiently clear so as to enable

[325] The mere conceptualisation of the use of AI systems in the procedure as 'limitation' is also questionable, which explains the placement of the word in inverted commas.

[326] See generally J Gerards and H Senden, 'The Structure of Fundamental Rights and the European Court of Human Rights' (2009) 7(4) *International Journal of Constitutional Law* 619.

[327] *Malone v the United Kingdom* App no 8691/79, 2 August 1984.

individuals to regulate their conduct accordingly.[328] When applied to AI systems, the key question can be formulated with reference to the *required level of precision of the legal basis*. According to the ECtHR, the level of precision required of the national regulatory framework depends to a considerable degree on: 'the content of the instrument in question, the field it is designed to cover and the number and status of those to whom it is addressed'.[329] The level of precision and detail can depend on the seriousness of the interference and the position of vulnerability.[330]

A compelling question arises: can the use of an AI system – one that may interfere with the right to private life but is permitted under the EU AI Act – be considered 'in accordance with the law' within the meaning of the European Convention on Human Rights (ECHR)? Put differently, do the regulatory standards set out in the EU AI Act satisfy the ECHR's legality requirement? This remains an open question. As detailed in Section 4, the AI Act contains numerous vague and ambiguous provisions that may fall short of the ECHR's thresholds of accessibility and foreseeability. Consequently, to meet the demands of the legality requirement, national authorities may be obliged to adopt more specific and detailed regulations in certain contexts.

Suitability

Given the public interests of immigration control and effective administrative procedures, the use of AI systems might be accepted as a suitable measure. As much important, the test of suitability in general is easily accepted as being met in the human rights law review performed by the Court.[331] It is rather the

[328] *Sunday Times v the United Kingdom* App no 6538/74, 26 April 1979, para 48; *Centrum för Rättvisa v Sweden* [GC] App no 35252/08, 25 May 2021, para 246.

[329] *S. and Marper v the United Kingdom* App no 30562/04 and 30566/04, 4 December 2008, para 96 and para 99: 'It reiterates that it is as essential, in this context, as in telephone tapping, secret surveillance and covert intelligence-gathering, to have clear, detailed rules governing the scope and application of measures, as well as minimum safeguards concerning, *inter alia*, duration, storage, usage, access of third parties, procedures for preserving the integrity and confidentiality of data and procedures for its destruction, thus providing sufficient guarantees against the risk of abuse and arbitrariness.'

[330] See for example the German case regarding the extraction of data from mobile devises, which was found unlawful by the domestic court. The form that the applicants were obliged to sign did not explain what data would be collected, how it would be processed, whether they could object to such processing or which authorities they could lodge a complaint. See Francesca Palmiotto and Derya Ozkul, '"Like Handing My Whole Life Over" The German Federal Administrative Court's Landmark Ruling on Mobile Phone Data Extraction in Asylum Procedures' "Like Handing My Whole Life Over" – Verfassungsblog

[331] Vladislava Stoyanova, 'Populism, Exceptionality and the Right of Migrants to Family Life Under the European Convention on Human Rights' (2018) *European Journal of Legal Studies*

test of proportionality that can challenge interference measures, which might lead to their assessment as being contrary to human rights law.

Proportionality

The proportionality test in the human rights law review seeks to respond to the question whether the measure of interference upon important interests (i.e. the harm inflicted upon these interests) is proportionate in light of the competing interests that might favour the measure. The ECtHR has formulated this review in the following way: '[a]n interference will be considered "necessary in a democratic society" for a legitimate aim if it answers a "pressing social need" and, in particular, if it is proportionate to the legitimate aim pursued and if the reasons adduced by the national authorities to justify it are "relevant and sufficient"'.[332]

An important interpretative technique used by the Court to assess proportionality is reference to external legal standards. In the context of data protection, such external legal standards can come from EU data protection law.[333] Fundamental principles of data protection developed in EU data protection law are transparency, the principle of purpose limitation and the principle of data minimisation. The first one would, for example, imply that if the affected persons have no information as to the risk analysis method used by the AI system, they might be unable to respond to any outputs generated by the system.[334] The AI Act and the CoE AI Framework Convention can be another source for external standards. If these have been complied with in the development and the deployment of the AI system, it might be easier to conclude that proportionality has been complied with. Here it is also pertinent to clarify that the legality test might be merged with the proportionality test. In other words, the former might not be addressed with a conclusive answer; the Court's reasoning might rather focus on analysing proportionality by also taking into account the relevant legal framework and its quality.[335]

Most importantly, however, proportionality in human rights law implies an analysis in light of the specific context and in light of the position of the specific individual whose interests have been affected. This relates to one of the distinctive characteristics of human rights law, as noted in Section 6.1 above, namely an individualized approach. This means that although *in abstract*, an

[332] *Fernandez Martinez v Spain* [GC] App no 56030/07, 12 June 2014, para 123.

[333] Lena Enqvist and Markus Naarrijärvi, 'Discretion, Automation, and Proportionality' in Markku Suski (eds) *The Rule of Law and Automated Decision-Making* (Springer, 2023) 147.

[334] See the reasoning of the Dutch Court in SyRI, para 6.90.

[335] *Centrum för Rättvisa v Sweden* [GC] App no 35252/08, 25 May 2021.

AI system might meet the proportionality test, its concrete application might still be disproportionate and thus contrary to human rights law. Human rights law therefore demands *in concreto* assessment. This could imply that the asylum authorities might have to be required under human rights law to decide (or at least to have a procedure in place) whether the use of the system is warranted in the specific case of the concrete asylum seeker.[336] Such *in concreto* assessment is possible to perform only when faced with a concrete case.

However, it is possible to undertake a more abstract and general assessment by posing the following question: Are residual risks – understood as limitations on the fundamental interests protected by Article 8 of the ECHR – proportionate? It is important to recall that once a residual risk is deemed 'acceptable' under Article 9(4) of the AI Act, individuals are exposed to that risk. This exposure can be viewed as a limitation on human rights, thereby triggering the need for a proportionality assessment under the ECHR framework.

Here it is also worth reminding that, as explained in Section 4, in the view of the AI Act, a high-risk system can pose *acceptable* residual risks.[337] If judged as such, this does not make the system contrary to the AI Act legal standards. Rather risk management measures need to be undertaken to address the residual risks. Such risk management measures have to be understood in light of the overall objective of the AI Act that any regulatory burdens and costs (importantly, risk management measures are such burdens) have to be proportionate to any risks.[338] This regulatory approach implies that residual risks to fundamental rights are accepted under the terms of the AI Act. Are they however acceptable and proportionate under human rights law? Not necessary. Individuals may be exposed to harm from residual risks – an outcome that may be permissible under the AI Act but could nonetheless violate human rights law.

[336] See also Francesca Palmiotto 'Procedural Fairness in Automated Asylum Procedures: Fundamental Rights for Fundamental Challenges'55 (2024) *Computer Law and Security Review* 1, 5.

[337] Article 9(5) AI Act.

[338] See the Proposal for a Regulation Laying down Harmonized Rules on AI, Explanatory Memorandum, COM(2021) 206 final, 21 April 2021, para 3.5.

It is important to also observe that the understanding that complete elimination of risks is not possible is reflected not only in the regulatory approach of the AI Act,[339] it is also reflected in human rights law.[340] This means that human rights law does not demand categorical elimination of risks, a requirement that in any case comes as completely unreasonable. Risks to fundamental interests (understood as limitations upon such interests as protected by Article 8) can be balanced. This encapsulates the essence of the proportionality review. The central question is not whether balancing should occur, but rather *what is being weighed on each side of the scale*. On one side, we find the fundamental interests of individuals. But what lies on the other? What are these human interests being balanced against?

As explained in Section 4, under the terms of the AI Act, residual risks can be considered acceptable if they are proportionate to the benefits of the 'high-risk' AI system. This 'depends partly on the cost and effectiveness of precaution taken relative to alternatives.'[341] The question that emerges at this point is whether the host society has the political will to pay the cost of precaution in the context of the asylum determination procedures. If any residual risks from the AI systems imply limitations only upon privacy interests, perhaps the cost of precaution might be viewed as unjustifiable. If any residual risks imply more severe harm, the cost of precaution might be easier to consider as acceptable. Overall, however, in the asylum determination procedure, shifting more risks to asylum-seekers might be easier to justify and, in this sense, the host society might not be willing to pay the cost of precaution.[342]

[339] See e.g. Cindy Jardine et al, 'Risk Management Frameworks for Human Health and Environmental Risks' (2003) 6(6) *Journal of Toxicology and Environmental Health,* Part B 569, 572, 633.

[340] Vladislava Stoyanova, 'Fault, Knowledge and Risk within the Framework of Positive Obligations under the European Convention on Human Rights' 2020 *Leiden Journal of International Law* 601.

[341] H. Fraser and J. Bello y Villarino, 'Where Residual Risks Reside: A Comparative Approach to Article 9(4) of the EU's Proposed AI Regulation'.

[342] There has been a discussion about the use of AI systems in the public sector with reference to the dual role of public institutions – they need to function in an efficient way, while at the same time they need to protect citizens from harm. This point is crucial because public institutions exercise coercive authority – individuals are often compelled to interact with them, without the freedom to opt out. The legitimacy of this coercion rests on its justification. See Johan Laus, Sandra Watcher and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the EU AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 *Regulation and Governance* 3, 5. In the asylum and migration context, however, the expectations for justification and legitimacy are normally lower.

7. Conclusion

AI systems may enhance the efficiency of asylum decision-making processes. Assessing the empirical validity of this proposition can be and should be the subject of empirical research. This study is, however, not empirical but analytical in nature. The first question that this study aimed to address was whether AI systems used in the area of asylum comply with human rights law. This compliance question could only be answered by first identifying the harm to fundamental interests that systems might pose. Given the focus of the study, the harm can be initially circumscribed to harm related to the procedure for assessing protection claims. This is without prejudice to other types of harm that specifically migrants and asylum-seeker might be exposed to. In light of this initial limitation, the harm can be conceptualized as harm to the fundamental interests protected by the right not to be subjected to *refoulement.*

The study clarified the difficulties in establishing causation between this harm and the use of systems. These difficulties can be however overcome if the harm is conceptualised as a procedural harm. Such a conceptualisation would be fully in harmony with how the European Court of Human Rights (ECtHR) handles non-refoulement cases. In particular, *the focus is on the procedural harm*. This explains why the key for addressing the compliance question is assessment about compliance with procedural guarantees (i.e. procedural positive obligations in human rights law). Certain basic procedure guarantees (i.e. quality of the decision-making process, timeliness, effectiveness, independence, involvement of affected individuals, clarity of the reasoning behind decisions) have been developed in human rights law. *If an AI system is involved, these guarantees should be complied with, including the involvement of the affected individuals and clarity of the decisions that have affected them.*

Another concern is privacy. Given the collection and usage of data in procedures where AI systems might be involved, harm to private life might be involved. This is not necessary harm related specifically and exclusively to the procedure for assessing protection claims. However, the specific vulnerability of asylum-seekers needs to be acknowledged here; this vulnerability exceeds the normal vulnerability that individuals who are subjected to administrative procedures and discretion might experience. The specific vulnerability of asylum-seekers is related to their protection claims, the expectation that they substantiate these claims (when they have limited access to material evidence) and the possibility for negative repercussions if they do not fully subject themselves to

procedures.[343] All of this reveals the intertwinement between harm to private life and risk of *refoulement*. In relationship to the right to private life, the harm can be conceptualized in the following way – *the use of the systems constitutes an interference with private life and therefore a limitation of private life.* If this is the case, the interference has to comply with the tests of legality, suitability and proportionality to be in compliance with States' negative obligations.

Having identified the harm and thus the relevant rights, the compliance question necessarily demands engagement with the obligations corresponding to these rights. Only by specifying these obligations can we determine whether AI supported decision-making technologies in the area of asylum can be developed in compliance with human rights law. The specification of these obligations requires specification of the concrete conduct demanded from the State.

Such specification is contextual, which implies that *human rights law requires context-specific assessment*. In this sense, human rights law is yet to develop to address harm that AI systems might pose. When human rights law is faced with such novel challenges, it resorts to external legal frameworks as a source for interpretation. For this reason, the regulatory frameworks developed by the EU Act and the CoE AI Framework Convention are so crucial. Most importantly, however, compliance with these frameworks does not necessary imply compliance with human rights law. The context-specific review that human rights law demands, might still lead to a conclusion that a right has been violated, despite full compliance with the EU and CoE AI legal frameworks. To be more precise, *these external legal frameworks might not meet the test of legality* that *inter alia* demands some level of precision of the legal basis that allows the use of systems.

The test of proportionality is at the heart of human rights law. It prompts an analysis whether any risks (including residual risks) necessarily posed by high-risk AI systems are proportionate to any benefits. This assessment implies consideration of the cost of precaution: the cost for addressing the risks. *In the context of the asylum determination procedures, the cost of precaution might be considered more difficult to justify. This in turn might imply shifting risks to the asylum-seekers.* Such a shift might become more and more difficult to substantiate with the increase of the seriousness of the potential harm. The bigger role an AI system has in the decision-making

[343] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 109: 'Situations of asymmetry of power or information can affect the freely given requirement of consent, hence implying certain limitations to its protective function in certain situations [].' See also para 18 of the Explanatory Report to the CoE AI Framework Convention.

process, the easier it might be to expose the causal link between the system and the harm, which necessarily affects the proportionality analyses. When there is a higher risk that *both* the right to private life *and* the right to *non-refoulement* will be compromised, the proportionality test is more likely to weigh in favour of the individual.

Given the above-described analysis demanded by human rights law as characterised by context dependency, proportionality review and normative/value-related judgments, the invocation of human rights law in the AI legal frameworks can be assessed as problematic. To be more precise, the idea that inheres in the AI Act that private actors as providers of systems can make normative assessment about the meaning of human rights, is concerning. Human rights law implies normative judgments that should be taken by bodies that have legitimacy.[344]

Finally, this study stressed some inherent characteristics as to the nature of the refugee status determination procedure. These raise questions as compliance with the legality and the proportionality tests. This brings us to the research question as to how these technologies might help in the development of improved decision-making in the area of asylum by increasing the overall quality of the procedure. It some respects, they might, which might be possible to empirically test and prove. However, this study highlighted how *AI supported decision-making presents distinctive problems for applying the legal standards in the procedure about assessment of protection needs.*

A major issue is that asylum decision-makers typically lack the means to verify whether their decisions (e.g. granting or rejecting protection) were correct. This absence of feedback means there is no reliable test data to evaluate AI systems during development or after being placed in operation. Moreover, the historical data for the development of the system might not be relevant for predicting future risks in applicants' countries of origin.

This brings us to a final question raised in the study's introduction: *How might these technologies themselves change the practice of asylum law.* Such a change seems possible, given the increased importance of data, the selection of data and the role of programmers in the design of the algorithms. This in turn implies *shift away from discretion of individual decision-makers and in favour of discretion in the design.*

[344] Johan Laus, Sandra Watcher and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the EU AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 *Regulation and Governance* 3, 6.

8. Recommendations

1. *Uphold Procedural Fairness and Transparency:* The use of AI systems in the refugee status determination procedure should comply with basic procedure guarantees (i.e. quality of the decision-making process, timeliness, effectiveness, independence, involvement of affected individuals, clarity of the reasoning behind decisions). The assessment of compliance with these guarantees is context dependent and should be considered holistically. If the guarantees are undermined, this is not conclusive that human rights law has been breached. The various procedural guarantees should be assessed as a whole, where a less stringent application of one guarantee can be balanced by a more rigorous application of another.
2. *Context-Specific Proportionality Assessments:* Human rights law demands context-dependent review of the use of AI systems in the decision-making process. This makes the human rights law obligations difficult to specify. Similarly, the assessment whether the use of a system is proportionate, given the risks of harm that it might cause, is also context dependent. At very general level, however, human rights law demands proportionality review, which in turn prompts an analysis whether any risks – including residual risks – necessarily posed by high-risk AI systems, are proportionate to any benefits. This assessment implies consideration of the cost of precaution: the cost for addressing the risks that the national authorities who use the systems have to bear.
3. *Prioritize Precaution to Protect Rights:* In the context of the asylum determination procedures, the cost of precaution might be considered more difficult to justify. This in turn might imply shifting risks to the asylum-seekers. Such a shift might become more and more difficult to substantiate with the increase of the seriousness of the potential harm. The bigger role an AI system has in the decision-making process, the easier it might be to expose the causal link between the system and the harm, which necessarily affects the proportionality analyses.
4. *Use AI Regulations as a Floor, Not a Ceiling:* When evaluating AI use, consider compliance with dedicated AI regulations such as the EU AI Act and the Council of Europe AI Framework Convention. These regulations are likely to affect the proportionality review and the legal analysis in human rights. However, compliance with the EU AI Act and with the Council of Europe AI Framework Convention does not necessary imply compliance with human rights law.

5. *Ensure Public Oversight and Accountability:* The EU AI Act classifies asylum and migration AI systems, as "high-risk", which in turn triggers a whole set of regulations. The classification, however, is open to various exceptions. Even if the classification is applied, the regulatory standards can be an object of various interpretations and therefore be given different meanings. It is crucial that government authorities – not private developers – take responsibility for interpreting and applying these standards in the asylum context given the risks of harm to fundamental interests. Value judgements should not be left to private actors, such as tech companies.
6. *Recognize the Limits of AI in Refugee Decisions:* AI should not be used in ways that conflict with fundamental legal principles of asylum law. AI supported decision-making poses unique challenges in applying legal standards within procedures for assessing protection needs. Its use in refugee status determination processes may, in some situations, conflict with the principles of legality and/or key procedural guarantees. In particular, the historical data for the development of the system might not be relevant for the assessment of the risk for future harm in the country of origin. In essence, determining refugee status involves forward-looking assessments of risk which is something past data may not capture.

List of previous publications

Report and Policy Brief 2014:1, *Radikala högerpartier och attityder till invandring i Europa*, Mikael Hjerm and Andrea Bohman.

Report and Policy Brief 2015:1, *Internationell migration och remitteringar i Etiopien*, Lisa Andersson.

Research Overview 2015:2, *Politiska remitteringar*, Emma Lundgren Jörum and Åsa Lundgren.

Research Overview 2015:3, *Integrationspolitik och arbetsmarknad*, Patrick Joyce.

Research Overview 2015:4, *Migration och företagens internationalisering*, Andreas Hatzigeorgiou and Magnus Lodefalk.

Report and Policy Brief 2015:5, *Svenskt medborgarskap: reglering och förändring i ett skandinaviskt perspektiv*, Mikael Spång.

Report and Policy Brief 2015:6, *Vem blir medborgare och vad händer sen? Naturalisering i Danmark, Norge och Sverige*, Pieter Bevelander, Jonas Helgertz, Bernt Bratsberg and Anna Tegunimataka.

Research Overview 2015:7, *Kategoriernas dilemman*, Per Strömblad and Gunnar Myrberg.

Report and Policy Brief 2015:8, *Valet och Vägen: Syriska flyktingar i Sverige*, Emma Jörum Lundgren.

Report and Policy Brief 2015:9, *Arbetskraftsinvandring efter 2008 års reform*, Catharina Calleman (red.) and Petra Herzfeld Olsson (red.).

Research Overview 2016:1, *Alla tiders migration!* Dick Harrison.

Report and Policy Brief 2016:2, *Invandringens arbetsmarknadseffekter*, Mattias Engdahl.

Report and Policy Brief 2016:3, *Irreguljär migration och Europas gränskontroller*, Ruben Andersson.

Research Overview 2016:4, *Diaspora – ett begrepp i utveckling*, Erik Olsson.

Research Overview 2016:5, *Migration within and from Africa*, Aderanti Adepoju.

Report and Policy Brief 2016:6, *Invandring, mediebilder och radikala högerpopulistiska partier i Norden*, Anders Hellström and Anna-Lena Lodenius.

Research Overview 2016:7, *Invandring och företagande*, Martin Klinthäll, Craig Mitchell, Tobias Schölin, Zoran Slavnić and Susanne Urban.

Report and Policy Brief 2016:8, *Invandringens effekter på Sveriges ekonomiska utveckling*, Bo Malmberg, Thomas Wimark, Jani Turunen and Linn Axelsson.

Research Overview 2017:1, *De invandringskritiska partiernas politiska inflytande i Europa*, Maria Tyrberg and Carl Dahlström.

Research Overview 2017:2, *Hatbrott med främlingsfientliga och rasistiska motiv*, Berit Wigerfelt and Anders S Wigerfelt.

Dissertation Series 2017:3, *Vägen till arbete. Utlandsföddas möte med den svenska arbetsmarknaden*, Moa Bursell, Mikael Hellström, Jennie K Larsson, Melissa Kelly, Martin Qvist and Caroline Tovatt.

Policy Brief 2017:4, *Integration och tillit – långsiktiga konsekvenser av den stora invandringen till Norge*, Grete Brochmann.

Research Overview 2017:5, *Invandringens historia – från "folkhemmet" till dagens Sverige*, Mikael Byström and Pär Frohnert.

Report and Policy Brief 2017:6, *Invandring i medierna – Hur rapporterade svenska tidningar åren 2010-2015?*, Jesper Strömbäck, Felicia Andersson and Evelina Nedlund.

Report and Policy Brief 2017:7, *Valdeltagande och representation – Om invandring och politisk integration i Sverige*, Pieter Bevelander (red.) and Mikael Spång (red.).

Report and Policy Brief 2017:8, *Responsibility Sharing for Refugees in the Middle East and North Africa*, Susan Martin.

Report and Policy Brief 2017:9, *Reforming the Common European Asylum System*, Bernd Parusel and Jan Schneider.

Report and Policy Brief 2017:10, *A Fair Share: Refugees and Responsibility-Sharing*, Alexander Betts, Cathryn Costello and Natascha Zaun.

Report and Policy Brief 2018:1, *Somali Diaspora Groups in Sweden – Engagement in Development and Relief Work in the Horn of Africa*, Nauja Kleist.

Report and Policy Brief 2018:2, *Akademiskt utbyte och internationell migration – En studie av stipendiater inom Svenska institutets Visbyprogram 1997–2015*, Andreas Åkerlund, Astrid Collsiöö and Mikael Börjesson.

Report and Policy Brief 2018:3, *Ensamkommande barns och ungas väg in i det svenska samhället*, Eskil Wadensjö and Aycan Çelikaksoy.

Report and Policy Brief 2018:4, *Attityder till invandring – en analys av förändringar och medieeffekter i Sverige 2014–2016*, Jesper Strömbäck and Nora Theorin.

Report and Policy Brief 2018:5, *Familj, medborgarskap, migration – Sveriges politik för anhöriginvandring i ett jämförande perspektiv*, Karin Borevi.

Dissertation Series 2018:6, *Barn och migration*, Mehek Muftee, Lisa Ottosson, Gunilla Jarkman Björn, Anna Åhlund, Eva Skowronski and Michael Lindblad.

Policy Brief 2018:7, *Människohandel och människosmuggling i den irreguljära migrationen*, Ryszard Piotrowicz.

Report and Policy Brief 2018:8, *Asylsökandes möte med Sverige – Lärdomar från en panelundersökning*, Peter Esaiasson and Jacob Sohlberg.

Policy Brief 2018:9, *Medborgarskapslagar – en global jämförelse*, Rainer Bauböck.

Report and Policy Brief 2019:1, *Bridging the Gaps – Linking Research to Public Debates and Policy-making on Migration and integration*, Martin Ruhs, Kristof Tamas and Joakim Palme.

Report and Policy Brief 2019:2, *Från Afrikas horn till Sverige: Smuggling, informella nätverk och diasporans engagemang*, Tekalign Ayalew Mengiste and Erik Olsson.

Report and Policy Brief 2019:3, *Thai berry pickers in Sweden – A migration corridor to a low-wage sector*, Charlotta Hedberg, Linn Axelsson and Manolo Abella.

Report and Policy Brief 2019:4, *Internationella studenter i Sverige – Avgiftsreformens påverkan på inflödet av studenter*, André Bryntesson and Mikael Börjesson.

Research Overview and Policy Brief 2019:5, *Migration and development – The role for development aid*, Robert E.B. Lucas.

Policy Brief 2019:6, *Åter till grunderna – Läsförmåga, immigration och arbetsmarknadsutfall i Sverige*, Jon Kristian Pareliussen.

Policy Brief 2019:7, *Ålder vid invandring och arbetsmarknadsintegration – det svenska exemplet*, Torun Österberg.

Policy Brief 2019:8, *Barn med posttraumatisk stress – utvärdering av en gruppintervention för ensamkommande flyktingbarn med symptom på posttraumatisk stress*, Anna Sarkadi.

Policy Brief 2019:9, *Suicidalt beteende och vård – skillnader mellan flyktingar, andra migranter och personer födda i Sverige*, Ellenor Mittendorfer-Rutz.

Policy Brief 2019:10, *Fri rörlighet för arbetstagare i EU och dess effekter på statsfinanserna*, Marcus Österman, Joakim Palme and Martin Ruhs.

Report and Policy Brief 2020:1, *De som inte får stanna: Att implementera återvändandepolitik*, Henrik Malm Lindberg.

Report and Policy Brief 2020:2, *Laglig migration för arbete och studier - Möjligheter att få uppehållstillstånd i Sverige för personer som saknar skyddsbehov*, Bernd Parusel.

Report and Policy Brief 2020:3, *Effekten av krig – Posttraumatisk stress och social tillit hos flyktingar*, Jonathan Hall and Dennis Kahn.

Report and Policy Brief 2020:4, *Åtgärder mot människosmuggling och människohandel: Är de förenliga med EU-stadgan om de grundläggande rättigheterna?*, Vladislava Stoyanova.

Policy Brief 2020:5, *Den reglerade invandringen och barnets bästa*, Louise Dane.

Policy Brief 2020:6, *Migrationspolitik, välfärd och jämlikhet*, Björn Östbring.

Dissertation Series 2020:7, *Migranters möte med svensk hälso- och sjukvård*, Juliet Aweko, Ulrika Byrskog, Annika Esscher, Robert Jonzon and Josefin Wångdahl.

Policy Brief 2020:8, *Språkkaféet som arena för språkträning*, Gunilla Jansson and Silvia Kunitz.

Policy Brief 2020:9, *Nyanlända elever och lärande på gymnasieskolans språkintroduktionsprogram*, Päivi Juvonen.

Report and Policy Brief 2021:1, *Idrott och hälsa bland flickor - Uppfattningar och erfarenheter bland föräldrar från Somalia, Eritrea, Syrien och Sverige*, Anders Kassman and Åsa Kneck.

Policy Brief 2021:2, *Miljonprogram, migranter och utsatthet för covid-19*, Erik Hansson, Lina Al-Nahar, Maria Albin, Eskil Jakobsson, Maria Magnusson and Kristina Jakobsson.

Report and Policy Brief 2021:3, *Lokalsamhälletillit i Sverige före och efter flyktingkrisen*, Susanne Wallman Lundåsen.

Research Overview 2021:4, *Tolkfunktionen i asylprocessen*, Cecilia Wadensjö, Hanna Sofia Rehnberg and Zoe Nikolaidou.

Dissertation Series 2021:5, *Tidsbegränsade uppehållstillstånd, egenföretagande och skolsegregation – Aktuella avhandlingar om utrikes födda på arbetsmarknaden*, Kristoffer Jutvik, Matilda Kilström, Elisabet Olme, Aliaksei Kazlou and Debbie Lau.

Report 2021:6, *Ungas uppväxtvillkor och integration*, Jan O. Jonsson, Susanne Åsman, Thomas Johansson, Ylva Odenbring, Sevgi Bayram Özdemir, Erik Lundberg, Metin Özdemir, Bo Malmberg and Eva Andersson.

Policy Brief 2021:7, *Invandring och välfärdsstaten*, Tina Goldschmidt.

Policy Brief 2021:8, *Interaktiv rasism på internet, i pressen och politiken*, Mattias Ekman.

Policy Brief 2021:9, *Polariserad demokrati. Hur uppfattade hot påverkar främlings-fientlighet och spänningar mellan olika grupper i samhället*, Emma A. Renström and Hanna Bäck.

Report and Policy Brief 2021:10, *De som skickades tillbaka: Återvändande och återintegration av avvisade asylsökande till Afghanistan och Irak*, Henrik Malm Lindberg, Constanza Vera-Larrucea and André Asplund.

Report and Policy Brief 2021:11, *Trade Agreements as a Venue for Migration Governance? Potential and Challenges for the European Union*, Sandra Lavenex and Tamirace Fakhoury.

Policy Brief 2021:12, *The evolving policy coherence for development. Risk or opportunity for the EU as a development actor?*, Anna Michalski.

Dissertation Series 2021:13, *Nya perspektiv på segregation: skola, psykisk hälsa och bosättningsmönster*, Maria Granvik Saminathen, Sara Brolin Låftman, Petra Löfstedt, Elisabet Olme, Dany Kessel, Louisa Vogiazides, Hernan Mondani, Charisse Johnson-Singh, Mikael Rostila, Antonio Ponce de Leon, Yvonne Forsell and Karin Engström.

Research Overview 2022:1, *Exillitteratur i Sverige 1969 till 2019*, Daniel Pedersen.

Research Overview 2022:2, *The impacts of migration for adaptation and vulnerability,* François Gemenne.

Policy Brief 2022:3, *How large will the Ukrainian refugee flow be, and which EU countries will they seek refuge in?* Mikael Elinder, Oscar Erixson and Olle Hammar.

Policy Brief 2022:4, *The Temporary Protection Directive: EU's response to the Ukrainian exodus: The 'why', 'who', 'what', 'where' and 'how' of temporary protection.* Eleni Karageorgiou and Vladislava Stoyanova.

Report 2022:5, *Invandring och integration i svensk opinion. Hur formas värderingar och verklighetsuppfattningar?*, Anders Westholm.

Report 2022:6, *Hur gemensamma är de gemensamma EU-reglerna? Svensk asylrätt i europeiskt perspektiv*:, Eleni Karageorgiou and Vladislava Stoyanova.

Policy Brief 2022:7, *Vilka är mest benägna att rösta i svenska lokala val?*, Pieter Bevelander.

Report 2022:8, *Nyanländas integration: En enkätstudie om språkstuderandes erfarenheter av livet i Sverige*, Sara Thalberg and Linus Liljeberg.

Research Overview 2022:9, *Climate Change, Displacement, Mobility and Migration The State of Evidence, Future Scenarios, Policy Options*, Mathias Czaika and Rainer Münz.

Policy Brief 2022:10, *(Kommunala) insatser för att underlätta arbetsmarknadsinträdet för flyktingar och deras anhöriga*, Mattias Engdahl, Anders Forslund and Ulrika Vikman.

Policy Brief 2023:1, *Att förstå klyftan mellan politiken för flyktingintegration och erfarenheterna av integration: Resultat från två EU-finansierade projekt, FOCUS och NIEM*, Nahikari Irastorza and Sayaka Osanami Törngren.

Research Overview 2023:2, *Migration, religion och integration*, Magdalena Nordin.

Policy Brief 2023:3, *Lokal Migrationspolitik: Om strukturer, aktörer och processer i svenska kommuner*, Jon Nyhlén and Gustav Lidén.

Report 2023:4, *Under ytan: Hur många och vilka vill lämna sina länder för att flytta till EU och Sverige?*, Mikael Elinder, Oscar Erixson and Olle Hammar.

Dissertation Series 2023:5, *Migration, kontroll och säkerhet i en europeisk kontext*. Louise Bengtsson, Jennie Brandén, Daniel Silberstein and Teresa Quintel.

Policy Brief 2023:6, *Laglig migration för arbete och studier. Möjligheter att få uppehållstillstånd i Sverige för personer som saknar skyddsbehov*, Bernd Parusel.

Policy Brief 2023:7, *Integrationsprocesser och traditionsförhandlingar inom religiösa samfund*, Magdalena Nordin.

Policy Brief 2023:8, *Ökar röstande aptiten på medborgarskap?*, Pieter Bevelander, Michaela Slotwinski and Alois Stutzer.

Report and Policy Brief 2023:9, *Etablering av nyanlända flyktingar: Effekter av en tidig och intensiv arbetsmarknadspolitisk insats*, Matz Dahlberg, Johan Egebark and Ulrika Vikman.

Dissertation Series 2023:10, *Hinder och möjligheter för etablering på den svenska arbetsmarknaden*, Andrey Tibajev, Anni Erlandsson, Lillit Ottosson and Louisa Vogiazides.

Policy Brief 2023:11, *Integration av unga i Sverige i ett flerdimensionellt perspektiv*, Jan O. Jonsson.

Policy Brief 2023:12, *Civilsamhällets bidrag i integration i bostadsområden med socioekonomiska utmaningar*, Gabriella Elgenius.

Policy Brief 2023:13, *Den reglerade invandringen och barnets bästa: en uppdatering efter Tidöavtalet*, Louise Dane.

Research Overview 2024:1, *Återvändande och (åter) integration för barn i familj*, Pinar Aslan Akay.

Research Overview 2024:2, *Det andra utanförskapet: Skuggsamhället i Sverige*, Zenia Hellgren.

Report and Policy Brief 2024:3, *Migrationslagstiftning i Norden – Ett restriktivt skifte*, Jonas Hinnfors and Ann-Chatrine Jungar.

Policy Brief 2024:4, *Hur formas den svenska migrationspolitiken?*, Henrik Malm Lindberg.

Report 2024:5, *Migranters välbefinnande och befolkningens attityder till migration. Sverige i ett europeiskt perspektiv*, Maria Cheung, Matz Dahlberg and Hyrije Hasani.

Report and Policy Brief 2024:6, *Arbetskraftsinvandring till Sverige från tredjeland*, Mattias Engdahl and Erik Sjödin.

Policy Brief 2024:7, *På flykt från Rysslands anfallskrig i det transnationella Europa. Mottagandet av skyddssökande från Ukraina i Sverige*, Oksana Shmulyar Gréen and Svitlana Odynets.

Report and Policy Brief 2024:8, *Återvändandets diplomati: Om samverkan mellan Sverige och diplomatiska beskickningar i återvändande och återtagande*, Constanza Vera-Larrucea and Iris Luthman.

Policy Brief 2024:9, *Understanding public opinions towards integration. Where does Sweden stand?*, Nahikari Irastorza.

Policy Brief 2024:10, *Likgiltiga platser och platser för gemenskap. Nyanlända flyktingars relationsskapande i små- och medelstora städer*, Måns Lundstedt.

Report 2024:11, *War Policies and Migration Aspirations in Russia,* Mikael Elinder, Oscar Erixson and Olle Hammar.

Policy Brief 2024:12, *Att väga och värdera invandrares levnadssätt: den svenska vandelsprövningen historiskt och i jämförelse med andra länder*, Andreas Asplén Lundstedt.

Policy Brief 2024:13, *Security of Residence for Refugees and Migrants in Sweden after the Tidö Agreement,* Elspeth Guild.

Research Overview 2024:14, *Etnisk diskriminering i rekryteringsprocesser – orsaker och åtgärder*, Pinar Aslan Akay and Maria Cheung.

Policy Brief 2024:15, *The European Trust Fund for Africa (EUTF) and the externalization of migration control*, Joseph Trawicki Anderson.

Policy Brief 2024:16, *Hälsa bland förvarstagna och alternativ till förvar,* Andreas Savelli.

Research Overview 2024:17, *Hur formas den svenska migrationspolitiken?,* Henrik Malm Lindberg.

Policy Brief 2025:1, *EUAA och Sveriges implementering av EU:s nya asyl- och migrationspakt,* Johan Ekstedt.

Report 2025:2, *Nordic cooperation within return and readmission*, Anna Hammarstedt and Iris Luthman.

Research Overview 2025:3, *Den byggda miljöns betydelse för att minska boendesegregation och ojämlika livsvillkor*, Ann Legeby and Lars Marcus

Dissertation Series 2025:4, *Tid, rum, fördomar och uppfattningar i migrationspolitiken – från EU:s externaliserade gränser till utsatthet på den svenska bostadsmarknaden,* Rikard Engblom, Kristina Anna Wejstål, Andreas Asplén Lundstedt and Mikaela Herbert.

Research Overview 2025:5 *Insatser för ökat valdeltagande i bostadsområden med socioekonomiska utmaningar*, Karl-Oskar Lindgren.

Policy Brief 2025:6 *Samordning av återintegrationsinsatser i unika sammanhang – en fallstudie om nordiskt samarbete i den kurdiska regionen i Irak (KR-I)*, Anna Hammarstedt, Iris Luthman and Jenny Bergsten.

Policy Brief 2025:7 *Stöd för psykisk hälsa och integration: Långsiktiga lösningar för ukrainare som flytt till Sverige*, Viktoriia Svidovska and Yaroslava Shven.

Policy Brief 2025:8 *Kommuners förändrade roller i arbetsmarknadsintegration*, Patrik Zapata, María José Zapata Campos and Emma Ek Österberg.

Research Overview och Policy Brief 2025:9, *The Role of European Host Countries in Voluntary Return Migration. A Systematic Review of the Evidence*, Andrea Voyer, Klara Nelin and Alice Zethraeus.

Report and Policy Brief 2025:10, *Hälsa och migrationsvilja i en global kontext*, Mikael Elinder and Oscar Erixson.

Report and Policy Brief 2025:11, *Frontex roll i återvändandeoperationer – Perspektiv från Sverige*, Daniel Silberstein, Suzanne Planchard and Henrik Malm Lindberg.

Artificial intelligence (AI) is reshaping migration, asylum and border policies by enabling automated decision-making. While AI promises efficiency, its use in asylum procedures raises serious human rights concerns, particularly because affected individuals are often in vulnerable positions. The EU AI Act classifies use of AI in asylum procedures as ”high risk”.

This report examines whether AI-supported decision-making in asylum cases aligns with human rights law, especially the rights to privacy and protection from refoulement. Key concerns include the possible procedural harm, since AI may compromise fairness, transparency and the involvement of applicants; privacy, since AI use must meet legal standards of legality and necessity amongst others. And, last but not least, proportionality. Even if AI complies with the EU AI Act, it may still violate human rights if the risks outweigh the benefits.

The author of this Delmi-report is Vladislava Stoyanova, Associate Professor of Public International Law at the Faculty of Law, Lund University.

The Migration Studies Delegation is an independent committee that initiates studies and supplies research results as a basis for future migration policy decisions and contribute to public debate.

Delmi

STATENS OFFENTLIGA UTREDNINGAR

Delegationen för Migrationsstudier Ju 2013:17