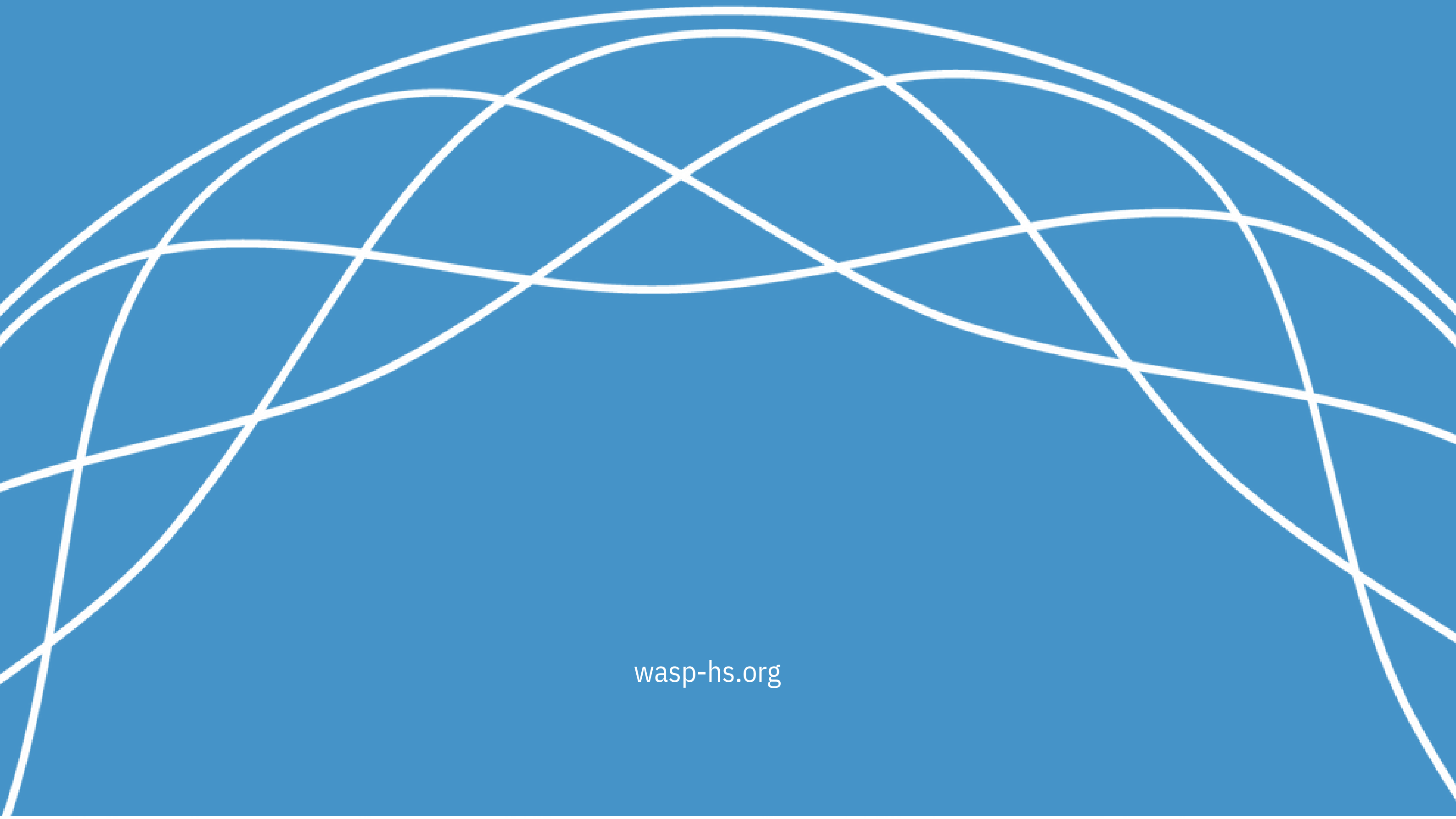
WASP—HS

COMMUNITY REFERENCE MEETING: CHALLENGES AND OPPORTUNITIES OF REGULATING AI

REPORT

August 2022



wasp-hs.org

Introduction

The development of regulatory frameworks to govern the development, design, and application of artificial intelligence is recently receiving much attention. In April 2021, the EU Parliament published a proposal, the AI Act (AIA), for the purpose of regulating the use of AI systems and services in the Union market. This proposal brings forward a regulatory vision based on European standards on human rights, democracy, and the rule of law. However, the effects of EU digital regulations usually transcend its confines. An example is the General Data Protection Regulation (GDPR), which rapidly became a world standard. The extraterritorial scope of AI should be analysed in the face of other AI governance models currently under development. The AIA adopts a risk-based approach that bans certain technologies, proposes strict regulations for "high-risk" ones, and imposes stringent transparency criteria for others. If adopted, the AIA will undoubtedly have a significant impact in the EU and beyond. A crucial question is whether we already have the technology to comply with the proposed regulation, and to what extent the requirements of this regulation can be enforceable.

For most of the past decade, public concerns about AI, and digital technology in general, have focused on the potential abuse of personal data. Data privacy and security are main drivers of the discussion on the ethical and societal impact of these technologies and have led to many efforts at regulation and legislation, the most influential being the EU's GDPR. However, the focus on data privacy is now shifting towards more comprehensive approaches that include identifying, interpreting and balancing different societal requirements such as inclusion, access, bias, explanation, and transparency, to name a few. As such, regulation of AI is increasingly based on the assessment of risks and opportunities, also including the risk of not using AI, and on the specification of trade-offs. In this meeting, we discussed how regulation would shape the AI technologies of the future and examined the interplay between national policies and the work of other organisations, by bringing together input and discussions from multidisciplinary stakeholders. The main takes from the meeting include:

- Without a clear definition of what AI regulation is aiming to regulate, legislation may be void, and hard to implement. The articulation of requirements and functionalities of the technology being regulated is fundamental.
- The (EU) legislation space in this area is becoming very 'crowded'. It is important to understand how to navigate, integrate and balance different regulatory directions.
- Participation and inclusion of all relevant stakeholders are crucial from the very onset of any design and implementation effort, be it regulation or digital technology.

> “This is a challenge that requires a multidisciplinary approach: law, technology and society experts must work together.”

WASP-HS Community Reference Meetings (CRMs)
CRMs are aimed at helping public and private organizations in Sweden with challenges and questions regarding their interests, as well as developments within WASP-HS. This is done to identify opportunities for collaboration between different sectors.

CHALLENGES AND OPPORTUNITIES OF REGULATING AI

Abstract Keynote Speech Catelijne Muller, LL.M

The European Union has been on a path towards regulating AI since 2018 when it presented its "AI for Europe Strategy" based on 3 pillars: (i) boost innovation, (ii) prepare for social-economic changes and (iii) ensure an appropriate ethical and legal framework. The third pillar has been pursued through the work of the EU High Level Expert Group on AI, that developed the Ethics Guidelines for Trustworthy AI, laying the groundwork for the proposal for an AI Regulation presented by the European Commission in 2021 (the "AI Act").

AI Act – Scope and Objective

The AI Act is broad in scope as it covers virtually all AI systems and domains and applies to both EU and global actors that want to introduce AI on the EU market. Both in the AI Act as well as in proposals of the co-legislators (EP and Council) however, several (temporary or full) exclusions from the AI Act are mentioned, such as existing high-risk AI systems, border control AI (temporary), general purpose AI, AI R&D, AI for national security. Many of these exclusions would however provide major loopholes and water down the protection of the AI Act considerably.

The objective of the AI Act is to protect people's health, safety and fundamental rights against the ill effects of AI and promoting trustworthy AI innovation. AI can have an impact on multiple fundamental rights[1], clustered into 4 families:

- Human dignity (integrity, privacy, fair trial)
- Freedom (expression, information, assembly)
- Fairness (non-discrimination, equal treatment)
- Social (education, healthy workplace, social benefits)

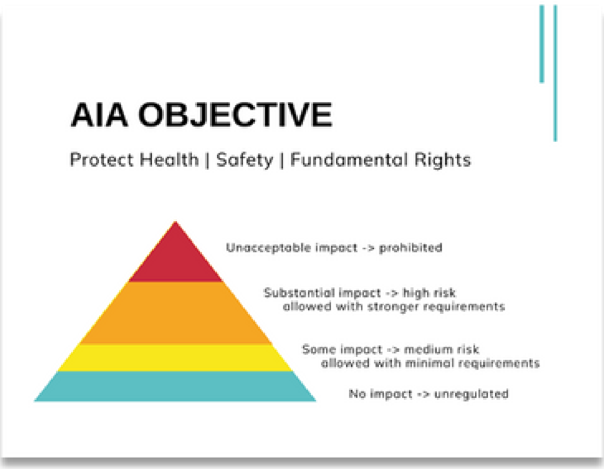
Risk pyramid

The protection of these rights is structured in the AI Act as a 'risk pyramid'. The higher the risk a certain AI system poses, the stricter the conditions and requirements for such a system and its developers will be. Some AI practices are even deemed to be too risky and prohibited.

Prohibitions

The AI Act prohibits only a very limited set of AI practices:

- Limited and rare cases of AI-driven manipulation
- Social scoring by public parties
- Biometric identification by or for law enforcement, but only the type that takes place remotely, 'in real time' and in publicly accessible spaces



AIA OBJECTIVE

Protect Health | Safety | Fundamental Rights

Unacceptable impact -> prohibited

Substantial impact -> high risk
allowed with stronger requirements

Some impact -> medium risk
allowed with minimal requirements

No impact -> unregulated

High-risk AI

The AI Act considers a large number of AI-practices high-risk, such as all other forms of biometric identification and categorization, AI that determines or informs access to education, work, essential private services and public services, AI used in law enforcement, migration, asylum, border control, the judiciary and democracy. Apart from these, all AI-systems that are part of products that are already regulated at the EU level (such as medical devices, toys, lifts etc.) are considered high-risk as well.

These AI-systems/practices are allowed, but they (and/or their developers/providers) need to meet a great number of requirements before they can be introduced into the EU internal market, such as a risk management system, high-quality data and proper data governance, technical documentation, record keeping and transparency of information, human oversight, accuracy, robustness and cybersecurity, critical events monitoring. Non-compliance with the requirements or the use of prohibited AI can result in hefty fines.

Legislative process

The legislative process is currently in the phase where the co-legislators (EP and Council) are forming their respective positions. These positions are expected to be finalized in early 2023, after which the so-called ‘Trilogue’, the negotiations between the EC, EP and Council, will start. These negotiations will likely last 1 to 1.5 years. If an agreement is reached in Trilogue, the AI Act comes into force shortly after that and after a transition period of 24 months becomes directly applicable in all EU Member States.

Catelijne Muller, LL.M
President ALLAI
Inquiries: welkom@allai.nl

References:

[1] Muller, C. The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law, CAHAI, Council of Europe, 2019

The AI Act – Comprehensive but is it Future-Proof?

Liane Colonna, Stockholm University; Cecilia Magnusson Sjöberg, Stockholm University

Main Challenges

- Legislators are faced with the difficult task of regulating the AI of the future as well as the AI that is currently available. This forces them to consider a technology neutral approach to regulation to "future proof" the legislation but this can lead to overregulation, threaten innovation and undermine the rule of law.
- It is difficult to achieve consistency in the myriad of laws applicable to AI that exist on the national, European, and international level: it is increasingly unclear which regulatory act has precedence over another one.
- It is challenging to reconcile a risk-based, product-safety approach to regulation which is focused on technical fixes like data quality, documentation, and auditability with a more rights-based approach that engages the individuals and communities that are impacted by the AI practices.
- The lack of semantic management and clear definitions that cut across disciplines creates legal obstacles.

For many years now, there have been intense discussions about whether AI needs specific regulation and, if so, what this regulation should look like. There is no doubt that the regulation of AI is a highly complex task and there are many difficult choices that must be made by legislators. Should a hard or soft law approach be taken? Is a rights-based or a risk-based approach more appropriate? How to avoid a situation where too much regulation hinders innovation? Should the law be technology-specific or technology neutral? Should there be a sectoral or a general approach to regulation? At the EU level, is a directive or a regulation a better legal instrument to apply? Should a control or a social protection model of governance be applied?

Ultimately, the AI Act takes both a horizontal (cross-sectoral) and vertical (sector-specific) approach. On one hand, it is an overarching legal framework that applies to all AI and, at least currently, defines AI broadly. On the other hand, it is focused sector-based high-risk applications. It also builds off existing sector-specific laws like medical device regulation and toy regulation.

How to precisely define "AI" is an extremely challenging task from a legal perspective, if at all feasible. Some commentators warn that it should not matter if it is"AI" or just some less advanced form of automation if the consequences of the computational system can have a devastating impact on fundamental rights, health, or safety of a person. If the government makes some incorrect or biased algorithmic decision that results in an individual needing to repay all her benefits and this sends her into poverty and she loses her child, should it matter if it was AI per se or some lesser form of computational technique? Commentators in favor of a broad definition of AI argue that it makes more sense to focus on the domain where the technology is used and the risks involved, rather than focus on a specific computational technique. Other commentators consider that a broad definition of AI will lead to legal uncertainty for developers, operators, and users of AI systems and ultimately to over-regulation. They insist that it is necessary to have a narrow and precise definition of AI in order to support innovation and legal certainty.

There are, as already mentioned, many different laws that are applicable to AI including, for example, national laws, international treaties, EU law and soft law instruments. It is clear that the AI Act is just one piece of a much larger puzzle. Besides existing legal frameworks (e.g., GDPR, NIS 1, the Product Liability Directive), there are also many new EU law proposals (e.g., NIS 2, Digital Markets Act, the European Health Data Space) and proposals that are expected to come soon (e.g., rules to address liability issues related to new technologies, including AI systems). The legal landscape surrounding AI is highly complex and there appears to be some confusing or conflicting rules. For example, medical technologies can be assigned to a range of risk classes under the Medical Device Regulation, but most will probably be classified as high-risk AI under the AI Act.

It is also challenging to understand how to reconcile the GDPR principle of data minimization with the strict data requirements e.g., robustness and accuracy in the AI Act. The AI Act should be fully aligned with existing legal frameworks to avoid inconsistency, fragmentation, and duplication of requirements.

The AI Act proposal takes a risk-based, product safety approach to regulation that is very technocratic, focusing on technical fixes like data quality, technical documentation, auditability. Many are worried that it fails to engage the individuals and communities that are impacted by the AI practices. The proposal also gives a large role to two private standardization organizations, and it can be questioned whether their roles are disproportionately large, especially when compared to the role of individuals.

The governance structure of the proposed AI Act involves a European as well as a national level. With many different actors in the space working in different locations, times frames and with possibly different information, it is easy to imagine a lack of coordination, particularly where it concerns the monitoring of risk. There is also a question about whether there is sufficient expertise and resources for monitoring and assessing at the national level.

The Participation Paradox in the Politics of AI

Michael Strange, Malmö University; Jason Tucker, Malmö University;
Jess Haynie-Lavelle, Malmö University; Dennis, Munetsi, Malmö University

Main Challenges

- AI systems are increasingly being used to shift decisions made by humans over to automated systems, potentially limiting the space for democratic participation. The risk that AI erodes democracy is exacerbated where most people are excluded from the ownership and production of AI technologies that will impact them.
- AI learns through datasets but, very often, that data excludes key parts of the population. Where marginalized groups are considered, datasets often contain derogatory terms, or exclude explanatory contextual information, that is hard to accurately categorise in a format that AI can process. Resulting biases within AI design raise concerns as to the quality and representativeness of AI-based decisions and their impact on society.
- There is very little two-way communication between the developers and users of AI-technologies such that the latter function only as personal data providers. Being largely excluded from the development of AI's role in human decision-making, everyday individuals may feel more marginalized and disinterested in building a healthy and sustainable society.
- Yet, AI's capacity for seeing patterns in big data provides new ways to reach parts of the population excluded from traditional policymaking. It can serve to identify structural discrimination and include information from those otherwise ignored in important decisions. AI could enhance public participation by both providing decision-makers with better data and helping to communicate complex decisions – and their consequences – to wider parts of the population.

The roundtable brought together a diverse and global range of experts from thinktanks, civil society, the private sector, and academia to discuss the challenges and opportunities for participation arising from the increasingly present and pervasive role of AI technologies within societal decision-making. The below points summarise our key conclusions from the discussion.

Democracy requires space for 'friction', that AI risks closing. Not all societal values are mutually supportive (e.g., freedom vs justice) in a way that AI can easily balance. While AI promises positive contributions, such as consistency and fairness, there needs to be space for genuine contestation. Overreliance on AI machine-learning – unless better designed to create meaningful processes of participation – will hollow out the decision-making process. Treating political decisions as purely technical has been proven to create apathy and resentment amongst those feeling 'left out'. The development of AI as a tool for public decision-making should only take place in tandem with new safeguards for participation rights and other instruments that foster democratic consultation. Human dignity needs to be a guiding principle for the development of AI.

Knowledge production and dissemination in AI needs to be more collaborative and redistributed with currently dominant actors serving as 'community connectors' in building multi-stakeholder partnerships throughout the development of AI systems from design through to usage. These partnerships should substantively represent a variety of actors who collaboratively become an integral part of building and maintaining AI systems. A diverse global population needs to be visible in not only datasets but also actively engaged in the design of how data is collected and utilized.

AI Machine-learning can make existing decisions more inclusive and transparent, allowing humans a better overview of systemic biases and other problems. In the example of participatory approaches towards AI usage in healthcare, data collection in real-time during interactions of patients and healthcare providers could enhance the quality of data and the participation of marginalised groups with limited or no access to other means of data collection due to various confounding social factors. Participants should be informed of data collection purposes and consent before participating in this alternative scenario.

Participatory AI requires better communication and education. Diverse actors need to be part of creating research objectives, and translating AI products and technical terminology, to be accessible and relevant for ordinary users. Representation will mean that decisions and actions taken reflect the collective efforts of all involved. There is a risk of ‘invisible’ exclusions due to a lack of communication between technology designers and the everyday individual. The disconnect has practical implications as interventions are often rejected in any context where developers and users are divided, undermining the potential of AI to bring positive change. Knowledge gaps can also hinder the equal participation of people without technical expertise. Capacity building through workshops in target communities can potentially enhance non-technical people’s understanding of AI interventions and their impact as well as ensure designers better appreciate their social impact.

Digital literacy needs to be combined with political and societal literacy. There is a lot of research on AI and its societal impact, but very little of this is read by the global public. Most of those who will be significantly impacted by AI are not actively debating its implementation. AI can bring many benefits, but only if its design involves people with the experience and knowledge of what works best for the local society. Even amongst educated persons designing AI systems, it is unusual that they are well trained in the requirements for sustainable societies. We all need to find new ways to ensure people are literate in the intersecting fields relevant to AI’s societal role. Only then can AI achieve its potential to be a force for societal good rather than exacerbate existing and worsening tensions.

Regulating the Use of Algorithms in Public Decision-Making

Sandra Friberg, Uppsala University; Yulia Razmetaeva, Yaroslav Mudryi National Law University/ Uppsala University; Natalia Filatova-Bilous, Yaroslav Mudryi National Law University/NGO Civil Law Platform

Main Challenges

- How to regulate the use of AI in public decision making.
- Placing responsibility for the implementation and deployment of algorithms in public decision making and for wrong and/or unfair decisions.
- Articulating requirements for AI-systems involved in public decision making, which includes requirements to be met by their creators as well as the procuring public authorities.
- Articulating requirements for the decision-making processes where AI is involved to maintain legal certainty.
- The relationship between private and public sectors in the implementation, deployment, and maintenance of AI-systems in public decision making - possible and necessary approaches.

The use of algorithmic decisions is growing in all areas, and this is especially significant in the context of public decision-making. Problems in applying algorithms to public decision-making are party based on concerns that AI can be (1) biased, (2) AI tend to highlight patterns, which can lead to exception cases being outlied, (3) algorithm decisions can be poorly explained (4) algorithms may be owned by private companies that refuse to disclose the details of AI decision making.

Regulating AI with an all-encompassing act like Regulation on Artificial Intelligence can be promising. At the same time, there are already several acts on AI that are in force in the EU, as well as in the national legislation of Member States. Consideration should be given to how the EU Member States already regulate the use of algorithms in public decision-making. In doing so, special consideration must be made to the interconnectivity to regulations on data protection and the protected interests and rights of individuals.

It is necessary to distinguish between the types of AI systems used for decision-making since not all "smart systems" involve the use of AI. Thus, it is worth returning to the potential definition of AI, what an algorithm and the use of algorithms in making public decisions mean. Individual public officials can be more or less operative in decision-making depending on the type of algorithmic or AI support system in use.

A question of high importance is whether we are looking for the subject behind the algorithm in all types of AI systems. Who is the government, the developers, the decision makers to include AI in decision making? Is the algorithm a tool and how trustworthy is it?

As an example, it was highlighted during the discussions that a police officer who decides on a case involving non-native speakers must trust a colleague interpreter, but it is still the police officer who makes the decision and is responsible for it. How much one should rely on algorithms in a similar situation in the future is an issue for debates. In this context, it is necessary to discuss how far, and to what extent public authorities can be responsible for articulating the tasks that an algorithm or an AI is supposed to fulfil. A developer must have a clearly defined assignment to develop a functional system, which in turn will entail potential liability in relation to the procuring authority. It is not difficult to envisage potential legal conflicts concerning a faulty system and the challenges in trying to avoid unclarity in this respect.

The use of AI that leads to unfair decisions towards individuals may require clear and easily accessible compensation mechanisms. During the discussions, it was pointed out that the way we should address the issue of liability depends on many factors. For example, if the decision is wrong and causing damages to a person, the state should be made liable towards the person and, afterwards, the state may put a subrogation or regression claim towards a front-end and (or) back-end operator of the AI-system used in the decision-making process. However, if the decision is wrong but beneficial for individuals, it is the State who suffers, e.g. in giving out subsidies although the requirements are not met. In those cases, the State should perhaps have an opportunity to lodge claims towards a front-end and(or) a back-end operator of the AI-system. On the other hand, it could instead be viewed as a responsibility of the State that the risk of overcompensation is not considered in the implementation of the algorithm.

Within public decision-making, there are also requirements of procedure and transparency: individuals must be able to understand how a decision is being made, and said reasoning behind the decision. It was pointed out that it – as a minimum – should be explained to people that in the individual decision-making process, an AI system has been used and to what extent the use of the AI influenced the final decision. This of course ties in with the rights for individuals to appeal the decision and perhaps claim damages for a faulty decision.

We also need to understand, and address, the issue of which public decisions can, and which cannot be made with the use of AI. This is the most challenging issue for regulators since it is difficult to clarify, in the AI Act, which kind of usage is allowed, and which kind of usage is forbidden. In the discussions, it also became evident that there is a need to explore how AI is affecting public decision-making in hidden ways. In particular, the introduction and implementation of algorithms may affect people in such a way that they tend to argue less and raise complex questions, but more to look for mathematically accurate decision. Looking to the future suggests that algorithmic public decision-making will reinforce algorithmic thinking. It is important to be aware of any risk of losing important public discussions and participation in public decision-making in the case of using algorithms since this of fundamental value in a democratic society.

WASP–HS

The vision of the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) is to realize excellent research and develop competence on the opportunities and challenges of artificial intelligence and autonomous systems with a strong investment in research in humanities and social science.

The WASP-HS program is planned to run 2019 – 2028 and will form an independent and parallel program to WASP, The Wallenberg Artificial Intelligence, Autonomous Systems and Software Program.

Request or more details: contact@wasp-hs.org

Authors

Liane Colonna, Department of Law, Stockholm University
Virginia Dignum, Department of Computing Science, Umeå University
Natalia Filatova-Bilous, Yaroslav Mudryi National Law University/NGO Civil Law Platform
Sandra Friberg, Department of Law, Uppsala University
Jess Haynie-Lavelle, Department of Global Political Studies, Malmö University
Cecilia Magnusson Sjöberg, Department of Law, Stockholm University
Catelijne Muller, ALLAI
Dennis, Munetsi, Department of Global Political Studies, Malmö University
Yulia Razmetaeva, Yaroslav Mudryi National Law University/ Uppsala University
Michael Strange, Department of Global Political Studies, Malmö University
Jason Tucker, Department of Global Political Studies, Malmö University

How to cite this report
WASP-HS. Community Reference Meeting: Challenges and Opportunities of Regulating AI. Report. August 2022.

Design and typesetting
Tatyana Sarayeva