E-PAPER

Regulating AI: Debating Approaches and Perspectives from Asia and Europe

Published by
Heinrich-Böll-Stiftung |
Association of Pacific Rim Universities

HEINRICH BÖLL STIFTUNG HONG KONG

APRU

Publisher:
Heinrich-Boell-Stiftung Asia Ltd. (hbs)
Unit E, 22/F, Derrick Industrial Building, 49 Wong Chuk Hang Road, Hong Kong SAR.
info@hk.boell.org

Association of Pacific Rim Universities (APRU)
Unit 902, Cyberport 2, 100 Cyberport Road, Hong Kong SAR.
info@apru.org

We would like to thank all contributors, including speakers, moderators, writers, designer, and organiser working group (between hbs and APRU) for making this dialogue project possible.

Synthesis report writer: Kal Joffres
Press release writer: Jens Kastner
Layout design: Avery Choi

"Regulating AI" webinar series working group
Axel Harneit-Sievers (hbs HK), Christina Schönleber (APRU), Lucia Siu (hbs HK), Tina Lin (APRU), Carol Mang (hbs HK)

Place of publication: https://hk.boell.org/ | https://apru.org/
Release date: January 2023
License: Creative Commons (CC BY-NC-SA 4.0),
https://creativecommons.org/licenses/by-nc-sa/4.0/

Contents

Overview

The Regulating AI: Debating Approaches and Perspectives from Asia and Europe webinar series was created to open an exchange discussing emerging approaches to the regulation of artificial intelligence (AI) in different parts of the world.

This exchange was hosted against the backdrop of two major shifts in the AI space:

- **Artificial intelligence has left the lab and entered our everyday lives**. AI is being deployed extensively across a variety of industries, from manufacturing, to finance, to social media. It is becoming part of government service delivery in areas such as policing, healthcare, and social protection. Millions of decisions – many of them invisible – are being driven by AI on a daily basis.

- **A critical mass of countries are shifting from moralising AI to regulating it**. The moralising phase of AI was characterised by expert groups conversations, white papers, and voluntary guidelines on the ethics of AI. Governments are now beginning to make use of existing legal frameworks and are establishing new ones to address AI. Key questions such as what kinds of AI constitute unacceptable risk or what kinds of explanations should AI be able to generate for its decisions need to be answered in a pragmatic and legal sense.

What happens in the EU matters for the rest of the world. The EU is a first mover in regulating the digital space. Just as the General Data Protection Regulation (GDPR)[1] has created a global standard in personal data protection regulation, the EU's work in regulating AI is expected to have global policy impact. While this webinar series explored regulatory approaches across a number of countries, the EU's rich set of regulatory frameworks and particularly the EU AI Act have been a primary source of discussion and learnings.

The AI Act is a proposed law on artificial intelligence. It is the first law on AI by a major regulator anywhere. The AI Act was first unveiled in 2021. The Act (i) lays out harmonised rules for the development, placing on the market, and use of AI in EU; (ii) draws heavily on the certification model of 'safe' products used for many non-AI products in the EU to regulate AI; and (iii) it does not replace but will overlap with the protections offered by the General Data Protection Regulation (GDPR). In 2021, a set of committees reviewed and proposed amendments to the Act. In 2023, the Council Presidency will be passed from France to

[1] Edwards, L. (April 2022). The EU AI Act:a summary of its significance and scope. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.

Sweden. The Act is anticipated to become a law by this time and will be enacted 2-3 years later[2].

International cooperation will be crucial to effectively regulating AI. Effectively regulating AI will require that governments go toe-to-toe with some of the world's largest corporations – corporations whose market capitalization exceeds the GDP of many countries. Governments are beginning to see the importance of working collectively to impose regulation and penalties. Ensuring AI serves the public interest will only happen if governments act collectively and persistently.

This synthesis introduces three foundational areas in AI regulation and key considerations discussed by AI experts from Europe and the Asia-Pacific region over the course of the webinar series:

- **Webinar 1: Risk-based Approach of AI Regulation**. AI applications can be categorised by the levels of risk they imply, with appropriate regulatory restrictions and exemptions specified in the regulatory framework. The EU's proposed AI Act is taking a significant step in defining the types of AIs with "unacceptable risks", as well as how these can be clearly defined. The first webinar focused on merits of a risk-based approach, bringing in perspectives from Toby Walsh (Scientia Professor of Artificial Intelligence at University of New South Wales); Alexandra Geese (Member of the European Parliament for the Greens EFA and Coordinator for the Greens EFA in the AI in the Digital Age Special Committee); and Jiro Kokuryo (Professor at the Faculty of Policy Management at Keio University).

- **Webinar 2: Explainable AI**. AI algorithmic designs may involve assumptions, priorities and principles that may appear opaque and incomprehensible to users and even operation managers. "Explainable AI", or "XAI", is an umbrella term for various AI applications to provide output that enables humans to understand why a system made a particular decision. This is seen as key to fostering public trust, informed consent and fair use of AI applications. The second webinar focuses on proposals of "explainable AI" and "trustworthy AI" with initiatives to create AI applications that are transparent, interpretable, and explainable to users and operations managers. It was joined by Liz Sonenberg (Professor of Information Systems at the University of Melbourne); Matthias Kettemann (Head of research programme, Hans-Bredow-Institute / HIIG); and Brian Lim (Assistant Professor in the Department of Computer Science at the National University of Singapore)

[2] Edwards, L. (April 2022). The EU AI Act:a summary of its significance and scope. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.

- **Webinar 3: Protection of Data Rights for Citizens and Users**. Protection of data rights for citizens and users is a hugely complex task, with risks deriving from both under-regulation and over-regulation of AI applications. The third session focused on having a balance with data rights, bringing in perspectives of Sarah Chander (Senior Policy Adviser at European Digital Rights); M. Jae Moon (Underwood Distinguished Professor and Director of the Institute for Future Government at Yonsei University); and Sankha Som (Chief Innovation Evangelist of Tata Consultancy Services).

About the hosts

The webinar series was co-hosted by the Heinrich-Böll Stiftung (hbs) and the Association of Pacific Rim Universities (APRU).

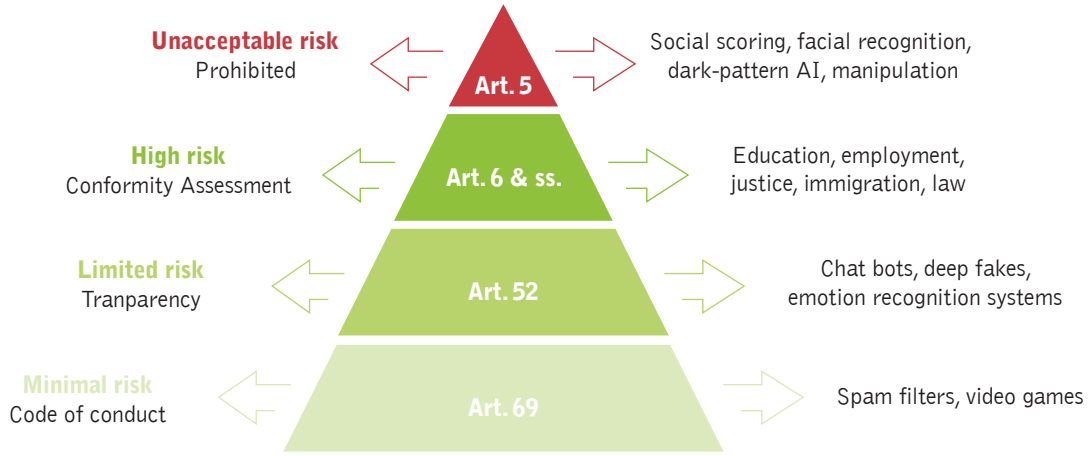
Heinrich-Böll Stiftung (hbs). The political foundation is affiliated with Germany's Green Party. It has a global network of more than 30 offices involved in the discussion of regulatory and governance issues surrounding digitalization. hbs deals with relevant European actors, including civil society and members of parliament, policy-makers and other experts involved in the EU's AI Law initiative.

Association of Pacific Rim Universities (APRU). The association has been pursuing debates in the field of AI policies and ethics since 2016. APRU in collaboration with UN ESCAP and Google set up the AI for Social Good network supporting governments and key stakeholders in developing insights on how best to develop governance approaches that will address challenges associated with AI, while maximising the technology's potential in the Asia Pacific region.

Risk-based approach of AI regulation

What it is and why it matters

The regulatory space for artificial intelligence in the EU is defined by a suite of laws that includes the General Data Protection Regulation (GDPR) Act and the Digital Services Act. At the centre of this suite is the EU AI Act, which categorises AI applications based on the levels of risk they pose:



A "risk-based" approach of AI regulation[3]

The risk classification plays the driving role in determining the regulatory environment for different AI systems. Some applications are generally prohibited with few exceptions, others will require certification and close monitoring, and yet others will see minimal oversight.

- **Unacceptable risk.** These are AI applications that break EU values (e.g. subliminal, manipulative, or exploitative systems that cause harm; real-time, remote biometric identification systems used in public spaces for law enforcement; and all forms of social scoring). AI applications in this category are prohibited.

- **High-risk.** High-risk AI applications are those that "evaluate consumer creditworthiness, assist with recruiting or managing employees, or use biometric identification, as well as others that are less relevant to business organisations"[4]. The list of applications under this category will be annually reviewed and updated.

---

[3] Edwards, L. (April 2022). The EU AI Act:a summary of its significance and scope. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.

[4] Benjamin, et al. (August 10, 2021). What the draft European Union AI regulations mean for business. QuantumNlack Ai by McKinsey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-the-draft-european-union-ai-regulations-mean-for-business.

- **Limited and minimal risk**[5]. These are AI applications that pose little to no risk on a person's safety or rights. Limited and minimal risk AI are commonly used for business operations (e.g. chatbots, emotion recognition, and biometric categorization systems).[6] These will be subject to the GDPR Compliance.

The EU AI Act is grounded in Europe's long standing philosophy of consumer protection. The consumer protection approach means not only certifying the end product but also regulating the inputs and processes used to create the product. The philosophy focuses on protecting peoples' safety and fundamental rights. This is achieved by treading a line between establishing boundaries on why and how AI systems are used while ensuring that regulation is not so burdensome as to hamper innovation.

The EU AI Act will influence AI regulation globally. The EU is a pioneer in regulation of the digital space. Just as many national policies on data privacy around the world have been influenced by GDPR, we can expect many will also be strongly influenced by the EU's approach to AI.

Key considerations

Classifying AI technologies by risk is challenging because these technologies are multi-purposed. Facial recognition, for example, has applications that benefit society and others that may be harmful. While remote biometric identification is generally prohibited under the EU AI Act, an exception is carved out for the case of missing children.

The actual risk of an application varies greatly from one context to another. For example, a system to prioritise medical appointments might be low risk in Denmark but a high risk in Germany. Denmark has a single-track appointment system, where everyone is part of the same kind of health insurance scheme. Germany has a dual-track appointment system, where people with private insurers tend to get appointments before people who use public health insurance. A prioritisation system trained on data from Germany will likely see people with private insurance or attributes correlated with having private insurance get appointments before those who have public insurance, undermining prioritisation by medical necessity. This issue in an appointment prioritisation system can have critical life and death outcomes. Following the EU's risk classification, a great deal of work is still needed to understand the

[5] Securiti. (November 1, 2022). European Commission's Proposed Artificial Intelligence Regulation. https://securiti.ai/blog/european-commissions-proposed-artificial-intelligence-regulation/.

[6] Edwards, L. (April 2022). The EU AI Act:a summary of its significance and scope. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.

risk of a specific application in a specific context. Regulatory frameworks must be careful about prohibiting entire classes of applications a priori.

Transparency is a crucial complement to regulating AI a *priori*. Transparency is essential to managing risk. It requires that the producer and implementers of an AI system set out the goals of the system, how it is optimised, and how it achieves those goals. When the "black box" of the AI system is opened up, people should be able to examine the data and algorithm. If this is not possible, the system must be regularly checked to understand whether the outcomes are fulfilling the expectations and goals set out at the beginning. All this must happen with the people who are most affected by these processes. Transparency is particularly crucial for AI applications that are not mature; applications in a variety of areas such as emotion recognition are not currently able to consistently generate accurate predictions or inferences.

AI applications that classify people and determine access to services often involve a great deal of risk. These systems are strongly positioned to reinforce or amplify existing inequities if not properly monitored. In particular, there is a strong case for including AI systems in the medical arena as part of the EU AI Act's high-risk classification.

There are contexts where transparency is not the best approach to managing risk. Some systems cannot be made transparent for technical reasons. AI applications involving computer vision, for example, are particularly challenging to be made transparent. There are situations where transparency is neither useful nor desired. We put our lives in the hands of doctors despite the fact they are not fully transparent. Instead, we rely on regulations and institutions that ensure that patients don't need to be medical experts so they can trust their doctors. Additional approaches beyond risk management will need to be developed in areas where transparency may not be the solution.

Some policymakers are considering layered disclosure to trusted regulators. There is an important discussion about whether businesses should be forced to disclose an algorithm to the public or whether those disclosures should be made to an appropriately trusted regulator. This could also apply to algorithms. The role of appropriately trusted regulators is expected to improve as they may be in a better position to deal with highly technical matters.

Implementing AI regulation is a challenge unlike any other regulatory challenge governments have faced before. Governments need to stay up to speed with new developments in AI, do real time risk assessments, and hold continuous dialogues due to the evolving nature of AI – a tall challenge given governments are already behind in developing frameworks for governing AI. AI already plays a significant role in determining millions of decisions, from how social media posts are prioritised to how delivery routes are optimised. Governments must contend with the fact these changes are happening at an unprecedented speed and scale.

The adaptive and evolutionary nature of AI means it requires significantly more monitoring than many other subjects of regulation. Unlike many other products regulated through consumer protection, the nature and behaviour of an AI system can change significantly over time. Whether a product violates human rights depends on how it evolves. A product that was certified can evolve to something entirely different a year later.

Building a better society with AI requires thinking differently about AI's role in the future. Many discussions about the role of AI in society start from assuming automation and working backward to examine how that might happen. The logic of automation and efficiency often leads to negative societal outcomes – disempowering people and reinforcing inequities. The starting point should be imagining a future where a societal issue has been resolved and then working backwards to consider the role that AI could play in making that future happen.

How AI is shifting the balance of power between the private and public sectors in public service delivery is underexplored. AI as a policy ideology strongly increases the private sector's role in delivering public services. The efficiency-driven logic of the private sector may not always be aligned with the public sector's focus on societal outcomes. Some of these challenges can be anticipated. Systems that privilege efficiency for one segment of the population can make things worse for another. Systems designed by the private sector can also represent private interests in subtle and sometimes difficult-to-detect ways, such as a mapping application that directs traffic through streets on which advertisers are based. The broader question of how increased reliance on privately-built AI systems can shift the balance of power in the delivery of government services needs further exploration.

Answering the difficult questions of AI is challenging because many of the players needed to answer them are not engaged. Participation in AI policy and ethics discussions have been driven by the private sector. Participation from civil society groups and ethics experts needs to be improved. There is a need for more balanced AI governance.

Some countries are focusing policy on achieving harmony rather than control over AI. Control is central to the Western philosophy of regulating AI. A more Eastern philosophy takes as a starting point that AI is complex; various elements of humans and machines are networked and influence each other to make behaviour of the system unpredictable and therefore impossible to control. Dynamic resilience is needed rather than control. This requires continuous monitoring of emerging risks, a willingness to rapidly adapt to the changing AI environment, and influencing technology to benefit humans. In Japan, the harmony-oriented approach to AI has been driven by voluntary guidelines and efforts to build greater understanding of the ethical dimensions of work amongst technologists.

Explainable artificial intelligence

What it is and why it matters

AI systems are often opaque. We don't understand how they make recommendations, what their assumptions are, nor the logic connecting the input into the system with the outputs. This technology promises to be at the centre of our lives not just in consumer products but in healthcare, social protection, banking, and overall policymaking. If we do not understand how these systems work, we cannot understand the kinds of biases or weaknesses those systems have and address them.

Explainable Artificial Intelligence (also known as XAI) is an umbrella term for various techniques that are undergoing trial so that AI provides outputs that enable humans to understand why a system has made particular decisions. It is more than a 'nice-to-have' in AI systems. EU regulators are working on making explainability a requirement for some AI applications and increasingly seeing the importance of rights to explanation. Successful implementation of policies around explainability will contribute to preventing people from feeling powerless against AI and from AI breeding mistrust and insecurity.

To illustrate how explainable AI works, consider a credit scoring system that has refused a loan to a consumer. Explainable AI would not only provide the result of the application but it would enable the customer to ask the system what they could do differently to get a positive outcome. The system might compare the profile of the customer with customers with similar but different attributes and make a recommendation. It might also indicate that the customer is holding three accounts with a cumulative debt of $5,000 and if those accounts were consolidated half the amount was repaid, the loan would be approved.

Explainable AI can lead to better decision-making. Human decision-making suffers from shortcuts and biases. Partnership between people and AI can help mitigate these issues. Many of the advances in the space of explainable AI are coming from regulation beyond the EU AI Act. The Digital Service Act includes a number of clauses for online platforms, such as Google, Meta, and Twitter. These companies are obliged to provide access or insight to data related to the algorithms they use[7]. The GDPR already outlines duties of those who have automated decision-making engines that use personal data to provide an explanation of the logic underlying the algorithm[8]. However, this package is far from complete. For example,

---

[7] EUR-Lex. (October 27, 2022). Document 32022R2065. http://data.europa.eu/eli/reg/2022/2065/oj.

[8] EUR-Lex. (May 5, 2016). Document 32016R0679. http://data.europa.eu/eli/reg/2016/679/o.j.

the GDPR does not require the explanation of an algorithm if a human is involved at some point in the decision-making process.

Explainability is most important for systems that make judgement calls. Clinicians and patients do not necessarily want technical explanations for how an AI diagnostic tool has made a diagnosis nor do they need to know the algorithms for image processing behind an MRI. They want to know: is it accurate? Is it approved by regulatory authorities? Is it clinically tested? When a system is making a recommendation or a judgement call (e.g. recommending a course of treatment), clinicians and patients will want to query the system to understand how that recommendation was produced to understand whether it is trustworthy.

Key considerations

Privacy and explanation can be at odds with each other. Privacy implies less shared data and explanation implies more shared data. AI models that provide explanations can be inverted to identify the source data, including confidential personal data. Privacy researchers have demonstrated that this can be done even when the information is not explicit[9]. One approach to addressing this challenge is to forgo providing explanations to the public and instead provide them to a technical group who are trusted and can interpret how the system works for the public. However, this assumes that a trusted authority or bridge can be put together and may not be possible in all contexts.

AI will need human assistance to defend itself around critical decisions. In sensitive applications or exceptional cases, a human will be needed to gain a deeper understanding of why the AI system has behaved the way it has. Experts need to be able to obtain more technical explanations. This will also open up a space for a whole new series of careers.

Explainable AI is in its infancy and generating useful explanations is tremendously challenging. There are technical challenges to explainability. Machine learning systems can be difficult to comprehend even for the developers of those systems. They can be built with thousands and sometimes millions of parameters. It becomes impossible to derive precise reasoning from these systems. They may use parameters that are not readily understood by humans. For example, can facial recognition provide meaningful explanations? Even simpler, rule-based systems that use decision-making rules with predefined outcomes can be challenging. An audit trace of rules that were applied to arrive at a decision is not generally meaningful to a user.

[9] Zhao, X., Zhang, W., Xiao, X., & Lim, B. Exploiting Explanations for Model Inversion Attacks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 682-692).

The design of explanations is too often left to lawyers or technologists. Any approach to designing explanations is an interdisciplinary endeavour that includes technology, law, psychology,[10] and design thinking. The design of explanations needs to be centred on the kinds of information people need, as well as how they understand it.

Explainability requires focusing on the psychology and human experience of what makes good explanations. Consider a medical application that predicts the probability of mortality of a patient using decision rules. While the system can show which rules were important and which ones were not in making the assessment, this is not how medical doctors understand decisions. The explainer not only needs to understand the rationale for the decision, they need to understand what the other person already knows and doesn't know about the situation – all with an appropriate level of detail. Explainable AI bridges the gap between how machines and humans think.

Effective strategies for designing explanations look at how people think and how they explain decisions to each other. Examining the kinds of explanations people give each other in a particular domain contain a great deal of useful information about generating understandable explanations. Doctors understand biological processes and causal mechanisms, not just variables. Explanations need to be relatable to concepts that people are familiar with and guided by underlying human processes of reasoning.[11]

Humans are sometimes the limiting element in explainable AI systems – and explainability can bridge the gap. Literature on decision support systems has already demonstrated that the way in which information is presented to people can have a significant impact on their decision-making performance. People whose objective performance is less than the software can also act as a bottleneck or gatekeeper by dismissing valid recommendations from AI systems. Explainability is essential to establishing effective partnerships between humans and AI and ensuring the implementation of good decisions.

Governments need to set benchmarks for what is considered as good explanation. What constitutes a good explanation varies from one domain to the next. There needs to be a framework where different communities of practice can derive standards that define the level of explainability needed before an AI system can be considered safe enough for a use case. The benchmarks for minimally acceptable explainability should be reviewed regularly as explainable AI technology improves.

---

[10] Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

[11] Zhang, W., & Lim, B. Y. (2022, April). Towards relatable explainable AI with the perceptual process. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-24).

Artificial intelligence and data rights

What it is and why it matters

Data is the lifeblood of AI systems. Many applications of AI – from evidence-based policy-making to detecting fraud in social protection systems – will not only require systematised data but also large quantities of personal data. Citizens are increasingly demanding services that recognize them as individuals and that can help them navigate the complex maze of government offerings to find the ones most relevant to them.

The importance of data rights is not just in the abstract analysis of big data but around making inferences, predictions, and decisions that have real impacts on peoples' lives. Data rights protect people's fundamental rights to the protection of their personal data. Data means any information that is related to or can be identified to an identified or identifiable natural person. How do we empower people in a context where data systems are opaque? How do we recognise that people face different kinds of marginalisation in terms of barriers and the use of AI systems? The use of data in government policy and decision-making implicates the wider set of fundamental rights: rights to social protection, non-discrimination, freedom of expression, and many more.

Key considerations

There needs to be a broad consensus about the use of data in governance. That consensus must not only address concerns of privacy and other fundamental rights; it must be focused on building trust between citizens and government.

Bias contained in datasets used to train AI is a fundamental societal problem. AI systems trained with datasets and proxies that poorly represent the relevant populations as well as AI systems trained on datasets that incorporate human biases lead to AI systems to behave in ways that are biased. This has been readily demonstrated in areas from policing to hiring.

Researchers believe there are already AI systems on the market that are making harmful recommendations due to biased data. The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a correctional assessment tool used across the US justice system to assess the risk that a criminal defendant will re-offend. The COMPAS is a broadly used artificial intelligence (AI) software system and its algorithm is proprietary and secret.[12]

---

[12] Carpenter, C. (February 21, 2021). The Threat of Black Box Algorithms - and How Business Leaders Can Survive Them. https://oxfordbusinessreview.org/the-threat-of-black-box-algorithms-and-how-business-leaders-can-survive-them/

While the system does not use race as an explicit feature, there is analysis that suggests that the algorithm operates in a racially-based way. Black offenders are twice as likely as white offenders to be labelled by the system as high risk while not actually going on to re-offend. The recommendations of the system are derived from the properties of over a hundred neutral-sounding characteristics that appear to be generating biased analyses.[13]

Data labelling is a crucial step in which data bias can be prevented. A key to ensuring that AI systems are not biased is regulating their inputs in terms of data. Data labelling is a critical step where bias can creep in. In data labelling, humans or AI systems identify and categorise text or images to provide context to machine learning models.

The presumption that bias can be combated at a technical level is frequently false. Some forms of bias can be addressed at a technical level, such as a facial recognition system that doesn't work well for people with darker skin tones. However, there are many types of systems that will always generate a harmful decision, prediction, or inference because bias is so deeply enmeshed with the training data. Policing data, for example, regularly reflect discriminatory behaviour. AI systems can scale up or supercharge these discriminatory behaviours.

Biases are like software bugs. It is impossible to guarantee that a system is free of bugs – but a great deal of effort is invested in ensuring that bugs are addressed and preventing them from being catastrophic. The real question is how a system is monitored for bias and how the system is dealt with once bias is discovered. Fortunately, AI can be used to look for biases in both datasets and to audit system results not only to look for bias but to examine deeper questions such as whether the system performs reliably.

Data rights needs to be concerned not just with individual rights but also impacts on broader society. Tdatahough data rights frameworks primarily focus on people who are the owners or subjects of data, there are broader questions to consider about the society-wide impact that AI and data can have in reinforcing marginalisation and inequality. Data rights must also consider questions of economic and social justice.

Existing and proposed legislation already includes a number of provisions related to data protection. Existing and proposed legislation already includes a number of provisions related to data protection. The EU AI Act includes a proposed obligation for implementers of high risk AI to do an impact assessment to look at how the system might impact the fundamental rights of people using the system. There is a need to increase the scope of these remedies so they cover the people affected – not just the users and data owners. There is also a

---

[13] Angwin, J. (May 23, 2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

need to take into account a wide range of vulnerabilities people may face, which can result in greater or less impact to individuals.

The right to withdraw consent for the use of data under GDPR poses some thorny questions for AI. The GDPR allows individuals to withdraw consent for the use of personal data.[14] AI models that are trained using this data are derivatives of the personal data that have been withdrawn. It is sometimes possible to "invert" these models to extract personal data that has been withdrawn. Some experts expect that GDPR will go beyond the data and require that companies destroy models they have built on the basis of withdrawn personal data.

Data rights lead to new security requirements for AI systems. The introduction of AI systems will lead to new kinds of attacks that can leak data, such as data poisoning, model poisoning, model inversion and stealing. Data poisoning is an attack wherein malicious and false information is injected into a machine learning model's training dataset and usually occurs during data collection. Model poisoning on the other hand involves tampering a machine learning model which allows for misclassification of data.[15] Model inversion and theft involves attackers reverse engineering an AI model by inputting enough data to generate a close approximation of a private training data set.[16] [17] Awareness of these possibilities is crucial as the technical systems and know-how for dealing with these are still very much in their infancy.

Many AI applications that may appear to be mundane are not. Today, most AI implementations by enterprise are focused on areas where efficiency gains are to be made, like invoice processing and trade finance. While these applications may appear mundane, this may be illusory. Questions must be asked about the broader societal impacts of AI. What are the broader impacts on power structures and justice? To whom are these applications accountable? Who is benefitting from gains in efficiency and who may be losing out?

---

[14] EUR-Lex. (May 4, 2016). "Document 02016R0679-20160504," http://data.europa.eu/eli/reg/2016/679/2016-05-04.

[15] Lyu, L. et. al. (November 2020). "Threats to Federated Learning" in Federated Learning pp. 3-16. Springer, Cham. DOI:10.1007/978-3-030-63076-8_1

[16] Zhaeo, X., et. al. (April 2021). Exploiting Explanations for Model Inversion Attacks. https://arxiv.org/pdf/2104.12669.pdf.

[17] Boenisch, F. (December 2020). Attacks against Machine Learning Privacy (Part 1): Model Inversion Attacks with the IBM-ART Framework https://franziska-boenisch.de/posts/2020/12/model-inversion/

Conclusion

AI has already moved out of the lab and begun impacting our lives in all kinds of ways, some more visible than others. Governments are shifting from moralising AI to regulating it. As we have engaged in this global exchange on policy approaches to AI, one theme has connected many of the comments: achieving AI regulation that contributes to thriving societies will require a rebalancing. That rebalancing needs to occur around on three dimensions:

Regulating a priori & managing live systems. While governments are fully capable of banning or restricting entire categories of AI uses, the risks posed by AI are so context-sensitive that regulating them a priori and regardless of context is a blunt instrument. Shifting the balance towards managing systems through transparency will require that governments build and expand capabilities for scanning the horizon and dealing with the rapid change. They will also require significant progress around AI explainability.

Individual rights & societal impacts. Policy discussions on AI have often focused on individuals' fundamental rights. These discussions need to be rebalanced for greater consideration of the broader societal impacts of AI: how data protection should not just be for those who are the subjects of data but also those who can be affected by how data are used; how AI trained on seemingly innocuous datasets can amplify marginalisation and inequalities; and how public services built on private sector AI systems can shift the balance of power between public and private interests. Many of these conversations have been driven by the private sector and technologists. They need to be rebalanced to reflect their true multidisciplinary nature, with participation from ethicists, civil society organisations, and impacted communities.

Risk & opportunity. Policy discussions centred on the risks of AI can sometimes lose sight of the opportunities AI offers for creating a better future. AI has the potential to help address human biases in decision-making and deliver a level of explainability that many of today's institutions cannot, from banks to government agencies. The role AI can play in building thriving societies goes far beyond the monochromatic focus on efficiency and automation often seen from the private sector. The opportunities of AI must be monitored and acted upon as rigorously as the risks.

Regulating AI: Debating Approaches and Perspectives from Asia and Europe

hbs HK and APRU want to give thanks to contributors of the webinar series:
1st webinar "Risk-based Approach of AI Regulation"
Speakers: Toby Walsh (University of New South Wales), Alexandra Geese (Member of European Parliament), Jiro Kokuryo (Keio University)
Moderator: Zora Siebert (hbs Brussels)

2nd webinar "Explainable AI"
Speakers: Liz Sonenberg (University of Melbourne), Matthias Kettemann (Hans-Bredow-Institute / HIIG), Brian Lim (National University of Singapore)
Moderator: Kal Joffres (Tandemic)

3rd webinar "Protection of Data Rights for Citizens and Users"
Speakers: Sarah Chander (European Digital Rights), M Jae Moon (Yonsei University), Sankha Som (Tata Consultancy Services)
Moderator: Axel Harneit-Sievers (hbs HK)

Disclaimer:
This report provides a synthesis that does not attribute particular statements to individual participant speakers, and does not in any way represent a consensus among participants. Rather, this is a report that highlights concerns, issues, and approaches towards solutions that emerge from the discussion and are presented here as a summary of the webinar series held, for further reflection and for ongoing debate.

This document contains additional references to support claims speakers made during the webinar sessions. These references were not provided by speakers. The references were selected based on relevance and recency.

The opinions expressed in this report arise from webinar speakers, moderators, and the synthesis report writer. They do not necessarily reflect the views of the Heinrich-Böll-Stiftung or the Association of Pacific Rim Universities.