Artificial Intelligence and the Rule of Law Massive Open Online Course

Booklet

1. Course Description

Artificial Intelligence (AI) has unprecedented potential to improve lives and livelihoods across the world, as well as to make remarkable progress towards achievement of the United Nations Sustainable Development Goals. However, while the socio-economic opportunities of AI are huge, they are inextricably connected to risks and challenges.

The Massive Open Online Course (MOOC) on AI and the Rule of Law is an introductory course on AI's adoption and impact for the rule of law, defined as a principle of governance in which all persons, institutions, and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced, and independently adjudicated, and which are consistent with international human rights, norms, and standards.

The course's scope is twofold. First, it explores the digitalization of justice systems and, specifically, the opportunities and risks with the increasing adoption of AI technologies across justice systems and contexts such as online courts. Second, it highlights AI's impact in the administration of justice, particularly regarding human rights and evolving AI ethics and governance issues increasingly preoccupying judicial operators. Although an important topic on its own, the course does not consider AI's impact on substantive law and whether or how legal doctrine needs to be addressed.

Leveraging examples and cases from around the world, the course engages judicial operators in a global and timely discussion around AI's impact and implications in upholding the rule of law.

The six modules the course unpacks are:

- An Introduction to Why Digital Transformation and AI Matter for Justice Systems
- AI Adoption across Justice Systems
- The Rise of Online Courts
- Algorithmic Bias and its Implications for Judicial Decision Making
- Safeguarding Human Rights in the Age of AI
- AI Ethics & Governance Concerning Judicial Operators

2. Course Objectives & Outcomes

The overall objectives of the MOOC are to:

1. Stimulate a participative dialogue with judicial operators and other relevant stakeholders on the adoption of responsible AI in the judicial system;
2. Promote knowledge of digital innovations in the justice system;

    3) Facilitate knowledge exchange and experience sharing among judicial operators and other relevant stakeholders on AI, existing norms and standards (hard and soft law) in the field, and its implications for human rights;

    4) Highlight existing case studies and best practices that translate legal and ethical principles into practice both in terms of the use of AI in justice systems and in cases involving AI impacting human rights.

The teaching and learning objectives of the MOOC are to:

1. Strengthen capacities of judicial operators to address AI's impact in their domain;
2. Ensure that judicial operators are equipped with the necessary information and knowledge concerning AI-based applications in justice systems;
3. Ensure that prosecuting services, in AI-related cases, are aware of the rule of law as it concerns AI technology and correlated risks;
4. Provide an opportunity for other relevant stakeholders, including policymakers, academics, social workers and non-profits, to build understanding of AI issues and the rule of law.

The target audience of the MOOC are judicial operators, including judges, prosecutors, public lawyers, and other stakeholders working for judiciary branches, prosecution services worldwide, and sectors including legal technology companies.

Other relevant audience groups include policymakers, academics, social workers, and civil society representatives.

3. Prerequisites

There are no explicit prerequisites, including the need for a technical background, to enroll in the course. However, judicial operators are the target audience and, hence, possessing a good understanding of judicial processes and the rule of law may help those participants better grasp the topics and/or case studies.

4. Essential Resources

The course requires access to a computer and/or a mobile device, and reliable broadband.

5. Pedagogical Approach

The course is based on a constructivist theory of learning, where individuals are provided the opportunity to construct their own sense of what is being learned by building internal connection or relationship among the ideas and facts being taught. It calls on individuals to be active information receivers, applying what they already know to the new information they are receiving in order to make sense of it and apply it in real life contexts.

Specifically, the course leverages elements from both project-based learning and inquiry-based learning to trigger individuals to absorb and reflect on the information being taught through learning objectives and driving questions that participants can methodically investigate to come to conclusions.

The modules are built on the 5 E's Model: Engage, Explore, Explain, Extend, and Evaluate. During the engagement stage, participants are introduced to the concept and become engaged in the new learning experience, tasked to make connections to prior knowledge and what is to be learned. During the exploration stage, participants have the opportunity to participate in an activity or a series of activities to directly practice what is being learned. During the explanation stage, participants start making patterns from the earlier stages in order to make sense of the new information. During the extension stage, participants expand their learning, practice skills and behavior, and make connections or applications to related concepts and in the world around them. During the evaluation stage, participants are invited to illustrate their built understanding and capacities.

6. Assessment

Assessment will be based on essential short quizzes (composed of 5 multiple choice questions) following each of the six modules. There will also be recommended material and online and offline activities which you can choose to take to enhance your understanding of the topics covered in each module.

Certificates of course completion will be provided for participants who have successfully completed the formative assessment for all six modules.

7. Course Structure

The course, its scope, and its objectives will be introduced through an overview video and the curriculum itself will be delivered through six modules, designed to promote reflection and discussion around key issues related to AI and the rule of law.

Specifically, each module will include:

- Video lectures by subject-matter experts
- Lecture summaries
- Lecture Script (WebVTT)
- List of recommended reading materials
- Module Assessment

- Relevant use cases and examples

Each module will be led by a facilitator, a relevant subject-matter expert, and feature 3 to 4 guest speakers.

Participants are expected to spend one to two hours per module and are recommended to take one module per week.

The course will be available in two formats:

- a closed 6-week course - open to participants upon registration - with live moderation and a certificate of completion following the course, and
- an openly available self-paced course available as part of UNESCO's online decision maker's toolkit.

Lastly, the course will be accessible in English, French, Spanish, Russian, Chinese, Arabic and Portuguese.

8. Course Modules

8.0 Welcome to the Course

A welcome video by UNESCO which will introduce the course, its objectives, and the six modules. The video will also present the vision behind the course, as well as how it fits within the wider UNESCO Global Judges Initiative.

8.1 MODULE 1: AN INTRODUCTION TO WHY DIGITAL TRANSFORMATION AND AI MATTER FOR JUSTICE SYSTEMS

8.1.1 Module Focus

This module will introduce why digital transformation and Artificial Intelligence (AI) matter for justice systems, both regarding how AI can be responsibly adopted across justice systems and how judicial operators, in their quest to uphold the rule of law, are affected by AI's adoption across societies. It will first unpack and define the terms 'digital' and 'Artificial Intelligence,' and provide an explanation of current technological capacities across justice systems. Next, it will explore digital transformation trends worldwide, and consider how AI interacts with access to justice in the African context. Finally, it will discuss how the global pandemic and associated lockdown measures have jeopardized access to justice worldwide.

8.1.2 Learning Objectives

From this module, participants will:

- Familiarize themselves with concepts of digitalization, digital transformation, Artificial Intelligence, and the rule of law
- Acquire a general understanding of recent digital and AI adoption trends across justice systems
- Contemplate the role of judicial operators in upholding the rule of law in the digital era

8.1.3 Lecturers

Cedric Wachholz | UNESCO, Chief of Section, Digital Innovation and Transformation
Grace Mutung'u | High Court of Kenya, Advocate | Kenya
Hon. Katherine Forrest | Cravath, Swaine & Moore LLP; Former Judge of the United States District Court for the Southern District of New York | USA
Hon. Diego Garcia-Sayan | UN Special Rapporteur on the Independence of Judges and Lawyers; Former Justice and Foreign Affairs Minister of Peru

8.1.4 Structure

Lecture 1.1: Why should justice systems care about digitalization and AI?

This presentation will introduce the module and its objectives. Recognizing AI's transformative potential, it will illustrate how justice systems and judicial operators are increasingly affected by AI's adoption worldwide. It will set the scope of the course from two angles: Firstly, by looking at how AI can be responsibly adopted across justice systems, and secondly as to how judicial operators are impacted by AI's adoption across societies.

Lecture Script (WebVTT)

Welcome to Module one. An introduction to why digital transformation and AI matter for justice systems. Lecture one: Why should justice systems care about digitalisation and AI? Hello, I'm Cedric Wachholz, chief of UNESCO's Digital Innovation and Transformation section. On behalf of UNESCO and our partners, I welcome you to this course on AI and the Rule of Law. Since 2013, UNESCO's Judges' Initiative has raised the capacities of judicial actors on international and regional standards on freedom of expression, on access to information, and on the safety of journalists. Over 23,000 judicial actors, including judges, prosecutors, and lawyers, have been trained on these issues.

We offered them a series of Massive Open Online Courses (MOOC), on-the-ground training and workshops, and a number of toolkits and guidelines. With this course, we are expanding our engagement with judicial actors on the human rights and legal implications of emerging technologies. Let me talk about why this is important, why we offer you this course on AI and the Rule of Law. Artificial intelligence has an unprecedented potential to improve lives, and to help achieve the United Nations Sustainable Development Goals. However, while the use of AI presents many socio-economic opportunities, there are also serious risks and challenges. These relate to bias, to discrimination, to lack of transparency, and loss of autonomy of judicial actors, among others. Rule of law and justice systems are a cornerstone of democratic societies, as you all know. As technologies like AI advance, and they really do permeate all spheres of life today, we need to ensure that justice systems are prepared to address the impact of AI on societies.

This module will introduce why digital transformation and AI matter for justice systems. I will be joined by Grace Mutung'u, researcher and advocate at the High Court of Kenya, who will talk about transformation and access to justice, including its gender dimensions. Judge Katherine Forrest will discuss the deployment of AI tools throughout justice systems. This includes the use in investigations for collecting and analyzing documentary, voice, or video evidence to predictive justice systems. The coronavirus pandemic and subsequent lockdowns affected operation of justice systems worldwide. Justice Diego Garcia-Sayan, the UN Special Rapporteur on the Independence of Judges and Lawyers, will underline some unanswered questions about guarantees of access to justice and due process in relation to the use of digital communication tools. Let me now provide some background. E-governance has shaped the way citizens and governments interact. The use of ICTs has improved the efficiency of government agencies, for example, by providing government services online. They've also enabled innovation and transparency in governance by moving us towards open government data. In a similar vein, ICTs are also used in justice systems for digitalization of court records, that

are made available online for greater transparency. What is the role of artificial intelligence or AI in this context? AI refers to machines capable of imitating certain functionalities of human intelligence, including such features as perception, learning, reasoning, problem solving, language interaction, and even producing creative work. As an example, AI solutions have contributed significantly to the international response to COVID. AI has aided contact tracing. It has also allowed for targeted support measures for marginalized groups and helped combat deadly online disinformation through tagging and filtering algorithms.

In recent years, there have been significant commercialization of AI in at least three domains: First, AI-Generated Content: AI systems can generate high-quality text, audio, and visual content to a level that it is difficult for humans to distinguish between synthetics and non-synthetic content. Secondly, Image Processing: Computer vision, a branch of computer science that works on enabling computers to see, identify and process images in the same way that human vision does, has seen immense progress in the past decade. We all know it in applications that also include autonomous vehicles. Thirdly, Language Processing: Natural Language Processing, or (NLP), is a branch of computer science, concerned with giving computers the ability to understand the text and spoken word in much the same way human beings can. NLP has advanced such that AI system with language capabilities now have meaningful economic impact through live translations, captioning and virtual voice assistants many of us use. But these advances in AI-generated content, image and language processing are also creating new challenges. For example, AI systems generating synthetic faces have rapidly improved. This has led to the proliferation of 'deepfakes' that erodes people's trust in information they encounter online. Some actors misuse deepfake technology to imitate authority figures to spread online disinformation, resulting in adverse implications for democracy and political stability. Concerns about the infringement of human rights, like the right to freedom of expression, right to privacy, the right to equality are growing rapidly. For instance, AI-powered facial recognition software has raised concerns about unwanted surveillance and the protection of privacy. AI filters on social media platforms have also threatened the freedom of expression through filtering and censorship.

What is the concrete impact on the role of the judiciary? Justice systems have adopted AI for a host of activities which include today predictive justice that makes use of algorithms to process cases supporting judges in the decision making, but also creating risks for bias and discrimination in decisions based on past data and proprietary algorithms. Law enforcement where predictive policing has become a widely used tool in some jurisdictions. Large sets of historical crime data are analyzed by AI to help law enforcement decide where to deploy police or to identify individuals who are more likely to commit or be a victim of crime. These systems, of course, have the risk of being biased and discriminatory. As the use of AI systems expand, legal concerns have emerged. Namely, it has raised concerns for fairness, accountability and transparency in decision making by automated or AI-enabled systems. Self-learning algorithms, for instance, may be trained by data sets that contain biased data. It might be biased from previous unfair decisions made with sociocultural prejudice, or it might be based on skewed data sets suffering from the lack of inclusive training data. The results can be terrible and flawed high level applications in the justice systems. There is also the issue of accountability. If any AI program makes a wrong call or judgment, does the responsibility fall on the programmers, judges, or the system as a whole? Not all AI decisions can be explained, so it is difficult to find the causes of issues as well as where the responsibility for them lies. Let me conclude by highlighting also that a judicial system embarking on the digital transformation journey, must focus on the use of technologies like AI in a manner that it is rights-

based, open, accessible, and informed by multi-stakeholder consultation. I hope this introduction has set the stage for deep diving into the use of the AI in justice systems. I hope you will enjoy the lectures, readings, and the quizzes.

Relevant Readings / Activities

- Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe. Available at: <https://www.turing.ac.uk/research/publications/ai-human-rights-democracy-and-rule-law-primer-prepared-council-europe>.
- Hildebrandt, M., (2020). *Law for Computer Scientists and Other Folk*. Available at: <https://lawforcomputerscientists.pubpub.org/>.
- Chinmayi, A. (2020). *AI and the Global South: Designing for Other Worlds*. The Oxford Handbook of Ethics of AI. Available at: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190067397.001.0001/oxfordhb-9780190067397-e-38>.

Lecture 1.2: What Are General Digital and AI Trends Across Justice Systems?

This presentation will illustrate trends in digital transformation in the African context. The presentation will look at how AI systems that developed in the context of advanced economies interact with challenges related to access to justice on the African continent. It will consider the need to take lessons from challenges exposed in the Global North to caution African countries on the challenges of AI and if it can help strengthen access to justice for people who may often find themselves at the margins.

Lecture Script (WebVTT)

Lecture two, what are general digital and AI trends across justice systems? Welcome to this session where we talk about digital transformation and AI and why it matters for justice systems. Now in the past decade or so, there has been a lot of digital transformation and you can see this from e-government services. A lot of services have moved online, even transactions among people. People now pay for services using systems like mobile money and this was even more prominent during the COVID-19 pandemic, where a lot of systems moved online. We now had a lot more countries adopting things like electronic signatures to facilitate online hearings. We now have a lot of countries where even paying for government services happens through mobile money and so even during the COVID-19 pandemic, countries that had already started using digital systems, there was some catalysation and now we're having a lot of more countries have doing things like online hearings. We have a lot more countries that are adopting electronic signatures to facilitate virtual or online hearings. We also have a lot of documents being digitized, as well as all these systems that are reading legal texts and analyzing it in ways that even surpass what maybe humans were doing before. And so, with all this transformation, we must ask ourselves, is it all a positive thing or should we be looking at this as a cautionary tale? In considering digital transformation, AI and the justice sector, you find that there's a lot of optimism among developers of AI systems as well as technology vendors, and they are importing a lot of systems from mostly the Global North to Africa.

And we must ask ourselves how this then intersects with existing issues with access to justice in the continent. In the whole digital transformation, there has been a lot of optimism about the role of technology in resolving societal problems, but we must be cautionary because you find that even as we apply AI in issues like in access to justice, there are already existing issues that have nothing to do with technology. Take, for example, the issue of women and land. In countries like Kenya, for example, we find that the bulk of cases in the judiciary have to do with land. Yet women have historically had problems accessing justice or accessing land in general. This is not a problem of technology per-se, but a problem with the political and cultural systems, and therefore any AI deployment must be aligned to these existing issues. And therefore, as we even go through this digital transformation, we have to look at it as a cautionary tale because a lot of the digital transformation in the AI that we are getting is very reliant on technology, even in places where people do not have access to technology in the first place. There is also a lot of centralization of data and datafication of issues, which could be problematic because it could end up, we could end up merging issues in centralizing even the judiciary that was always independent and separate from other arms of government. And there are examples in countries where, for example, people require digital ID to access court systems, which could be problematic for, for example, if somebody wanted to go to court to seek for their rights to identity in the first place. And so, the question there is that any system that is being implemented, people have to be aware that they are already

existing issues, structural issues that need to be resolved before we apply technology as a solution because technology in itself cannot be a solution, but rather a tool to help in betterment of life.

And so, as we go through the digital transformation and use of more AI systems in the justice sector, it's important to think about issues such as ownership of technology. Can we develop technology that speaks in our language that helps to resolve our own problems? Can we develop technology that gives us the space to participate in our own scientific development and enjoy the benefits of science? And we can also take in lessons from the Global North, where they've already deployed AI in the justice systems for much longer. And we've already seen issues such as bias, such as systemic racism, and these are already issues that we can push on ourselves and make sure that we develop systems that are free from that. Because at the end of the day and from a social justice perspective, technology, or AI, for that matter, is only beneficial if it can help improve the lives of the people who are normally at the margins. And these are the people who come to seek justice from the courts.

Relevant Readings / Activities

- The Law Society UK (2018). *AI: Artificial intelligence and the legal profession*. Available at: <https://www.lawsociety.org.uk/en/topics/research/ai-artificial-intelligence-and-the-legal-profession>.
- Margetts, H. and Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the WebPhil. Trans. R. Soc A.3712012038220120382. Available at: <http://doi.org/10.1098/rsta.2012.0382>
- Ogonzo, F., Gitonga, J., Wairengi, A.. Utilizing AI to Improve Efficiency of the Environment and Land Court in the Kenyan Judiciary. Available at: <http://ceur-ws.org/Vol-2888/paper9.pdf>.

Lecture 1.3: How is the role of judicial operators in upholding the rule of law transforming in the age of AI?

This presentation will define the notion of the rule of law. It will explore how AI development and deployment worldwide is impacting the role of judicial operators and justice systems in upholding the rule of law. It will dive deeper into the role of judicial operators in the digital era as well as the competencies required of today's judicial operators to navigate digital transformation.

Lecture Script (WebVTT)

How is the role of judicial operators in upholding the rule of law transforming in the age of AI? I am Katherine Forrest, and today I'm going to be speaking to you about why AI matters in the justice system and how it's impacting judicial operators. I want to start by talking about how AI today in the justice system is not the stuff of science fiction. It is being used today and every day throughout the justice system worldwide. And we must, as judicial operators, understand the way in which the rule of law is impacted by the use of AI. So, let's talk about first what the rule of law is, of course, and then how AI is impacting that. Judges and judicial operators are tasked with upholding the rule of law, of determining when human actions are in breach of the law or when human actions are not in breach of the law. AI is software. It is sophisticated software that learns, it learns from patterns and probability.

But AI is actually made by human progenitors. It is designed by humans. It is created by humans. The factors that go into making an AI tool are chosen by humans. The data sets are created and then also chosen by humans. So how is AI used today by judicial operators? It's used in the investigation phase. It's used in the policing phase. It's used in the phase of identifying voice evidence, DNA evidence, physical evidence, video evidence. It's used today to determine whether or not a particular individual might actually comply with the terms of release or not. It's used to determine and predict judicial decision making. So, AI tool is deployed widely today. What do judicial operators have to know and understand to be able to bring fairness and accuracy to the utilization of AI? First, human judicial operators have to understand what AI tools are being used. AI tools again are software, and not all software is the same. AI tools are being used, as we've said, to predict probabilities and patterns. So, the very first task for a judicial operator is to understand; is an AI tool being actually used? Secondly, the judicial operators need to understand when there are fairness and accuracy issues coming into play with the AI tools. How do they do this? Well, they've got to understand something about the creation of the AI tool, something about the input to the AI tool, something about the selection of the data set for the AI tool.

In upholding the rule of law, the means to the end matter. How a judge or a judicial operator arrives at a decision matter. Understanding the logic of a decision matters. That means that we must be able to have AI tools that are trained to be able to explain to the human operators how the AI tool is arriving at a decision, and therefore how the AI tool is able to achieve accuracy and fairness. AI tools, at the end of the day, can provide efficiencies. They can provide a way of reducing the need to deploy certain kinds of human resources. But AI tools also have the possibility of bringing to bear the human biases that the designers chose and or resulting in the judicial operators not understanding how the AI tools work. What we have to do is understand what is the logic? Are the AI tools in fact being used to arrive at decisions and to apply human discretion ultimately to the

decision making that is being arrived at in association and in conjunction with the AI tools to allow those tools to best fulfil their mission to help us as judicial operators not to take over.

Relevant Readings / Activities

- Susskind, Richard E. (2013). *Tomorrow's lawyers: an introduction to your future.* Oxford, UK: Oxford University Press

Lecture 1.4: What are the consequences of COVID-19 on the increasing digitalization of justice and access to justice?

This presentation will explore how the global pandemic and subsequent lockdown measures have affected the administration of justice worldwide and jeopardized access to justice. First, it will address the impact of the pandemic on individuals, institutions, and societies and how it affected the working of the courts. Second, it will consider how the pandemic and consequent national lockdown policies led to worldwide judicial paralysis, and how digital innovation became a permanent feature in judicial systems around the world. However, despite some positive outcomes, unanswered questions about guarantees of access to justice, due process, and judicial guarantees in relation to the use of digital communication tools remain. The presentation concludes with recommendations to address the digital divide and its ramifications for access to justice.

Lecture Script (WebVTT)

Lecture four, what are the consequences of COVID-19 on the increasing digitalisation of justice and access to justice? Hello. This presentation will explore how the global pandemic and correlated quarantine measures have affected the administration of justice worldwide and jeopardized the right of societies to a functioning and independent judiciary. Among the different conclusions that are included in the report I have presented to the Human Rights Council, there are three especially important ones.[1] First, that the impact of the pandemic on individuals, institutions and societies is unprecedented. Lockdowns in isolation and social distancing measures have affected severely the work of the courts, delayed its proceedings and the provision of legal assistance. Second, innovation in tele-workings have become permanent features of the working environment in judicial systems worldwide.

The use of virtual means for various types of proceedings has provided a positive response to widespread and prolonged judicial paralysis. However, there are unanswered questions about guarantees of access to justice, due process and judicial guarantees in relation to the use of digital communication tools. And third, the digital divide. Difficulty or the impossibility in accessing the internet in some forms of information technology make it challenging to ensure equitable access to judicial systems. Insufficient geographical coverage and access to technology, and the lack of training for operators and users gravely undermine access to justice. Among the several recommendations I have presented, there are three that are especially oriented to the digital divide.

First, countries should promote appropriate policies to achieve universal access to the internet and information technologies that bridge the digital divide and ensure real and effective access to justice by guaranteeing connectivity and availability of the necessary software and equipment, as well, of course, as adequate training. Consideration should be given specially to developing accessible and differentiated information materials for vulnerable groups.

Second, countries should provide tax and fiscal incentives to boost private and public investments, both to increase broadband and 5G supply and to subsidize its access in the poorest sectors, including, for example, infrastructure sharing between providers. Widespread 5G in first and second tier urban centers, according to World Bank estimates, would cost 0.17% of GDP per year. Much less than the several percentage points of investments normally expended in transportation energy.

[1] United Nations Human Rights Council (UNHRC) Report by Special Rapporteur Diego García-Sayán (9 April 2021) UN Doc A/HRC/47/35. Available at: <https://www.undocs.org/A/HRC/47/35>

Finally, third, the technology used in providing judicial services should ensure respect for the privacy, confidentiality and security of the information transmitted.

The privacy of attorney-client interactions must be strongly and especially guaranteed for due process.

8.1.5 Module Assessment

1. What is the potential impact of artificial intelligence?
   a) Improving lives and livelihoods worldwide
   b) Advancing progress towards the UN Sustainable Goals
   c) Surfacing risks related to bias, discrimination, transparency, and autonomy
   d) All of the above

2. Which legal concern is not linked to AI development and adoption?
   a) Fairness
   b) Versatility
   c) Accountability
   d) Transparency

3. What brought about digital transformation worldwide?
   a) Increasing computing power
   b) Increasing availability of data
   c) Both increasing computing power and availability of data
   d) None of the above

4. How is the role of a judicial operator different in the age of AI?
   a) It is not different
   b) Judicial operators should be able to develop AI systems
   c) judicial operations should have basic understanding of AI's potential and limitations
   d) Judicial operators should solely rely on AI for making decisions

5. What is the digital divide challenge for justice?
   a) An emerging challenge to ensure equitable access to judicial systems worldwide
   b) A separation of cases into digital and non-digital categories
   c) A movement to prevent digital transformation for the administration of justice
   d) A split between judges who use AI in courts and those who do not

8.2 MODULE 2: AI ADOPTION ACROSS JUSTICE SYSTEMS

8.2.1 Module Focus

The adoption of AI across justice systems can increase the effectiveness of judicial services for citizens, facilitate better access to justice, and decrease costs. Along with the many benefits of AI adoption in the justice context, this module will present the associated risks and demonstrate how judicial operators can assess the benefits and risks of AI adoption. Next, it will focus on examples of responsible AI adoption in justice contexts. Examples include AI's application for civil and criminal investigations (i.e., risk assessment, docket management, case management, jury selection, contract review, electronic discovery, legal research) and law enforcement (i.e., digital evidence, facial recognition, predictive policing). It will conclude with a discussion on the potential of AI for the Supreme Court of India and different opportunities and risks for the country's judicial system.

8.2.2 Learning Objectives

From this module, participants will:

- Acquire an understanding of the benefits and risks of AI adoption in justice systems
- Familiarize themselves with concrete examples of AI adoption for civil and criminal investigations, as well as law enforcement
- Contemplate the different rates of AI adoption in different contexts, including the Supreme Court of India

8.2.3 Lecturers

Hon. Benes Z Aldana | The National Judicial College, President; Former Chief Trial Judge of the Coast Guard | USA

Jhalak Kakkar | National Law University of Delhi, Executive Director | India

Justice Dhananjaya Y. Chandrachud | Supreme Court of India, Judge and Head of e-Courts | India

8.2.4 Structure

Lecture 2.1: What are the benefits and risks of adopting AI in justice contexts?

This presentation will provide an overview of the benefits of the use of AI for justice, focusing on access, quality of judicial services and improved efficiency. It will acknowledge the risks of AI and the challenges judicial operators must consider when using AI systems. Consequently, it will propose risk assessment strategies for judicial operators to make informed decisions about AI applications. The presentation will identify a set of questions intended to illustrate concerns related to the rollout of AI in courts.

Lecture Script (WebVTT)

Welcome to Module two AI adoption across justice system. Lecture 1: What are the benefits and risks of adopting AI in justice contexts? Hello, I'm Benes Aldana, president of the National Judicial College, the National Judicial College is the leading provider of judicial education in the United States and trains judges and other judicial branch employees, both in the United States and around the world. We are pleased to participate in these important initiatives involving artificial intelligence and its application to improving access to justice for all. It is my privilege to introduce Module two, which will focus primarily on practical illustrations of how AI can be responsibly adopted in judicial contexts. Technology in our lifetimes has evolved rapidly.

Fifty years ago, when one spoke about computers, one referred to a mainframe machine that occupied an entire building and took hours to process data inquiries. Not only has that landscape in computer technology altered almost beyond recognition, but the actual pace of technological change continues to accelerate.

For that reason, this module and indeed this entire MOOC can only serve as a snapshot of the state of the art in 2021.Five years from now, the nature and capabilities of AI may be indeed almost certainly will be very different. Much of the literature in AI and the law is decidedly futurist in nature and speculates about the effect of AI developments that do not currently exist and which may or may not ever come about. The two variants to which I refer are a rule-based approach; decision making based on complex algorithms, designed to enable machine learning. Machine learning, which is based on detecting useful patterns in large amounts of data, is the predominant form of AI currently in vogue. It has been used successfully to enable machines to play chess and Go translating languages and more recently, driving cars. It has also been successfully deployed in products that produce legal documents for lay-people like LegalZoom. The other approach similar cases warrant treatment approach.
This is brute force calculation using data to analyze massive amounts of public and government information in a matter and at speeds that are beyond the capabilities of human beings. This approach looks for patterns in large bodies of data to identify relationships and correlations from which the computer can draw conclusions. This is useful for prompt and effective resolution of large volumes of relatively simple cases. AI applications in our society have been made possible by the creation of massive data polls, tracking behaviour, a large amount, and numbers of people such as location data on mobile phones and online shopping and browsing patterns. Data reliant AI has also been enhanced by programming that mimics the way the human brain processes information. These so-called neural networks employ various mathematical complex forms of regression analysis to evaluate data with the goal of assigning what the programming has defined as good results. Random

forests, as the name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual decision tree in the Random Forest generates a predicted result, and the forest operates as a kind of consensus decision making heuristic so that the result with the most decision tree support wins. In order for this to perform well, the programming for the decision tree models has to work better than random guessing, and the predictions made by individual decision trees must have low correlations with each other.

Deep learning is an important element of data science, which includes statistics and predictive modelling. Whereas traditional machine learning algorithms are linear, deep learning algorithms are stacked in the hierarchy of increasing complexity and abstraction in much the same way as a child learns language and to distinguish between different categories of objects. Implementation of this type of AI is only feasible where sufficiently large bodies of data exist. Datafication has made that possible in many areas because technologies such as GPS or the internet can create data trails previously unimaginable. Nonetheless, in some areas, enough data does not exist, and machine learning cannot find statistically valid relationship.

China, where dockets are measured and tens of millions, is a good candidate for implementing this technology, as are India and Brazil. But we must be cognizant of the fact that in some societies there may be countervailing policies. For example, the marshalling of large amounts of personal data creates an inherent conflict between privacy concerns and the goals of machine learning. Privacy advocates often advocate for the non-collection of, or frequent purge of, personal data. Some recent court uses of AI. In Brazil, there is a backlog of 78 million lawsuits waiting for final and binding decision. Notably, Brazil spends almost two percent of its GDP on the judicial branch, which is more than any other country in the world.

There have been many attempts to tackle the crisis of the Brazilian justice system, but thus far all of them have failed. For instance, despite the efforts to promote alternative dispute resolution (ADR) mechanisms in Brazil, to date only 10 percent of the lawsuits end in a settlement. Implementation of a technological solution for Brazilian courts seems to be the only way out. The Brazilian National Council of Justice authorized the ninety-two courts it oversees to adopt AI developed by each court's own internal team. Many of these are reportedly geared toward identifying the themes of precedents so as not to make different decisions for similar cases. To alleviate this currently uncoordinated AI landscape, work is underway to design a collaborative governance structure that would strategically integrate all AI initiatives in the Brazilian judiciary. A recent, somewhat controversial development designed to expedite the processing of appeals before Brazilian Supreme Court. In 2019 alone, more than fifty thousand cases were filed in that court. The experiment AI solution goes by the name of VICTOR. Currently, appellants are not informed when VICTOR is used since its pilot version is randomly picking up appeals to evaluate. Some people contend that this is a violation of Brazilian laws. The most recent of which the Brazilian data protection law, which entered into force in May 2021, provides that automated decision making should be fair, transparent, and informed. How this issue will be resolved remains unclear.

In Estonia, lower valued claims have been delegated to an online powered exclusively by AI. In some respects, like AI mediation. Note that Estonian society may be well-suited to the concept of having a robot judge in some circumstances. Partly because 1.3 million residents already use a national ID and are use of online media

services such as e-voting and digital tax filing. Government databases are interconnected, and Estonian residents can also check who has been accessing their information by logging into a government digital portal.

In China, a typical local judge handles 400 cases per year. In the aggregate in 2018 local courts accepted 28 million cases concluded 25 million. The Supreme People's Court in China handles 35 thousand cases per year and is in the process of centralizing data 120 million legal documents. Contrast the U.S. Supreme Court in which 7 thousand to 8 thousand cases are filed each year and about 70 to 80, approximately one percent, are heard. The Chinese government is committed to rolling out smart courts as part of its push to be pre-eminent in AI generally and as part of a desire to improve the efficiency of and promote confidence in their court system. Internet courts have been established in China within the past two years. For example, there is the Beijing Internet Court for Commercial Disputes, which features a robot judge. One logs in to the court's website and a robot judge will identify the nature of the claim and once can also upload data evidence.

In Norway, the Law Commission on Public Administration Act, appointed in 2015, issued a report in 2019 recommending automated decision-making in administrative proceedings in order to improve efficiency, particularly where case volumes are large. Automated proceedings can also enhance implementation of rights and obligations, for example, by automatically making decisions that grant benefits when the conditions are met. This can particularly benefit the most disadvantaged in society.

In India, as we will hear from Justice Chandrachud, The Supreme Court of India has introduced a system of AI that would assist in better administration of justice delivery. This is known as a Supreme Court Portal for Assistance in Courts Efficiency or SUPACE. The judiciary has identified the potential of AI in cataloguing large numbers of judicial prior decisions for better handling of case data, whether for understanding the factual matrix of particular cases or dynamic research of precedents. The court's AI committee has resolved to implement the SUPACE tool on an experimental basis, with judges dealing with criminal matters in Bombay and Delhi High Courts. India has reportedly lodged a Supreme Court mobile app that will translate the court's judgments into nine regional languages. In addition, India is using AI for the resolution of low-level offenses such as traffic infractions.

In the United States AI is used for risk assessment in some jurisdictions for pre-trial release determinations in criminal cases. But this is controversial, predominantly for two reasons. Allegations of racial disputes in outcomes and due process of law concerns. In Wisconsin the court held that algorithmically generated risk may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community. The case challenge Wisconsin's use of closed source risk assessment software. In the sentencing of one Eric Loomis, Loomis was driving a car that had been used in a shooting. He was arrested and pleaded guilty to eluding an officer. In determining his sentence, the judge looked at his criminal record, as well as a score assigned by a tool called COMPAS, developed by a private company, which uses proprietary algorithm that consider some of the answers to a 137-item questionnaire. COMPAS has been used in other states as well, including New York, California and parts of Florida. COMPAS classified Loomis as a high risk of re-offending, and Loomis was sentenced to six years in prison. The gravamen of the appeal was at the trial judge in considering the outcome of an algorithm whose inner workings were secretive could not be examined and therefore violated due process. This slide seeks to encapsulate what is essentially common ground for

most judicial systems, whether common law or civil law, in terms of identifying the core functions of the judiciary. Notwithstanding that common ground, there are significant differences as one goes from one legal system to another in terms of public perceptions of the judiciary. The legal effects of their judgments and even perception differences within a single legal system when it comes to judges serving a different types or levels of courts. These differences suggest that implementation of AI may be better suited to some judicial systems than others. But that is not the end of the challenges facing rollout of AI in the courts.

The next slides identify a set of eight questions intended to illustrate these challenges. These are offered for your consideration as food for thought. How effectively can AI deal with evidentiary considerations, such as judging credibility of witnesses, determining whether evidence probative value is outweighed by its prejudicial impact? What is the impact of an AI judge on public participation in the judicial process and hence public acceptance of its legitimacy as parties, spectators, witnesses, jurors, victims of crimes? Can AI overcome the black box characterization of lack of transparency? Can AI effectively partake of the majesty and dignity of the judiciary as a branch of government? Will AI be perceived as capable of presiding over disputes with a depth of knowledge, experience and emotional intelligence of human judges? Can AI lawfully be deployed in criminal matters other than perhaps traffic infractions? Is the role of judicial officers in safe-guarding human rights changing as a result of AI? What do issues related to algorithmic bias mean for the administration of justice? Can AI be deployed consistent with the protection of personal data? In that regard, is AI a viable alternative for disputes involving European Union nationals, such as, is it precluded at least without that person's consent, by operation of the General Data Protection Regulation?

We hope you have enjoyed the short introduction to some of the issues highlighted by Module two and that you will keep them in the back of your mind as you listen to our next two presenters.

Relevant Readings / Activities:

- Faggella, D., (2021). *AI in Law and Legal Practice - A Comprehensive View of 35 Current Applications.* [online] Emerj. Available at: <https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/>.

Lecture 2.2: How can AI be used for civil and criminal investigations, as well as for law enforcement?

This presentation will provide information about AI in the judiciary and law enforcement by giving practical examples of potentials and risks. It will consider examples such as AI's use for predictive policing through technologies like facial recognition and the use of digital evidence in prosecution. After examining the related concerns, the video will present mechanisms that could be adopted to assess the risk of AI systems and to regulate them effectively.

Lecture Script (WebVTT)

How can AI be used for civil and criminal investigations as well as for law enforcement? Hello. I will be talking to you about the use of AI in the judiciary and the law enforcement. Artificial intelligence systems are increasingly playing a significant role in the discharge of critical democratic functions ranging from justice delivery to welfare delivery to its adoption by law enforcement. For instance, administrative functions, which form a crucial part of any judge's role, can be simplified with the use of AI, whether it's for case management or scheduling of hearing dates or docket management. Globally, AI systems are also being used within the judicial system to undertake more complex tasks, such as helping with jury selection, contract review and e-discovery. Increasingly, we are seeing the development of AI systems that are undertaking highly conceptual tasks, like providing support for legal research, undertaking risk assessment of offenders and even making decisions in certain types of cases.

On the law enforcement side, we are seeing the adoption of AI systems for predictive policing through technology such as facial recognition technology, as well as the use of digital evidence in prosecution. Adoption of facial recognition technology has given rise to significant concerns around its impact, both on the privacy of individuals, as well as its broader negative implications for society at large. And globally civil society organisations have been calling for a moratorium on the use of such technology. While justice delivery and law enforcement must move together with modern times and harness technology, this has to be done keeping in mind the challenges and risks that arise with the adoption of such technologies and ensuring that any such technology used by the judiciary or law enforcement is in consonance with human rights standards, constitutional values as well as rule of law.

This adoption should be done on the basis of extensive public consultations with a wide cross-section of stakeholder groups from civil society, academic groups, technical experts, public sector representatives, as well as the private sector. Besides consultation, openness and transparency in the technology being adopted is paramount. Often AI systems that are used by the judiciary and law enforcement are designed by private entities and are proprietary technologies. Given that the use of AI in the judiciary can introduce bias that results in denial of justice, discriminatory risk assessments and wrongful convictions, careful thought has to be given to the approaches used to develop AI for the judiciary. The judiciary must enable the scrutiny of all such AI systems by the public, as well as experts to ensure that they engender high levels of trust in the judicial system and ensure that the deployment of these technologies aligns with constitutional values.

At present, many countries such as Brazil, the US, the UK, India, Argentina, Colombia, Estonia, countries from all around the world are experimenting with the deployment of AI systems in various parts of the judicial process, ranging from administrative functioning to supporting decision making by judges. Some of the

examples of the use of AI and judiciary include the COMPAS software, which is an AI system in the US used to assess the chances of recidivism.

The UK is testing a similar AI system called HART, which works on a similar model to predict the risk of recidivism and thereby recommend low risk offenders to counselling programmes. However, the use of these kinds of systems have given rise to significant concerns around amplifying societal bias in sentencing decisions due to historically biased datasets that these systems have been trained on. Another interesting example of an AI system being used by the judiciary is the testing being undertaken in Brazil around a tool called VICTOR. This system conducts preliminary case analysis based on national language processing tools to exclude extraneous appeals and ensure that only questions that are truly relevant to society are heard for appeal. The system has been subject of much criticism, for the lack of transparency in decision making and for giving potentially distorted results. Countries such as Estonia have initiated pilot programs to use AI to settle small claims disputes.

The UK and US are experimenting with the use of Chatbot lawyers who can help people challenge parking fines and violations free of cost. There are many benefits that come with the leveraging of AI technology in the judicial system, such as better case management, docket management, enhanced e-discovery. However, as we have just discussed a number of examples, there are also certain challenges and concerns which come with the adoption of such technologies by the judiciary. Let us discuss some of these challenges and mechanisms that can be adopted to assess the risk of these systems and regulate these systems effectively. A key concern with the adoption of AI systems is the lack of transparency and explainability due to something known as the black box effect, as well as concerns around bias and fairness of such systems.

This could affect not only the legitimacy of the system, but also public trust in the use of such technology and serious thought needs to be given to the various aspects of judicial administration and decision making that these AI systems are integrated into because we have to ensure that justice delivery is not negatively impacted in any manner whatsoever. Additionally, there are concerns around the collection and use of significant amounts of data, a significant proportion of which may be personal data that is being used to train these AI systems. A robust personal data protection framework is essential to protect the fundamental rights and freedoms of litigants.

Besides this, the judiciary should consider mechanisms such as setting up a data protection and privacy office to deal with the governance mechanisms around such data. One of the most significant challenges on the use of AI is the tendency of AI systems to amplify existing biases in our society, as well as enhance discrimination against vulnerable and marginalized populations. The risks of this bias and discriminatory outcomes are magnified in the context of justice delivery and can have significant outcomes and consequences on the rights of those involved. AI systems are usually trained on existing datasets, which tend to be historically biased and discriminatory. Besides this, prejudices held by programmers involved in developing the AI system, exclusion of historically marginalized populations from datasets, information gaps all contribute to introducing bias in AI systems. It is therefore essential to develop regulatory and evaluative frameworks that require the adoption of risk assessment mechanisms and ensure that the design of AI systems is in consonance with our fundamental rights and constitutional values. Besides regulatory frameworks designed by legislatures around the world, it is equally important that the judiciary actively identify best practices. and evaluation mechanisms to adhere to in the design, as well as the deployment of the AI systems in the judicial system.

Some key areas include identifying the context in and tasks for which it is suitable to deploy AI systems of the judiciary, standard setting on the design of such systems. Rules requiring regular risk assessments, algorithmic audits and ex ante and post facto fundamental rights impact assessments. Guidelines on the use of AI tools in courts and other relevant human oversight mechanisms. Finally, the design and deployment of algorithmic systems by the judiciary must also account for the ongoing challenges of access to internet and digital infrastructure that we see, especially in the Global South.

Citizens that do not have access to digital infrastructure could be excluded from accessing online dispute resolution tools and mechanisms that are being integrated into judicial systems to simplify processes, as well as reduce legal costs. Hence, the adoption of AI and other technology tools by the judiciary must be cognizant of the risks of enhancing existing inequalities. For those who do not have access to digital infrastructure or are not digitally literate. Globally, there's a great deal of work which is being done by legal policy and technical experts and researchers around developing regulatory frameworks and evaluating mechanisms to adequately protect personal data, address the challenges of explainability, transparency and bias that we see emerging from the use of AI systems. It is imperative that the judiciary continue to engage with these stakeholder groups to ensure that equitable and human rights consonant AI systems are adopted. While the adoption of AI systems can be considered to enhance the efficiency of the judiciary, it needs to be done in a way that ensures that the decision making and autonomy of judges is not eroded in any way whatsoever, and that the delivery of justice and upholding of constitutional rights is given primacy.

Relevant Readings / Activities

- Explore the AI Risk Assessment Tool: https://bja.ojp.gov/program/psrac/selection/risk-assessment-landscape#acxbh
- Chelioudakis, E., (2020). *Risk Assessment Tools in Criminal Justice: Is There a Need for Such Tools in Europe and Would Their Use Comply with European Data Protection Law?* Australian National University Journal of Law &Technology (ANU JOLT), Vol. 1, Issue 2, 2020, Available at: <https://ssrn.com/abstract=3743757>.
- Pearsall, B., (n.d.) *Predictive Policing: The Future of Law Enforcement?* Available at: <https://www.ojp.gov/pdffiles1/nij/230414.pdf>.
- Directorate-General for Justice and Consumers (European Commission) (2019). *Liability for artificial intelligence and other emerging digital technologies.* Available at: <https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en/format-PDF>.
- Williams, Rebecca A., (2018). *Rethinking Deference for Algorithmic Decision-Making*. Oxford Legal Studies Research Paper No. 7/2019, Available at SSRN: <https://ssrn.com/abstract=3242482> or <http://dx.doi.org/10.2139/ssrn.3242482>.

Lecture 2.3: What is the potential of AI technologies for the Supreme Court of India?

This presentation will consider AI's potential for the justice system in India by exploring both the opportunities and risks in adopting AI for civil and criminal litigation, as well as law enforcement. To contextualize the needs of the judiciary and the uses of AI in India, Russia, China, Mexico, USA, Abu Dhabi and Argentina are discussed to illustrate the ways in which AI is being adopted in justice systems. Proponents argue that the use of AI will lead to transparent, impartial judicial outcomes by eliminating subjective bias. However, overreliance on AI, without human intervention, may lead us to a paradoxical situation where in the quest for equality, AI-generated judicial outcomes may overlook the very factors that cause gender, religious, cast and class discrimination in India. The lecturer concludes that the pursuit of justice needs to build on experience and re-imagine societal structures with the interface of law and technology.

Lecture Script (WebVTT)

What is the potential of AI technologies for the Supreme Court of India?

The past year has been an incredible learning experience for the Indian judiciary in adapting to new challenges through the use of technology and reimagining the structures of governance. During the course of my presentation, I will touch upon the experience of integrating technology, specifically artificial intelligence in the administration of justice in India and the opportunities it offers for the future, as well as its limitations.

The Indian judicial system's tryst with technology began in 2005 with the launch of the E-Courts Project to ensure technology enabled courts. First, the creation of infrastructure for 17,000 courts at the district level. 25 High Courts and the Supreme Court of India. Second, during the COVID-19 pandemic from March 2020 to March 2021, the high courts in the district courts across India have heard a total of 8.64 million cases through video conferencing facilities. Richard Susskind, in his book titled 'Online Courts in the Future of Justice', notes that "technology should bring about a transformation and not just automation". The E-Courts project also laid emphasis on increasing access to justice through efficient court management. Contextualizing the judicial needs in India, I must highlight some salient points.

Let us contextualize the needs of the judiciary in India. First, we need to address the requirements of a country with a population of 1.3 billion people. The Indian judiciary consists of a three-tiered structure with courts in every district of the country, a High Court in every state and the Supreme Court at the federal level. Second, let's look at the pendency of cases in India. 39.72 million cases in the district courts, 5.73 million cases at the High Courts, and 69,476 cases in the Supreme Court of India. Uses of AI in India. First, we have virtual courts, which provide for online dispute resolution for settling traffic violations and other petty offences, which eliminates the physical presence of the litigator in the court. Second, the AI Committee of the Supreme Court has launched a software which relies on natural language processing to translate orders and judgments in English to regional languages. Third, another AI portal was launched for data mining, legal research and projecting case progress. Fourth, we have developed AI solutions for automation of the scrutiny of cases in the Supreme Court by categorizing cases into different subject matters, identifying cases involving common questions of law and marking objections to defective petitions.

Around the world, AI has been used in providing automated legal advice and timelines for case resolution in courts in Russia, China and Mexico. In predicting case outcomes in the US, Abu Dhabi and Argentina and in assisting substantive decision making in the US. I believe India can also draw from some of these experiences.

Government as one of the biggest litigants can use predictive technologies to assess whether it needs to pursue certain cases at all. Transcription of court proceedings can assist in criminal trials. Speech translation such that proceedings, which are conducted in English, can be translated into regional languages and we can increase access to justice by providing automated legal responses to citizens in rural areas who lack legal literacy. There are certain risks about AI. Substantive justice tools which have been used in the UK and the US in determining sentencing in criminal cases have not been adopted in India. These tools have been severely criticized for racial profiling and automation bias. The use of AI to audit judicial behaviour or for ascertaining predictive outcomes, may also lead to profiling of judges and bench hopping by litigants for a favourable outcome. The proponents of AI argue that it can lead to transparent, impartial judicial outcomes, eliminating subjective bias.

However, over-reliance on AI without human intervention may lead us to a paradoxical situation where in the quest for perfect equality, AI-generated judicial outcomes may overlook the very factors that cause gender, religious, caste and class discrimination. In conclusion, the possibility of advancing justice with the advent of technology is unravelling day by day. In 1881, Justice Oliver Wendell Holmes said that "the life of the law has not been logic, it has been experience". The pursuit of justice requires us to build on this experience and re-imagine our societal structures with the interface of law and technology.

Relevant Readings / Activities:

- Bokil, M., Khare, A., Sonavane, N., Bej, S., Janarthanan, V., (2021). Settled Habits, New Tricks: Casteist Policing Meets Big Tech in India. Available at: <https://longreads.tni.org/stateofpower/settled-habits-new-tricks-casteist-policing-meets-big-tech-in-indiaL>.
- Sengupta, A., Jauhar, A., Misra, V., (2021). Responsible AI for the Indian Justice System – A Strategy Paper. Vidhi Centre for Legal Policy. Available at: <https://vidhilegalpolicy.in/research/responsible-ai-for-the-indian-justice-system-a-strategy-paper/>.

8.2.6. Module Assessment

1. Which regions have already adopted AI in their justice systems?
   a) North America
   b) Asia
   c) Europe
   d) Global

2. Which is not a current AI application for the administration of justice?
   a) Docket management
   b) Replacement of lawyers
   c) Support for legal research
   d) E-discovery

3. What is a 'rules-based approach'?
   a) Identification of human-constructed rules to regulate AI
   b) Elaboration of decision-making based on complex algorithms
   c) Employability of AI systems for docket management
   d) Adoption of AI technologies using global protocol

4. Which tools can help mitigate the risks of AI adoption?
   a) Human oversight mechanisms
   b) Algorithmic audits
   c) Risk assessments
   d) All of the above

5. What are the potential risks of AI in the Supreme Court of India?
   a) Automation bias
   b) Dismissal of physical courts
   c) Increase in crime
   d) Corruption

8.2.7 Use cases and examples

UK- Pyrrho Investments Ltd v MWB Property Ltd (2016)

Pyrrho Investments Limited v MWB Property Limited, is the first British case to consider the use of 'predictive coding' during the electronic (e-discovery) process of document disclosure.[2] Predictive coding is a process whereby software is trained to assess the relevance of copious quantities of electronically stored information. Although not completely replacing manual review (for example, to train the software and check relevance or privilege) the use of such technology can vastly cut down the cost of and time for large disclosure exercises. The judgment by the High Court of the United Kingdom approved the use of predictive coding software, involving only a limited manual review of the results, for the purposes of e-disclosure. This decision was based on several arguments; first, several jurisdictions have proven such software to be useful in appropriate cases, second, the parties agreed on use of the software, third, there is no evidence suggesting a lower degree of accuracy, or a more consistent and cheaper alternative, and finally, there is no legislation prohibiting its use. This was the first time that an English court expressly considered and endorsed the use of software to provide disclosure in civil procedure.

***Canada - Drummond v The Cadillac Fairview Corp. Ltd.* - Ontario Superior Court of Justice (2018)**

In the 2018 case *Drummond v The Cadillac Fairview Corp. Ltd.* the Court commented on the importance of AI in assisting lawyers in their legal research.[3] Ltd. Cadillac Fairview objected to the disbursement for legal research costs.[4] The judge denied the reimbursement and decided that the hours the lawyers spent on the law database should be included in the counsel fees, stating that: *"in reality, computer-assisted legal research is a necessity for the contemporary practice of law and computer assisted legal research is here to stay with further advances in AI to be anticipated and encouraged. Properly done, computer assisted legal research provides more comprehension and accuracy."* Quoting Shane Katz, who acted for plaintiff in the case, *"legal research is an important practice of law, and using AI can help make the research more efficient, as well as let lawyers use their time better"*.[5]

Canada - Cass v 1410088 Ontario Inc. (2018)

In the subsequent case of *Cass v 1410088 Ontario Inc.* the Ontario Superior Court of Justice commented on the use of AI in reducing preparation costs, notably, the number of paid resources used to engage in legal research. The Court of Justice denied a legal research disbursement for several reasons and mentioned that *"if AI sources were employed, no doubt counsel's preparation time would have been significantly reduced"*.[6] The courts' longing for the implementation and exercise of legal research tools such as AI, in reducing client costs and creating a more efficient legal process, is not new. Since the advent of computer technology, lawyers have rightfully capitalized on its ability to assist in organization, quick retrieval of case law and case summaries, and more recently, the use of tools like e-discovery to reduce client costs and the average lawyer's workload in the preparation of affidavits and examinations for discovery.

India – AI in the Judiciary

In India, AI is already being used in virtual courts that provide for online dispute resolution to settle traffic violations and other petty offenses, which eliminates the physical presence of the litigator in the court.[7] Further, the AI committee of the Indian Supreme Court launched a software that relies on natural language processing to translate orders and judgments to regional languages.[8] Recently, other AI portals were launched for data mining, legal research and predicting case progress.

---

[2] *Pyrrho Investments Ltd v MWB Property Ltd*, England and Wales High Court, EWHC 256 (Ch), 16 February 2016.
[3] *Drummond v The Cadillac Fairview Corp. Ltd.*, ONSC 5350, Ontario Superior Court of Justice, 13 September 2018. Jo. *Drummond v The Cadillac Fairview Corp. Ltd.*, ONSC 4509, Ontario Superior Court of Justice, 10 July 2018.
[4] *Drummond v The Cadillac Fairview Corp. Ltd.*, ONSC 5350, 2018, §9
[5] *Drummond v The Cadillac Fairview Corp. Ltd.,* ONSC 5350, 2018, §10 jo. Intahchomphoo Channarong, Vellino Andre, Gundersen Erik, et al., *References to Artificial Intelligence in Canada's Court Cases*, Cambridge University Press, 11 May 2020. Available at <https://www.researchgate.net/publication/341291571_References_to_Artificial_Intelligence_in_Canada%27s_Court_Cases>.
q
[6] *Cass v 1410088 Ontario Inc., ONSC 6959*, (CanLII) 51145/09, the Superior Court of Justice Ontario Canada, 22 November 2018, §34.
[7] Kartik Pant; *AI in the courts*, The Indian Express, 2021. Available at <https://indianexpress.com/article/opinion/artificial-intelligence-in-the-courts-7399436/>.
[8] Shanti S.; *Behind SUPACE: The AI Portal of The Supreme Court of India*, AIM, 2021. Available at < https://analyticsindiamag.com/behind-supace-the-ai-portal-of-the-supreme-court-of-india/>.

Brazil – AI in the Judiciary

The AI tool VICTOR is the result of the Brazilian Supreme Court's initiative to deepen the discussion on the applications of AI in the judiciary. It takes care of the largest and most complex AI project of the judiciary and of the entire Brazilian public administration. In the initial phase, VICTOR can read all the appeals that go up to the Supreme court and identify which ones are linked to certain topics of general repercussion.[9] This action represents only a small but important part of the initial stage of processing appeals in the Court, but it involves a high complexity level in machine learning. In the Judiciary of Pernambuco, an AI system examines new tax enforcement actions and decides which ones are in accordance with procedural rules and which ones are dismissed due to the statute of limitations.[10] The Federal Appeals Court for the 3rd Region started, on July 3, 2020, implementing the AI program called SIGMA to assist in preparing reports, decisions, and judgments in the Electronic Judicial Process system. The use of AI in the judiciary was recently regulated by the National Council of Justice and will be done through SYNAPSES, a virtual platform that will centralize technology initiatives.[11]

Norway – AI in the Judiciary

The Law Commission on Public Administration Act issued a report in 2019 recommending automated decision-making in judicial administrative proceedings to improve efficiency, particularly where case volumes are large. Accordingly, "*Automated administrative proceedings can also enhance implementation of rights and obligations; for example, by automatically making decisions that grant benefits when the conditions are met. This can particularly benefit the most disadvantaged in society*".[12]

Estonia – AI in the Judiciary

Estonia's pilot AI judge came after the Estonian Ministry of Justice requested from its chief data officer Velsberg Ott to assist in designing a 'robot judge' that could decide small claim disputes of less than 7,000 Euros.[13] In concept, the two parties to a dispute upload documents and other relevant information to a court platform, and the AI issues a decision that can be appealed to a human judge. Estonia is among the countries with the highest ranks in the UN's 2018 and 2020 E-Government Development Index, as its citizens and public servants can access a wide range of services online using secure digital IDs, including making payments, accessing full health records, and internet voting. Today, 99% of the public services are available online 24/7, 30% of Estonians use i-Voting, and the country estimates the reduced bureaucracy has saved 800 years of working time. Up until now, lower-valued claims have been delegated to an online court powered exclusively by AI.[14]

---

[9] Becker Daniel and Isabela Ferrari; *VICTOR, the Brazilian Supreme Court's Artificial Intelligence: a beauty or a beast?* 2020. Available at <https://sifocc.org/app/uploads/2020/06/Victor-Beauty-or-the-Beast.pdf>.

[10] Fausto Martin De Sanctis; *Artificial Intelligence and Innovation in Brazilian Justice*, International Annals of Criminology (2021), 1-10. Available at <https://www.amb.com.br/publicacoes/artigo-artificial-intelligence-and-innovation-in-brazilian-justice/>.

[11] Ibid.9

[12] Weitzenboeck Emily M.; *Simplification of Administrative Procedures through Fully Automated Decision-Making: The Case of Norway, Special Issue in Simplification of Administrative Procedures – In search of Efficiency, Public Interest Protection and Legitimate Expectations*, published online MDPI, 7 December 2021. Available at <https://www.mdpi.com/2076-3387/11/4/149/htm>.

[13] Tara Vasdani; *Estonia set to introduce 'AI judge' in small claims court to clear court backlog*, The Lawyer's Daily, April 10, 2019. Available at https://www.thelawyersdaily.ca/articles/11582 jo. Saaqib Ahmad Malik; *Estonia Is Developing a Robotic Judge for Its Courts to clear Case Backlogs*, WE, August 2019. Available at <https://wonderfulengineering.com/estonia-is-developing-a-robotic-judge-for-its-courts-to-clear-case-backlogs/>.

[14] 2018 United Nations E-Government Survey jo. 2020 United Nations E-Government Survey

8.3 MODULE 3: THE RISE OF ONLINE COURTS

8.3.1 Module Focus

This module will deep dive into AI's deployment in one of the most prominent judicial contexts: the courtroom. It will examine trends over the past few decades from online dispute resolutions to online courts, and the salience they have gained as result of the global pandemic. It will also discuss technological, economic, legal, and socio-ethical constraints that prevent AI adoption from fully materializing. Through context-specific use cases, it will shed light on both the opportunities and the risks the same AI applications may have. It will look at two examples: the British Columbia Civil Resolution Tribunal and smart courts in China.

8.3.2 Learning Objectives

From this module, participants will:

- Familiarize themselves with the concept of online courts and the new possibilities for justice they offer compared to traditional courts
- Build an understanding of Online Dispute Resolution Systems (ODR)
- Consider the effects of the pandemic on the 'online courts' movement'
- Contemplate lessons learned from the British Columbia Civil Resolution Tribunal
- Explore what smart courts look like in China

8.3.3 Lecturers

Judge Isabela Ferrari | Second Regional Federal Court of Brazil, Federal Judge | Brazil
Professor Orna Rabinovich-Einy | University of Haifa | Israel
Shannon Salter | Civil Resolution Tribunal, Chair | Canada
Fuhio Sun | Supreme People's Court of China, Deputy Director of the Information Center | China

8.3.4 Structure

Lecture 3.1: What are online courts?

The presentation will introduce the module and its objectives. It will explain what an online court is and review the development of online courts from online dispute resolutions processes. Finally, it will provide an overview of the benefits and challenges associated with online courts with respect to quality and access to justice.

Lecture Script (WebVTT)

Welcome to Module three, the rise of online courts. My name is Isabella Ferrari, I'm a Brazilian federal judge and in this module, we're addressing the rise of online courts.

In this short video we're answering 5 central questions. What is, after all, an online court? How do they differ from courts that simply operate online? What are Online Dispute Resolution systems? How can we adopt AI tools to help judicial operators? And finally, what are the pros and cons of the online courts movement? The first thing we need to know is that online courts are much more than courts that simply operate online. When you talk about online courts, we're talking about a movement that is not only bringing courts to a new space, but especially re-thinking the way we offer justice and the way we handle conflicts. If you wanted to draw a line, we would realize that the inspiration to online courts came from the private sector. When platforms such as eBay, Amazon, PayPal, Alibaba, etc. decided to implement their own Online Dispute Resolution systems. And then we go to our third fundamental question.

What is an Online Dispute Resolution system? An ODR? ODR's are digital spaces where parties can convene to resolve their dispute (or to have it resolved). Just to give you an example, if you think about ODR's in these private platforms, they usually operate in the following way. Once there is a dispute, some of them provide information to one or both of the parties. This is important because we have a neutral source of information. Therefore, we put the parties in the same page, reducing the possible asymmetry of information between them as they start to have similar expectations concerning the result of the dispute. After that, the parties usually have the opportunity to negotiate the conflict directly. Sometimes, the information provided before is a sufficient stimulus to allow the parties to get to an agreement by their own. If it's not the case, many platforms have a face of online mediation in which a person or an automated system provide the box with suggestions of baselines to agreements. If even that does not work, the platforms usually have the case decided either by a human or a software in the decision sometimes is immediately executed (because many of these platforms have the credit cards of their users and the authorization to do so of course). As ODR tools in the private sector have been a huge success, judicial operators, usually pressed by the increase in litigation and backlog, they started to reflect on the way of adopting ODR inside their own field. So usually, the movement of online courts start for the countries in the following way.

In the first movement, some of the acts are available online, filling or responding petitions, doing hearings for example, and of course, the pandemic gave traction to this trend as you can see in Professor

Rabinovich's class. In my view, these courts that are simply available online are not properly online courts yet. We can actually talk about online court when ODR and therefore data analysis are added to the game.

When this kind of tools are implemented to help case management. So for example, when you have a system that makes predictions regarding the time of the lawsuit or the rate of the success that demand, or when the system offers legal information to its users based on data analysis. The British Columbia Civil Resolution Tribunal, for example, that is a groundbreaking reference when we think about online courts and the topic of one of the classes of our module, has an automated system that provides its users with legal advice after they have answered some questions about their case. The system also provides models of documents that might be used by the parties, like pre-written notices.

In online courts, technology has also an important role in pushing agreements. Strategies like providing information and letting the parties negotiate their conflict directly or make them go through mediation before adjudication are important to reduce the backlog that overwhelms the court systems

In these courts new ways of communicating between the parties and the judge are usually officially allowed, such as email, WhatsApp or similar apps and video calls. To give you a better and more concrete idea of how ODR tools might be adopted on courts, therefore, giving birth to online courts included a class on the BCCRT taught by Shannon Salter, its chair, and materials on the Chinese smart courts.

Regarding ODR in courts, I would say that our biggest concern should be: fight the digital divide and care about cyber security. The digital divide is how we named the gulf between those who have access to computers and the internet and those who do not. Solving or at least narrowing the digital divide gap among legal services and beyond, demands a multi-stakeholder effort that should involve the judiciary, legislators, and government. Cybersecurity needs also to be a concern when you change the way you deliver justice, especially to guarantee the availability of the services as well as their integrity.

We already had news of attacks that hacked court systems, encrypting online lawsuits, and changing the content of judicial decisions. Though ODR is usually where countries start their internal movement towards online courts, this is not the only pillar of online courts. They advance to another level when we start to adopt AI tools to assist judges in their decisions.

But how does AI operate? Machine learning algorithms find hidden patterns on data they're fed with. They learn through this activity and can replicate human knowledge. As highlighted by Professor Lawrence Solum; in legal systems AI might be adopted to provide non-binding recommendation to judges, for example, regarding the precedents they might adopt in their decisions or even draft decisions.

To provide judgments that must be followed unless the human judge provides a legally sufficient reason to override the judgment and the most frightening possibility for many, replace human judges in some cases. Although this seems dystopic, it's good to remember that Estonia is already developing a robot judge to decide small claims. And if we think about that, it kind of makes sense. We're talking about lawsuits that involve not civil or political issues, but only proprietarian rights and of small amounts in

lawsuits that usually repeat themselves a lot. So there's not much room or need for human creativity, and it's not extremely dangerous if occasionally they're ruled in the wrong way.

What might also happen when we're talking about human decision making? Robot judges can also help us foster some of the rule of law values such as accessibility, efficiency, consistency. But it's impossible to talk about algorithmic decision making without mentioning its biggest risk: Bias.

Once algorithms learn based on the data they're fed with and considering this data is produced in the midst of the biased societies we live in. It is natural that they also learn and thus replicate this biased pattern, therefore targeting historically vulnerable groups such as women or black people. Bias is such a serious concern that we dedicate a full module to deepen this discussion. The next one. So, in a nutshell, the online courts movement is based on two pillars, integrated ODR into the judiciary and adopting AI tools to help judicial operators.

In a sense they are much more than courts that simply operate online. We're talking about re-thinking the way legal services are delivered. This seems to be a global movement fostered even more by the pandemic. Online courts bring new possibilities of fostering agreements, reducing mitigation, increasing the consistency in decisions, and promoting timely and predictable resolution of disputes, thus advancing the rule of law. At the same time, when you think about ODR, we must consider digital divide and cybersecurity concerns. And when we think about AI tools assisting or substituting judges, we must fight algorithmic bias that usually originates in the data algorithms are fed with.

Therefore, in the scenario of ongoing change and looking at the future of the judiciary, the fundamental question does not seem to be whether we are integrating technology in the judiciary, but where and how we're doing so and the safeguards we must take. In this debate you play a central role. We call you to use your voice. Thank you for being with me and enjoy the rest of the course.

Lecture 3.2: How has the pandemic affected online courts?

The presentation will consider the surge of online courts because of the pandemic and related lockdown policies worldwide. Even though online courts are not new, the pandemic transformed their scope and direction. After differentiating Online Dispute Resolution (ODR) between court ODR and remote ODR, the presentation will consider both the benefits and challenges. In the end, it will provide remarks on remote ODR, the way forward, and how to design a novel dispute resolution landscape.

Lecture Script (WebVTT)

Lecture two. How has the pandemic affected online courts? Hi, I'm delighted to speak to you about online courts and the impact of COVID.

When we think of courts, we typically imagine a large and impressive building and inside of it a judge dressed in a robe holding formal proceedings and delivering a recent decision at the end of it. The reality of courts, however, is quite different, and it's been dramatically different during COVID. With the onslaught of COVID many courts worldwide were shut down for significant periods of time and offered their proceedings online via video, audio and in writing. While this change was dramatic in scope and in the types of activities that were being performed online, the roots of online courts actually preceded COVID and started in the second decade of the 21st century. Nevertheless, the impact of Covid on the trajectory of online courts was dramatic. Many new courts embraced new technology and began offering online court proceedings for new types of disputes, embracing new types of proceedings across various mediums and in different types of platforms.

So, what are online court proceedings? Online court proceedings involve the use of digital technology to redesign and deliver court proceedings that were formally offered in person, in a courtroom, in the presence of a judge and the parties. Here at least some of the parties interact remotely asynchronously or synchronously. The roots of online courts are in the phenomenon of

Online Dispute Resolution or (ODR) which emerged in the 1990's in the e-commerce setting, where new types of disputes arose and had no effective avenues of redress. As online interaction grew and access to the internet shifted from PCs to smartphones, ODR also grew in scope and shifted from online private settings to offline ones, private and public. And the meaning of ODR changed. If initially in ODR, what we tried to do was to replicate, mimic traditional dispute resolution processes and move them online like negotiation, mediation or arbitration; now, ODR was being understood as the attempt to re-imagine all types of proceedings and focus them around user needs. So in terms of the courts, that has meant not to think of court proceedings as being designed by lawyers and for lawyers, but actually about making them accessible to the lay-person, allowing them to effectively navigate the system and reach fair outcomes. This has meant the court ODR has traditionally relied on asynchronous written communication. By contrast, remote proceedings which emerged during COVID relied mostly on video, and the motivation for the adoption was not so much the desire to enhance access to justice and make these processes centered around users. But the need to sustain court activity and address the accumulating caseload associated with COVID.

So online court proceedings come with obvious benefits, such as increased efficiency and reduced costs, right? Or you don't need to attend court physically. But where properly designed, they actually draw on what we call the fourth party, the technology that assists the Third party and the disputants not only to enhance efficiency,

but actually to enhance court capacity to make these processes more accessible and even to make them more fair, by using automation, tailoring information to parties needs, offering prefixed options and easily accessible language, and by studying the data that is accumulated through court ODR. Of course, online proceedings also came with many challenges and critiques. Some of these challenges were about the digital divide and the concerns, especially the parties coming from disempowered groups, would not be able to effectively steer their way through these processes. Also, concerns were raised about the lack of media richness and the increased privatization that comes with the reliance on platforms that offer court ODR.

In essence, what the critics were concerned about was that a court ODR would offer second class justice as opposed to the idea, or should we say, idealized alternative of traditional face to face dispute resolution. So what the critics and proponents of ODR were doing, was raising an old argument some of you may be familiar with, from the early days in which alternative dispute resolution was adopted in the courts. They were asking whether increased access came at a cost to justice. Whether added efficiency necessarily reduced from fairness, and the proponents of ODR were suggesting that perhaps the unique qualities of digital communication could actually reduce what has seemed like an inherent trade-off between efficiency and fairness, between access and justice. Some of the next few slides show you specific design choices made by online courts and tribunals that attempt to reduce the tradeoff between access and justice and between efficiency and fairness. Let's look at them and see to what extent they are successful in doing so. So, look at the entry point to online courts and tribunals and how different and attractive and easy to understand this looks as compared to traditional processes.

As you can see here, ODR is not a mere replication of familiar processes, but it's actually re-designing the court journey and offering a new type of process diagnosis, which allows parties to better understand their needs, their rights and their options. In the next two slides, you can see how technology assists parties in conducting negotiation and mediation online, as well as offers them automated tools to create legal documents such as letters and agreements. So as you saw, the onslaught of COVID has expanded both court ODR and remote proceedings. While these phenomena are tied and share many of the same characteristics, there are also distinct in important ways.

The motivation that brought them about and the goals they advance are very different. While court ODR places an emphasis on re-designing and re-imagining the court journey and enhancing access to justice, remote proceedings are focused on sustaining court activity and enhancing their efficiency. Therefore, they also rely on different mediums. While court ODR relies on asynchronous written communication, remote proceedings rely on video. And finally, each of these design schemes raises different advantages as well as challenges in terms of the efficiency fairness trade off. So, as we can see, when we introduce the technology into the courtroom, the devil is in the design. If we want to ensure that what we create aligns more with the goals of court ODR as opposed to remote proceedings, we need to re-imagine proceedings and not replicate existing ones. We need to advance both efficiency and fairness. We need to realize that the change in medium is not a mere technicality but comes with a different mix of benefits and challenges. And above all, we want to ensure that we understand that design is an ongoing process in which we constantly learn and improve in designing our novel dispute resolution landscape.

Relevant Readings / Activities:

- RABINOVICH-EINY, Orna and Katsh, Ethan, The New New Courts (2017). 67 Amer. U. L. Rev. 165-215

(2017), Available at SSRN: https://ssrn.com/abstract=3508460

- RABINOVICH-EINY, Orna and Katsh, Ethan, Artificial Intelligence and the Future of Dispute Resolution: The Age of AI-DR (2021). Orna Rabinovich-Einy and Ethan Katsh, "Artificial Intelligence and the Future of Dispute Resolution: The Age of AI-DR" in Online Dispute Resolution: Theory and Practice (Mohamed Abdel Wahab, Daniel Rainey & Ethan Katsh, eds.) Eleven International Publishing, (forthcoming, 2021). , Available at SSRN: https://ssrn.com/abstract=3830033 or http://dx.doi.org/10.2139/ssrn.3830033
- Susskind, R., (2020). *The Future of Courts.* Harvard Law School. Available at: <https://thepractice.law.harvard.edu/article/the-future-of-courts/>.
- JUSTICE. *JUSTICE COVID 19 Response.* Available at: <ttps://justice.org.uk/our-work/justice-covid-19-response/>.

Lecture 3.3: What can we learn from the British Columbia Civil Resolution Tribunal?

The lecture will present the case of the Civil Resolution Tribunal, Canada's first online tribunal located in British Columbia. As it is the first example, in the world, of Online Dispute Resolution (ODR) being incorporated into the public justice system, the lecture will provide insight into the landscape of online dispute resolution and the goal, role and actions undertaken by the tribunal since its launch.

Lecture Script (WebVTT)

What can we learn from the British Columbia Civil Resolution Tribunal? Welcome to an introduction to the Civil Resolution, or CRT. The CRT was the first online Tribunal in Canada when we opened our doors in 2016 and in fact the first example in the world of ODR integrated into the public justice system. But creating an ODR platform is not as simple as using Zoom technology or email. Rather, it's about fundamental culture change that requires us to listen to the experiences of people who use the system and then design around it. Our goal at the CRT is to bring the public justice system to people and accommodate their various circumstances and backgrounds. One of the ways that we do this is offering people free upfront legal information and tools to help them resolve their problem without even starting a claim. If they do have to start a claim at the CRT, we offer them the opportunity to negotiate, we help them with mediation, we try to help them solve their problem collaboratively wherever possible, and we use adjudication where a Tribunal member or judge makes a binding court order as a very last resort. One of the other things we do to increase accessibility is we use responsive design so that the platform works well on smartphones as well as tablets and desktops. When nontechnology feature we've incorporated across the Tribunal is using plain language. Everything we write is written at the average reading level of a 10-year-old to ensure that it's understandable to the people who have to read it. Another thing that we do to make sure that we are meeting our mandate is we survey people who have gone through the tribunal experience and asked them what they thought of it. We asked them questions like: "Did we treat you fairly?", “Did we reach a decision in a timely manner?”, “Did we treat you in a professional manner?" And we take that feedback and publish it every month on our website to ensure that we're being accountable and transparent, and to ensure that we have a ready way to improve over time. So, the key message is that online courts are not as simple as using Zoom platforms or email. Rather, what online courts and ODR invites us to do is to re-imagine the public justice system in a way that accounts for the wide variety of needs and circumstances of the people that we serve. And many of the ways that the Tribunal that we've accounted for these circumstances require us to go offline and meet with people, talk with them, and really centre their experience within the design of the Tribunal. Thank you, and if you'd like more information about the CRT, please visit our website at Civilresolutionbc.ca.

Relevant Readings / Activities:

- Explore: https://civilresolutionbc.ca/

Lecture 3.4: What do smart courts look like in China?

This presentation will look at the smart court system in China. Chinese courts are using AI and other digital technologies to improve the country’s judicial capacity and processes in the following ways: 1) making judicial activities more transparent; 2) making judicial services more convenient; 3) making trial work more intelligent; 4) making enforcement more efficient; 5) making information management more standardized; 6) making

judicial decisions more accurate. The presentation will also discuss the use of big data and AI in China's future smart court system.

Lecture Script (WebVTT)

What do smart courts look like in China? Hello everyone! I'm Sun Fuhui, Deputy Director General of IT Center of the Supreme People's Court (hereafter "SPC") of China. It is a great honor to share with you the progress of smart court construction in China. In recent years, Chinese courts have actively embraced modern science and technology, vigorously promoted the in-depth integration of information technology and judicial work to build a smart court, and effectively promoted the modernization of the judicial system and judicial capacity.

It is mainly reflected in six aspects:

First, make judicial activities more transparent. The SPC has built various judicial transparency platforms to support the openness of judicial process, court hearing, judgments, verdicts, and enforcement information. The level of standardization, institutionalization, and informatization of judicial openness has been significantly improved. Every day, tens of thousands of court hearings are broadcast live to the public through the internet, millions of judgments and verdicts are made open to the public, and the progress of cases is timely and actively made accessible to the parties and their agents, to enable concerned people to participate and learn about the cases to the greatest extent possible.

Second, make judicial services more convenient. Chinese courts have been vigorously developing the "Internet + Litigation service" and have completed China Mobile MiniCourt to enable "litigation on fingertips" and full coverage of cross-jurisdictional filing. The Chinese courts promote e-Litigation and have enabled online litigation activities such as case filing, court hearing, mediation, and documents serving. In 2020, 10.8 million civil first instance cases were filed online, accounting for 54% of the total. The Chinese courts have established Internet Courts in Hangzhou, Beijing, and Guangzhou to explore litigation rules and trial mechanisms that are applicable to the internet. Judicial interpretation of online litigation rules of people's courts was promulgated by the SPC to make clear that online litigation activities share the same legal effect as offline litigation activities.

Third, make the trial work more intelligently. The Chinese courts promote online case handling based on e-files and have developed various intelligent auxiliary applications. The intelligent auxiliary applications not only support generation of summons and other standard documents with a single click, intelligently assist judges to generate judgments, and reduce judges' workload by more than 30%, but also intelligently identify cases based on e-files, automatically push similar cases and prompt related cases to judges. In addition, they can automatically identify the sound in the hearing and convert it into trial transcripts which improves adjudication efficiency by more than 25%.

Fourth, make enforcement more efficient. The SPC has built an enforcement informatization system to comprehensively and efficiently support the enforcement work. The enforcement command center is used to enable the integrated management of decision analysis, enforcement command, case handling, supervision, and management. The enforcement process information management system is used to solve the problem of standardized enforcement management; the network enforcement investigation system is used to shatter the obstacles of finding properties and looking for judgment debtors; the online auction platform is used to solve

the challenge of property liquidation, and the credit punishment system is used to solve the problem of dishonesty punishment.

Fifth, make information management more standardized. An information network covering more than 3,500 courts and more than 10,000 detached tribunal has been built to support the court personnel across the country to work via “the same network”. The AI system enables the e-management of government affairs such as document management, events management, and work communication, which greatly improves the efficiency and reduces the cost. The SPC also utilizes the trial supervision system to enable the intelligent inspection of court hearings and litigation services of courts at all levels and urge courts across the country to constantly improve professional ethics in judicial works.

Sixth, make judicial decisions more accurate. The SPC has built a judicial big data management and service platform to enable the auto-convergence, real-time update and dynamic analysis of case info of all courts in the country; to fully support the automatic generation of judicial statistics and intelligent analysis of trial situation; to deeply carry out judicial big data analysis and reveal the legal development behind the case; and to provide decision-making reference for social governance. Going forward, the SPC will provide more intelligent judicial services and support for users with the help of 5G, big data, artificial intelligence, blockchain and other advanced technologies. We are willing to learn from the experience of smart courts in other countries and make contributions to the rule of law of all humankind!

Relevant Readings / Activities:

- Shi C, Sourdin T, Li B. The Smart Court – A New Pathway to Justice in China? International Journal for Court Administration. 2021;12(1):4. Available at: <http://doi.org/10.36745/ijca.367>.

8.3.6 Module Assessment

1. What does the movement towards online courts entail?
    a. A shift towards online conference calls with judges
    b. A movement in the way we offer justice and handle conflicts
    c. A short period of conducting court during the pandemic
    d. A movement of using digital technologies in physical courts

2. What is not a challenge of the online courts movement?
    a. Algorithmic bias
    b. The digital divide
    c. Cybersecurity
    d. *Freedom of opinion and expression*

3. How has Covid-19 affected online courts?
    a. Educated judges about AI risks
    b. Increased online proceedings
    c. Limited the trust towards AI adoption in justice
    d. No affect

4. What is the British Columbia Civil Resolution Tribunal's core value?
    a. Inclusivity
    b. Innovation
    c. Effectiveness
    d. Patience

5. Which is not a goal related to the modernization of smart courts in China?
    a. To make judicial systems more effective
    b. To make judicial systems more universal
    c. To make judicial systems more accurate
    d. To make judicial systems more convenient

8.3.5 Use cases and examples

US - *Wisconsin v Loomis, Wisconsin Supreme Court (2016)*

The 2016 United-States Wisconsin v. Loomis is a pillar case for this module as the court examined the validity of using the AI, COMPAS, risk-assessment software in the sentencing of an individual.[15] In February 2013, Eric Loomis was found driving a car that had been used in a shooting. He was arrested and pleaded guilty to eluding an officer. In determining his sentence, a judge looked not just at his criminal record, but also at a score assigned by the COMPAS tool. His risk scores indicated that he presented a substantial risk of recidivism, and the Judge gave Loomis a long sentence, partly based on this report. Loomis filed a motion for post-conviction relief requesting a new sentencing and argued that the circuit court's consideration of the COMPAS assessment violated his due process rights. Loomis further asserted that the court erroneously exercised its discretion by improperly assuming that the factual bases for the read-in charges were true. In denying the post-conviction motion, the circuit court explained that it used the risk assessment software to corroborate its findings and that it would have imposed the same sentence regardless of whether it considered the risk scores. Loomis appealed to the Wisconsin Supreme Court to see whether the circuit court's consideration of a COMPAS risk assessment violated his constitutional right to due process. The Supreme Court rejected his appeal and ruled for the continued use of COMPAS to aid judges with sentencing decisions.[16] However, the court did express hesitation about the future use of the software in sentencing without highlighting the limitations of the tool.

China- First digital Court, the Hangzhou Internet Court (2017)

China has been adopting technologies into its administration of justice for years. The country's 'smart courts' or 'e-justice' policy framework promotes technologies, such as AI and blockchain, to be used in legal practice and court systems. China's first digital Court, *the Hangzhou Internet Court*, launched its full online litigation platform in August 2017 to hear civil and administrative cases.[17] Consequently, the Supreme People's Court issued an authoritative opinion calling to accelerate building smart courts.[18] The Court noted in 2019; that *"across the broad spread of the internet, problems such as unbalanced development, inadequate rules and inequitable order have become more evident. Citizens, corporations, and social organizations now expect greater access to justice and to guarantee socio-economic development in the digital age, which urges the courts to harness technology to keep pace with the rapidly changing demands of society"*.[19] Later, in September 2019, as Alibaba developed the first AI assistant judge to help judges during trial hearings, significantly reducing the time from filing to closing, the AI debuted in the People's Court of Shangcheng District, Hangzhou City.

China - Shanghai N°2 Intermediate People's Court

This was the first case in China where the '206 system' was used in a criminal trial. This system digitalizes courtroom processes (video evidence, witness), and the AI transcribes courtroom discussions.[20] The 206 system is an integrated AI assistive system for criminal cases that can help judges find facts, authenticate evidence, protect the right to appeal and judge impartially on a trial (Guo Weiqing, President of Shanghai No2 Court, 2019).[21] The system was used for the first time, that day, as the court heard a robbery and a murder case. Inside the courtroom, a screen was placed in front of all people present at the trial, including in the public gallery. When the judge, public prosecutor or defender asked the 206 system, it displayed all related evidence on the screen. Though the case was complicated, the 206-system displayed the case in a clear and comprehensive manner.

China - Hainan High People's Court

The Hainan High People's Court introduced an intelligent system to standardize sentencing. It employs big data and AI technologies, such as natural language processing, knowledge graphs and deep learning, to automatically

[15] State of Wisconsin v Loomis, Circuit Court, 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749
[16] *Wisconsin v Eric Loomis,* 2015AP157–CR, j Wisconsin Supreme Court, 13 July 2016
[17] Hangzhou Internet Court, China. Available at <https://www.netcourt.gov.cn/?lang=En>.
[18] Changqing Shi, Sourdin Tania, Li B.; *The Smart Court – A New Pathway to Justice in China?* International Journal for Court Administration, 2021. Available at <https://www.iacajournal.org/articles/10.36745/ijca.367/>.
[19] The Supreme People's Court; Chinese Courts and the Internet Judiciary, 2019, p. 59. Chinese text. Available at <http://english.court.gov.cn/pdf/ChineseCourtsandInternetJudiciary.pdf>.
[20] Wei Jiang; *China uses AI assistive tech on court trial for first time,* China Daily, 2019. Available at <https://www.chinadaily.com.cn/a/201901/24/WS5c4959f9a3106c65c34e64ea.html>.
[21] Ibid.

pick out key facts in a case and formulate a written judgment, after data analysis based on decisions from previous cases. The Hainan high court is encouraging lower-level courts across Hainan province to use the system, in a bid to increase efficiency and advance standardization in judicial services.[22]

Online Courts in Azerbaijan

Over the last decade, as a result of modernization efforts in the justice sector, Azerbaijan is viewed as a leading example of court efficiency, transparency and accessibility at the international level.[23] The European Commission for the Efficiency of Justice (CEPEJ) has recognized Azerbaijan for its technological innovations in the judicial system. Its smart courts and simplified judicial procedures are considered fast and business friendly.[24] The use of e-court technologies has increased judicial productivity by at least threefold. Simplified business processes are resulting in operational budget savings. During the COVID-19 pandemic, these advances have proven to be an asset, helping courts to remain operational and offer remote services to citizens and businesses in an easily accessible and transparent manner, while successfully fulfilling rule of law responsibilities.

Online Courts in India

Virtual courts in India have emerged as a temporary response to the COVID-19 pandemic. Now, a group of judges and lawyers wants to include virtual courts in normal court proceedings even after the health crisis is over. Since the 2005 E-court project, the Hon'ble apex Court addressed the issue of delivery of justice during the COVID-19 lockdown. A bench consisting of CJI Bobde and Justices Chandrachud and L Nageswara Rao issued a direction on *Guidelines for Court Functioning Through Video Conferencing During COVID-19 Pandemic* regarding measures to be taken by courts to reduce the physical presence of all litigants within court premises by adapting the social distancing guidelines.[25] These guidelines were issued by invoking Article 142 of the Constitution of India as an extraordinary jurisdiction. In the aforementioned order, the District Courts were directed to adopt virtual court hearing through modes prescribed by the concerned High court and to provide video conferencing facilities for litigants who lack resources.

India - Krishna Veni Nagam v Harish Nagam (2014)

A two-judge Bench in *Krishna Veni Nagam v. Harish Nagam,* while seeking transfer of a case when both parties were not located within the jurisdiction of the same court, referred the parties to participate in the matrimonial dispute through video conferencing.[26] While allowing the above-mentioned transfer petition, the difficulties faced by the litigants living beyond the local jurisdiction was acknowledged by the Hon'ble Apex Court that *"it is appropriate to use videoconferencing technology where both the parties have equal difficulty due to lack of place convenient to both the parties. Proceedings may be conducted on videoconferencing, obviating the needs of the party to appear in person, wherever one or both the parties make a request for use of videoconferencing"* .[27]

India - Santhini v Vijaya Venketesh (2017)

The case *Krishna Veni Nagam v. Harish Nagam* was overruled by the Supreme Court of India in *Santhini v. Vijaya Venketesh*. Chief Justice of India, Dipak Mishra and Justice AK Khanwilkar held that *"in transfer petition, video conferencing cannot be directed"*.[28] However, Justice Chandrachud wrote the judgment in favor of the use of modern technology and video conferencing. Justice Chandrachud highlighted, in the dissenting opinion, the benefits of video conferencing. *"The Family Courts Act, 1984 was enacted at a point in time when modern technology which enabled persons separated by spatial distances to communicate with each other face to face was not fully developed. There is no reason for a court which sets precedent for the nation to exclude the application of technology to facilitate*

---

[22] Yuan Shenggao; *AI-assisted sentencing speeds up cases in judicial system*, China Daily, 2019. Available at< https://www.chinadaily.com.cn/cndy/2019-04/18/content_37459601.htm>.

[23] The World Bank; *Azerbaijan: e-Courts and the Digitization of Justice*, June 29, 2021.

[24] World Bank Group; *Judicial Services and Smart Infrastructure Project -Azerbaijan: e-Courts and the joint achievements in digitalization of justice*, coordinator Judge Dr. Ramin Gurbanov, President of the European Commission for the Efficiency of Justice (CEPEJ), Council of Europe. Available at <https://thedocs.worldbank.org/en/doc/3ecf7262788a3ec69c8a45bbd3342a28-0080022021/related/29-06-21-Presentation-e-Court-WB-Dr-Ramin-Gurbanov.pdf>.

[25] The Supreme Court of India, Civil Original Jurisdiction, Suo Motu Writ (Civil) No.5/2020, In Re: Guidelines for Court Functioning Through Video Conferencing During Covid-19 Pandemic, 2020. Available at < https://main.sci.gov.in/supremecourt/2020/10853/10853_2020_0_1_21588_Judgement_06-Apr-2020.pdf>.

[26] *Krishna Veni Nagam v Harish Nigam,* Transfer petition (CIVIL) NO. 1912 OF 2014, 9 March, 2017, §1

[27] *Krishna Veni Nagam v Harish Nigam,* Transfer petition (CIVIL) NO. 1912 OF 2014, 9 March, 2017, §12

[28] *Santhini v Vijaya Venketesh,* Transfer Petition (Civil) No.1278 OF 2016 and transfer Petition (Civil) No. 422 of 2017, 9 October 2017

the judicial process. Imposing an unwavering requirement of personal and physical presence (and exclusion of facilitative technological tools such as video conferencing) will result in a denial of justice".[29] As Sharad Arvind Bobde, the chief justice of India, stated "there is no looking back" and the way forward will be a combination of virtual courts and physical courts, "the new and the old".[30]

Online Courts in South Africa

South-African courts are also tapping into opportunities for innovation within the legal system and judiciary. On the 10th of January 2020 the Judge President Mlambo issued a practice directive for the full implementation of the Court Online Evidence Management Application (CaseLines).[31] This is an advanced cloud -based collaboration solution that is aimed at providing a platform for Law Firms and Litigants to file documents to the Courts electronically (E-Filing over the Internet from anywhere).

At the beginning of 2020, the Gauteng Division and Gauteng Local Division of the High Court, Pretoria and Johannesburg implemented CaseLines. The aim of the system is to enable litigants to file and upload pleadings and other documents electronically and to present their case and argument during court proceedings. The implementation of the system was timely as the world was hit by a pandemic that insisted on limited human contact. The system broadly functions by way of case creation, party/legal representative invitation, document filing and uploading and case presentation. It enables litigants to file and upload pleadings and other documents electronically and to present their case and argument during Court proceedings, with Judges given the opportunity to efficiently and securely prepare and review evidence online and follow evidence presented digitally during the court hearing. In addition, video conferencing technology has also been adopted by South African courts to ensure that the legal system does not grind to a complete halt during the COVID-19 pandemic and extended lockdowns.[32]

[29] *Santhini* v *Vijaya Venketesh,* Transfer Petition (Civil) No.1278 OF 2016 and transfer Petition (Civil) No. 422 of 2017, Dissenting Opinion Justice Chandrachud, 10 March 2017.

[30] Vaid Dharvi; *How coronavirus is propelling the rise of online courts in India*, DW, June 2020. Available at <https://www.dw.com/en/how-coronavirus-is-propelling-the-rise-of-online-courts-in-india/a-53774109>.

[31] Judge President Mlambo's first Practice Directive of 2020 concerning the full implementation of the Caselines digital/electronic system, into effect in the Johannesburg and Pretoria High Courts on 27th January 2020. Available at <https://www.ppv.co.za/wp-content/uploads/2020/01/Judge-President's-Practice-Directive-1-of-2020.pdf>.

[32] Africa Judges and Jurists Forum; A policy Brief: Digital Transformation of court processes in Southern Africa: A Human Rights Approach, June 2021. Available at <https://africajurists.org/wp-content/uploads/2021/publications/21.06-Digital-Transformation-of-Court-Processes-in-Southern-Africa-AJJF-Final.pdf>.

8.4 MODULE 4: ALGORITHMIC BIAS AND ITS IMPLICATIONS FOR JUDICIAL DECISION MAKING

8.4.1 Module Focus

This module will define and unpack the issue of algorithmic bias and consider its implications on the principle of the rule of law. It will explore the sources and origins of algorithmic biases in AI systems used in justice systems. It will explore the implications of algorithmic bias for the digital divide, gender diversity, pluralism, equality, non-discrimination, and accountability. It will consider use cases like gender- and racially- disparate risk classification in software used by criminal judges to evaluate recidivism risks. Lastly, it will explore how algorithmic biases can be prevented, detected, and addressed through a human-centric approach.

8.4.2 Learning Objectives

From this module, participants will:

- Familiarize themselves with concepts of algorithmic bias, fairness, and transparency
- Build an understanding of the different origins and sources of bias, including the digital divide, data divide and capacity divide.
- Familiarize themselves with emerging solutions for preventing, detecting, and addressing algorithmic bias

8.4.3 Lecturers

Diogo Cortiz | Pontificia Universidade Catolica de São Paulo, Professor; Researcher at NIC.br | Brazil
Ameen Jauhar | Vidhi Centre for Legal Policy, Senior Resident Fellow and Lead | India
Bianca Kremer | Oxford Faculty of Law, Doctoral Researcher | Brazil

8.4.4 Structure

Lecture 4.1: What is algorithmic bias and what are its origins and sources?

This presentation will explain the issue of algorithmic bias. Algorithmic bias leads AI systems to systematically make unfair decisions that impact certain individuals or groups. The presentation discusses the ProPublica study on biases in predictive algorithms used for parole decisions. The video will illustrate the steps involved in training AI systems to demonstrate how algorithmic bias can originate from the data. It will conclude by discussing ways in which biases can be reduced.

Lecture Script (WebVTT)

What is algorithmic bias and what are its origins and sources? In this module, we'll discuss about bias, justice, and algorithms. And we start I'll give you an example of a landmark study about bias in AI.
That is the research called machine bias that was performed by ProPublica, and this study investigated, a system used in the US to calculate criminal recidivism. So basically, this system used data set from historical criminal systems in the US to calculate the probability of a person to commit a crime again. And it was used by different actors in the country and in this research, they found that the system was biased, had a bias based on race. So basically, the result of the systems had a high rate to mark black people as a high-risk person and the system had a bias to mark a white person as a low-risk person. But it was a longitudinal study.

So, they compared the results of the systems after two years, and they identified that the black people that the system marked as a high risk, in fact did not commit a crime anymore. But there were many white people that were marked as a low risk that in fact committed a crime. So basically, it's one example of bias in justice using an algorithm. In this section, we will discuss about bias in algorithms, and I will start with a question. Is data the only source of bias in AI? And to give you an answer, I think that first we need to discuss the whole pipeline of an AI.

In the pipeline of training, an AI system we will have three different things. The first one is the dataset that you have, that you use to train a model. Then we'll have the training algorithms. That is the recipe of how you use the data to learn. And finally, we will have the model training that is the outcome of the training process. And as you can see in the image, we start from the data. And if the dataset has any kind of bias, it will be present in the training model. So yes, data is one of the sources of bias in AI, but it's the only one, and we will discuss more with some examples. I'll give you an example of data set in AI. In this the fictional dataset commonly used in academic and research field twisted AI models and to teach about machine learning. This dataset is about the Titanic accident. So, we have information about all the passenger, for example, name, age, sex and if the passenger survived or died in the accident.

So now we use this dataset to train different models. And you can discuss if choosing the right features could reduce or amplify bias in AI. This is the first module I'll give you an example of how we train an AI. So, I took the Titanic dataset and using a decision tree I trained an AI to help us to determine if a person would survive or die in the accident. To train this model I used three main features that the class that the passenger was travelling on, the fare and the age. And as you can see here, we have a trained module. And because a decision tree is an open box or a white box if you prefer, we can investigate further the behaviour of the module. And

as you can see here in the graph, the class feature is the most important feature for this model to decide if a person would survive or die in the accident.

So basically, travelling in the first class, second class, third class is the most important decision for the AI model in this scenario. In this second example, I will show you a different model. So, I took the same dataset in the same trained algorithm, that's the decision tree but now I've managed to add a fourth feature. That's the sex. So now in the training process, our algorithm is using four different features that is: class, fare, age, but also sex. As you can see in the image, it's a completely different model, if you compare it to the previous one, and because the decision tree is an open model, we can investigate the impact of different features in the decision. So basically, you can see that now sex is the most important feature for this module. If you compare it to the previous one, with the first model class was the most important feature. And basically, we can see that we use the same dataset, the same training algorithm. But now we have two differently trained models. One that uses sex as an important feature and the other one you use the class. And based on the specificity of the scenario, one algorithm could be more important for us or could amplify or reduce a bias in the AI. So, this is an example to show that, of course, bias is present in the data, but it could also be amplified or reduced based on the design of our model.

Few considerations about bias in AI. Bias is almost always present, and we need to be aware of it. As we could see in the example, bias is hard to identify and hard to fix because it's not just a technical term that you can identify only using math or algorithms, because we need to understand the whole scenario to which it would be applied. For example, in those two models that we trained the previous sections, we have one model that is not using gender, sex, as a feature to decide in the other one is using this kind of information. And depending on the situation, one could reduce or amplify bias. And we will also need to investigate the social scenario that we apply artificial intelligence to, to decide which model is better. Thank you and enjoy Module four.

Relevant Readings / Activities:

- Cortiz, D., (2019). Artificial Intelligence: fairness, justice and consequences. Available at: <https://cetic.br/pt/publicacao/year-xii-n-1-artificial-intelligence-fairness-justice-and-consequences/>.
- Mayson, S., (2019). *Bias in, Bias out.* The Yale Law Journal. Available at: <https://www.yalelawjournal.org/article/bias-in-bias-out>.
- Watch Coded Bias and reflect using this discussion guide

Lecture 4.2: What risks does algorithmic bias surface for justice?

This presentation will focus on the risks of algorithmic bias for justice. First, it will begin by explaining some terms like Narrow AI, and the difference between data bias and design bias. Second, the presentation will discuss the harms posed by algorithmic bias such as reinforcement of historical biases, impact on civil rights and liberties, the false presumption of AI neutrality, and the black-box problem or the lack of explainability on how algorithms arrive at certain decisions. Finally, it will provide two prominent examples of these harms with predictive justice tools such as COMPAS (US) and HART (UK), and the potential harms associated with the use of facial recognition in the criminal justice system.

Lecture Script (WebVTT)

What risks does algorithmic bias surface for justice? I will give a brief overview of what we mean by algorithmic biases. The relation with AI as it currently exists and how this is likely to pan out in the legal and justice systems. I want to start quickly by discussing what we mean by AI or Artificial Intelligence. It is a collective term that we use for technologies like intelligent algorithms, autonomous systems, robotics, and other tech that can mimic human cognition. AI is further categorized into narrow or general AI But for this presentation, I'm limiting the discussion to narrow AI which basically refers to task-specific algorithms designed to perform set functions. Algorithmic bias is also of two types.

There are data biases and design biases. The task specific algorithms that I just mentioned usually rely on machine or deep learning techniques and require large datasets to be trained for their accuracy. Thus, if there are any biases in such datasets, they are likely to be hard coded into and perpetuated by the algorithm. The other form is designed biases where the innate prejudices of a developer can subconsciously or inadvertently transfer into the coding of the algorithm. I will come back to these in just a minute when I discuss some actual use cases. But before that, I also want to list out the broad harms that stem from biased algorithms. First, as I mentioned, they can reinforce systemic or historical biases.

This, in turn, can result in depriving, especially vulnerable populations of their civil rights and liberties, because of biased algorithmic decision making. Third, the idea that AI or algorithms counter human bias is probably false. And finally, there is the black box problem or the lack of explainability of how algorithms arrive at certain decisions or conclusions. From a judicial standpoint this is particularly problematic as it hampers the ability of an individual to challenge any decision that is made by an algorithm against them, unlike a well-reasoned or detailed judicial order. It also makes it nearly impossible for them to prove that the decision-making was biased.

Now, I want to quickly cover two prominent use cases within the legal and justice systems, which showcase some of these harms.

First, there are the predictive justice tools, which, as the name suggests, are used for predicting outcomes. COMPAS in the United States and HART in the UK are two such tools. COMPAS is a recidivism assessment algorithm used in parole proceedings and was empirically demonstrated by an organization called ProPublica to be biased against African Americans and was found only to be accurate 61% of the time, of its predictions. COMPAS was challenged in the Wisconsin Supreme Court in 2013 by a man named Eric Loomis, who was convicted to six years in prison after this algorithm determined him as a high-risk parole applicant. Loomis argued that the black box of the algorithm did not afford any adequate reasoning for his labelling as a high risk. However, the Supreme Court did not hear this great petition or grant the writ of certiorari.

Similarly, HART, which is used in the UK to determine the risk of future offences, has also displayed what I earlier called design bias. It has the erroneous use of postal codes, for instance, as a proxy to determine potential risk

of an individual. This relegates entire communities as high risk, especially those living in poorer neighborhoods, and presumptively impugns anyone from these neighborhoods as a serious potential offender.

The other problematic technology is that of facial recognition, which though not part of the judicial system, is becoming commonly integrated into law enforcement and policing and thus becoming a part of the criminal justice system. Like any algorithm, facial recognition also depends on large amounts of data sets, and these data sets can either make or break the final algorithm.

Even the most advanced facial recognition tools have been found to be inaccurate like that of Amazon, which is all recognition, and was found to be biased towards darker skin tones, especially amongst women. Such inaccurate algorithms can prove to be quite prejudicial, endangering individuals with arrest, detention or even criminal prosecution. That's having real-life consequences. For instance, there have been cases in the United States where an African American man was last year misidentified by a facial recognition tool and arrested during the peak of the Black Lives Matter movement. Similarly, some years back, a South Asian student at Brown University was misidentified as a terrorist suspect and taken into custody. In fact, as a first, the European Union's proposed aid legislation has determined facial recognition to be a High-Risk technology, requiring stringent regulation and exceptional usage. I hope through this brief overview, I have been able to highlight some of the critical debates around the use of AI or algorithms within the legal and justice systems. More importantly, my hope is that I have sufficiently emphasized the need for an active, open and robust dialogue around these issues by all concerned stakeholders across the globe.

Relevant Readings / Activities:

- Buolamwini, G., Gebru, T., (2018). "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91. Available at: <https://proceedings.mlr.press/v81/buolamwini18a.html>.
- Oswald, M., Grace, J., Urwin, S., Barnes, G., (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, Information & Communications Technology Law, 27:2, 223-250, Available at: < https://doi.org/10.1080/13600834.2018.1458455>.
- Green B., (2018). *"Fair" Risk Assessments: A Precarious Approach for Criminal Justice Reform.* 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018). Available at: <https://scholar.harvard.edu/bgreen/publications/%E2%80%9Cfair%E2%80%9D-risk-assessments-precarious-approach-criminal-justice-reform>.

Lecture 4.3: What does algorithmic bias mean for inequalities and how can it be addressed?

This lecture will present the need for inclusive AI systems by providing a quick introduction to AI algorithms, machine learning, deep learning, AI neutrality, data and algorithmic bias. It will introduce the concepts of fairness, transparency and equality in the adoption and regulation of innovative technologies to present how algorithmic bias can be tackled through a human-centric perspective. The lecture will also discuss how algorithmic bias can be prevented, detected, and addressed through various pathways, ranging from technological to regulatory.

Lecture Script (WebVTT)

What does algorithmic bias mean for inequalities and how it can be addressed? "Sometimes respecting people means making sure your systems are inclusive. At times it means respecting people's privacy by not collecting any data, and it always means respecting the dignity of an individual." These words were spoken by Joy Buolamwini, the computer science and digital activist who founded the Algorithmic Justice League, an organization that looks to challenge bias in decision making software, highlighting the social implications and potential harms of AI. In many places around the world, artificial intelligence technologies are gradually being deployed in several moments of our lives. We can say that the concept of AI is related to the capacity of technological solutions to carry out activities in a way considered intelligent. For example, reasoning, perception of environment and even analysis skills for decision making. AI can also learn for themselves thanks to learning systems that analyze large volumes of data, enabling them to expand their knowledge. In this expansion process, systems are not neutral or objective in how they calculate, sort and present data. Algorithms can affect people's lives by spreading fake information or perpetuating bias by constructing data sets based on perceived identities and stereotypes. But what does algorithmic bias mean, and how could it be tackled through a human-centric perspective?

My name is Bianca Kremer. I'm a Brazilian lecturer, researcher and legal adviser in the fields of law and technology. I've been a digital rights activist concerned with fairness and equality in the adoption and regulation of new technologies through a gender and race perspective. The vast majority of AI applications are based on deep learning, a category of algorithms that Kai Fu-Lee explains in his book 'AI Superpowers' "use massive amounts of data from a specific domain to make a decision that optimizes for a desired outcome." It does this by training itself to recognize deeply buried patterns and correlations connecting the many data points to the desired outcome. It is mostly applied in the fields of insurance, loans, beside recruitments, retails and even public security systems. Facial recognition technologies, for example, have proven to be powerful tools for mass surveillance that are being deployed under narratives about public security and innovation. In other words, sometimes when we talk about patterns, we talk about bias. And the reason is simple: pattern-finding processes work better- are more accurate- when the data is labelled with the desired outcome. And labelling is permeated by subjective values.

Considering that algorithms are sequences of instructions that command the computer, machine learning algorithms can exhibit complex behaviours or behaviours that only humans would. This is where algorithmic bias come in reproducing trends in social behaviour. Even the negative ones. In Brazil, for example, a recent study showed that 90.5 percent of those arrested by facial recognition technology systems are black, but Brazil is the third country with the highest rate of incarceration in the world. A rate that expresses a context of

historical exclusion. And the point is, the lack of transparency about databases is combined with the history of colonial violence, which is unfortunately very similar in several places around the world. It doesn't always appear as an obvious premise. For this, it would be necessary to recognize that technology does not find a neutral and aseptic soil, but an entire structure of inequality that receives it. Because technology is produced by people and applied by people to people. There is no neutral technology because there is no technology outside of a social context.

But if deployed technologies happen to reinforce discrimination against historically oppressed and persecuted populations like women, indigenous, black, people with disabilities, LGBTQAI+ community... What could we do about it? A suggestion is to offer a human centric perspective that takes into consideration the different experiences of people and territorialities in the enjoyment and development of new technologies. Not every innovation is good or should be implemented just because it exists elsewhere in the world. Why not designing tech regulation based on the possibility of redefining the meaning and purpose of technology? Taking as a starting point particular political, economic and cultural tensions, open to multiple ways of being and living in the hyper connected world. A new praxis in the debates on technological production and digital rights that considers other contributions besides global north as possible and valuable political places of enunciation about what is considerable or even relevant in terms of technology. A commitment to diversity of gender, race, class, sexuality, disability, and religion in a globalized world divided by colonial heritage and power dynamics. Algorithmic bias reflects social constructions of differences. Only by recognizing them and the existence of a data- driven economy that takes advantage of this logic of difference for profit, will combating algorithmic bias be possible? Nothing about us without us. Axé muntu! Thank you.

Relevant Readings / Activities:

- Barocas, S. Hardt, M., Narayanan, A., (2019). *Fairness and Machine Learning*. Fairmlbook.org. Available at: <http://www.fairmlbook.org>.
- JUSTICE. *Preventing Digital Exclusion from Online Justice.* Available at: <https://justice.org.uk/our-work/assisted-digital/>.
- Hao, K. and Stray, J., (2019). *Can you make AI fairer than a judge? Play our courtroom algorithm game*. [online] MIT Technology Review. Available at: <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>.
- Barabas et al., (2018). *Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment.* Available at: <https://arxiv.org/pdf/1712.08238.pdf >

8.4.6 Module Assessment

1. What is algorithmic bias?
    a. Unequal redistribution of algorithms
    b. Phenomenon where AI systems make decisions that are systematically unfair to certain groups of people
    c. Historical phenomenon of segregating certain individuals in justice
    d. Selection of algorithms based on subjective values

2. Which of the below is not a source of bias in AI?
    a. Data
    b. Computing power
    c. Training algorithm
    d. The model

3. What are potential harms of algorithmic bias for justice?
    a. Potential to reinforce human and systemic biases
    b. Lack of explainability of algorithmic decisions
    c. Presumption of neutral and a strong 'automation bias'
    d. All of the above

4. What is the name of the predictive justice tool that exemplifies bias?
    a. COMPAS
    b. Brit
    c. Judge-bot
    d. AGIN

5. Which communities do AI systems reenforce discrimination against?
    a. Europeans
    b. Historically oppressed individuals
    c. Policymakers
    d. Technologists

8.4.5 Use cases and Examples

ProPublica on COMPAS and algorithmic bias (2016)

ProPublica, a nonprofit news organization, critically analyzed the AI risk-assessment software known as COMPAS, to forecast which criminals are most likely to reoffend, and found evidence of so-called 'machine bias'.[33] Guided by these AI-powered assessments, judges in courtrooms throughout the United States would generate conclusions on the future of defendants and convicts, determining everything from bail to sentences. Based on survey questions, the software estimates how likely a defendant is to re-offend. ProPublica compared COMPAS risk assessments with how often people reoffended and discovered that the algorithm was in a way able to predict the tendency of a convicted criminal to reoffend. However, when the algorithm was wrong, the results were displayed differently for people of color. ProPublica discovered that the COMPAS assessments were likely to consider people of color twice as much a higher risk despite their criminal history displaying lower probabilities to reoffend.

Italy - Deliveroo Case (2021)

In January 2021, the Court of Bologna ruled that the reputational-ranking algorithm used by Deliveroo to assess riders was discriminatory.[34] This algorithm was used to determine the reliability of the workers, which would, for example, be negatively affected if they cancelled a previously booked shift within 24 hours prior to the start. As it became clear that more 'reliable' drivers were offered more opportunities in busier blocks of time, and those that had to miss shifts, due to an emergency or serious illness in the past, would get fewer opportunities in the future. According to the Court, the fact that the algorithm did not consider the grounds for cancellation, unfairly penalized workers with legitimate reasons for not working, and amounted to discrimination.[35] Conclusively, the Bologna Court ruled that the reputational-ranking algorithm used by Deliveroo discriminated against delivery workers by breaching local labour laws as it did not differentiate between legitimate and illegitimate reasons for cancelling. This case is indicative of an increased willingness on behalf of regulators, the judicial system, labor unions, and workers across the continent to tackle Blackbox algorithms and for increased awareness of how such algorithms can potentially be abused to circumvent traditional labor protections.

Italy - Foodinho Case (2021)

In the *Foodinho case,* the Italian Data Protection Authority found that the algorithm used by delivery platform Foodinho by Glovoapp23 - to rank riders and allocate deliveries - was discriminatory.[36] Foodinho's use of automated processing to assign riders with a score and prioritize those with higher ratings in the assignment of delivery slots, was considered discriminatory. As the 'excellency score' is assigned through an automated mathematical formula - mostly based on feedback and delivery rates - but negative feedback outweighs positive feedback, the system penalizes riders who do not reach certain delivery thresholds.

According to the Italian Data Protection Authority, Foodinho's management violated Article 22(3) of the GDPR, mainly because; Foodinho made decisions about its riders based solely on automated decision making, by analyzing or predicting aspects of their professional performance, behavior, and their location and movements. Additionally, Foodinho did not adopt any technical and organizational measures to verify the accuracy of its

---

[33] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner; Machine Bias: *There's software used across the country to predict future criminals and it's biased against blacks*, ProPublica, May 2016. Available at <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

[34] Keane Jonathan; *Deliveroo Rating Algorithm Was Unfair to Riders, Italian Court Rules*, Forbes, January 2021. Available at <https://www.forbes.com/sites/jonathankeane/2021/01/05/italian-court-finds-deliveroo-rating-algorithm-was-unfair-to-riders/?sh=2937afc122a1>.

[35] *FILCAMS CGIL Bologna, NIDIL CGIL BOLOGNA, FILT CGIL Bologna v Deliveroo Italia S.R.L.*, N.R.G. 2949/2019, Ordinary Court of Bologna Labor Section, 5 January 2021.

[36] Injunction order against Foodinho s.r.l., Order No 234, *Italian SA v (Garante per la protezione dei dati personali Italy) v Foodinho*, 9675440, 10 June 2021.

algorithmic system, nor the accuracy, relevance, and adequacy of the data used by this system in relation to the purposes pursued. In addition, Foodinho did not adopt measures aimed at reducing the risk of distorted or discriminatory effects in the context of both the scoring system and the orders assigning system where riders receive fewer work opportunities in case of low or sporadic deliveries. In addition to violating article 22(3) GDPR, The Italian Data Protection Authority found that Foodinho had violated Articles 5(1)(a), (c) and (e),13, 25, 30, 32, 35 and 37 of the General Data Protection Regulation.[37] It fined Foodinho 2.6 million EUR for its use of performance algorithms in connection with its employees.

US Health Care Risk Algorithm (2019)

On the 25th of October 2019, Obermeyer et al. published a study that found that a health care risk-prediction algorithm, used on more than 200 million people in the US, demonstrated racial bias as it relied on a faulty metric for determining need.[38] This algorithm was used to compute who qualifies for extra care, and the authors found that the algorithm discriminated against people of color. While the authors agreed that race itself was not a variable used in the algorithm, the bias occurred because the algorithm used health costs as a proxy for health needs. For distinct reasons, patients of color incurred lower health-care costs with the same conditions on average. Reformulating the algorithm so that it no longer uses costs as a proxy for needs, could eliminate the racial bias in predicting who needs extra care. Thankfully, researchers worked with Optum to reduce the level of bias by 80%.[39]

[37] Abstract of Italian SA (Garante per la protezione dei dati personali) order as issued against Foodinho S.r.l., 9677611, 5 July 2021.
[38] Ziad Obermeyer, Brian Powers, Christine Vogeli, Sendhil Mullainathan; Dissecting racial bias in an algorithm used to manage the health of populations, Science, 25 October 2019.
Available at< https://www.science.org/doi/10.1126/science.aax2342>.
[39] Paul Kari; Healthcare algorithm used across America has dramatic racial biases, The Guardian, 25 October 2019. Available at <https://www.theguardian.com/society/2019/oct/25/healthcare-algorithm-racial-biases-optum>.

8.5 MODULE 5: SAFEGUARDING HUMAN RIGHTS IN THE AGE OF AI

8.5.1 Module Focus

This module will provide an overview of key issues related to AI's impact and implications for human rights as well as the legal challenges associated when safeguarding them. Amongst others, it will explore AI's impact on the rights to freedom of opinion and expression (about content personalization, filter bubbles, echo chambers, and online content moderation), the right to privacy (regarding data privacy, surveillance, and facial recognition), and the right to equality (regarding algorithmic bias and due process of the law). For each of these human rights, it will draw upon examples from around the world to consider the legal implications surfaced in the age of AI.

8.5.2 Learning Objectives

From this module, participants will:

- Contemplate AI's impacts on human rights including the rights to freedom of opinion and expression, the right to equality, and the right to privacy
- Familiarize themselves with concepts of content personalization, filter bubbles, echo chambers, and online content moderation
- Familiarize themselves with concepts of data privacy, surveillance, and facial recognition
- Contemplate what AI means for equality, particularly regarding gender diversity and the Global South
- Acquire information on relevant existing or proposed national and regional legislative approaches to safeguard human rights

8.5.3 Lecturers

Cecile Aptel | United Nations Institute for Disarmament Research, Deputy Director; Professor at The Fletcher School at Tufts University
Teki Akuetteh Falconer | African Digital Rights Hub, Executive Director | Ghana
Hon. Ricardo Perez Manrique | Inter-American Court of Human Rights, Judge

8.5.4 Structure

Lecture 5.1: How does AI violate and/or help protect human rights, particularly the rights to freedom of opinion and expression?

This presentation will introduce the module and its objectives. It will provide an overview of AI's impact on human rights, sustainable development, and its legal implications by acknowledging the potential of AI to help ensure human rights but also jeopardize them. To maximize the benefits of AI while minimizing its harmful potential, the development and deployment of new technologies must be framed through a human right based approach and apply core principles of equality, nondiscrimination, participation, and accountability. It will discuss the role of judicial operators in respecting and ensuring international human rights law and explore AI's impact on the rights to freedom of opinion and expression. It will address concepts like content curation or personalization, filter bubbles, echo chambers, online content moderation and transparency.

Lecture Script (WebVTT)

Welcome to Module Five, safeguarding human rights in the age of AI. Lecture one, how does AI violate and/or help protect human rights, particularly the rights to freedom of opinion and expression?

International human rights law; a body of law agreed by states and monitored by national, regional and international mechanisms, provides a key guiding framework for societies in shaping the responses to the challenges of an ever changing technological environment, and also a universal legal framework to guide States and businesses when designing, developing and using artificial intelligence (AI). This module will provide an overview of some of the issues and legal challenges related to the impact of AI and the implications for international human rights law. This module will explore AI's impact on the rights to freedom of opinion and expression, the right to privacy, and the right to non-discrimination and equality drawing upon examples from around the world.

The five learning objectives of these modules are; one, contemplate AI's impact on human rights, including the right to freedom of opinion and expression, the right to equality and the right to privacy. Two, introduce the concepts of content personalization, filter bubbles, echo chambers and online content moderation. Three, present the concept of data privacy, surveillance and facial recognition. Fourth, contemplate what AI means for equality and non-discrimination, particularly in regard to gender diversity and the Global South. And fifth, provide information on some existing or proposed national and regional legislative approaches to safeguard human rights.

To introduce this model, here is an overview of the impact of AI on human rights and sustainable development. Like other technological innovations, AI can be developed and used to advance the fulfilment of human rights and the Sustainable Development Goals. The potential of AI to analyze vast amounts of information could contribute to developing more effective policies and actions in a number of fields. AI may, for example, lead to improvements in the provision of health, water and energy services to the population. It also offers the potential of increased productivity beneficial for the economy. It may also contribute to solving some highly complex problems, such as climate change, the conservation of life on land and under water.

Yet, AI can also jeopardize human rights and sustainable development. The effects of the AI depend on a multitude of factors, including how it is designed and developed, the modalities of its application and use, and the ways in which they inform or make decisions. Let us consider three examples of the potential negative effects of AI on the SDGs. First, the uneven distribution of smart technologies could increase inequalities between countries and also within countries with increased unemployment rates. As tasks carried out by humans are replaced by machines, education requirements needed to operate those smart systems increase. Second, the use of AI, particularly in the online sphere through content creation, could contribute to political polarization and the reduction of social cohesion. Third, the increased reliance on ICT technologies in general, especially AI, could have negative consequences for the environment. AI has a large carbon footprint because of its dependence on vast data processing, and that's problematic, particularly where the power to sustain these high energy demands come from non-renewable resources. Given its widespread use across sectors, AI can also potentially contribute to violations of virtually all human rights.

Just a few examples; the use of AI in determining suitable candidates for a job impact the right to work. Persons identified as suspect through the use of facial recognition technology that has made a false identification and that are detained on this basis will see several of the human rights violated, including the right to privacy and the right to liberty, and even possibly the right to a fair trial. In many of these and other scenarios, the right to non-discrimination may be infringed upon because of discriminatory outputs or use of the technology. Facial recognition technology, for example, is known to currently perform worse for non-white and women. A particular challenge for the effective protection and promotion of human rights is the opacity of many AI systems.

As the UN High Commissioner for Human Rights has pointed out, the complexity of algorithms and models underlying AI systems, as well as the secrecy of governments and private actors, are factors that undermine meaningful ways for the public to understand the effects of AI systems on human rights and society. To maximise the benefits of AI's use while minimizing its harmful potential, the development and deployment of new technologies must be framed by human rights. Such a human rights-based approach to AI requires the applications of core principles, including equality, non-discrimination, participation and accountability. Principles that are also at the heart of the Sustainable Development Goals. Accordingly, the UN High Commissioner for Human Rights in 2021 called for a ban of AI applications that cannot be operated in compliance with international human rights law and imposed moratoriums on the sale and use of AI systems that carry a high risk for the enjoyment of human rights, unless and until adequate safeguards to protect human rights are in place. Human rights treaties impose obligations on States. This extends to all branches of the States, including the government, the executive, the judiciary. The judiciaries obligations to implement international human rights law is expressed through particular rights, such as the right to a fair trial, but also the judiciary, judges and law enforcement officials have a more general role and responsibility in controlling that the human rights obligations of States are observed by other branches of governments. Importantly, the executive. Courts have a role in determining the legality of the use of new technologies such as AI under human rights law and courts can also help in clarifying the conditions for its permissible use through the interpretation and application of international human rights law. The freedoms of opinion and expression form part of international human rights law, which is binding on states to global and regional human rights treaties. The right to freedom of opinion is absolute.

While restrictions to the freedom of expressions are legitimate only when they are provided by law, pursuing a legitimate aim and are necessary and proportionate. State may also be under an obligation to exercise due diligence to prevent human rights abuse committed by private actors. Private companies such as social media companies and telecommunications companies are central actors in the online space. The UN Guiding Principles on Business and Human Rights contains authoritative guidelines on the human rights responsibilities of such companies. As noted by the UN Human Rights Committee, the two rights, freedom of expression and freedom of opinion are indispensable conditions for the full development of the person. They constitute the foundation for every free and democratic society, and they also form a basis for the full enjoyment of a wider range of other human rights.

They are a necessary condition for the realisation of the principles of transparency and accountability. In the digital age, the protection of freedom of expression online and the right to privacy, which are interlinked are fundamental. As the public square becomes digitalized and takes place largely on social media platforms, AI technologies, such as machine-learning, are central to how information is shaped online. As such, they can impact the rights to freedom of opinion and expression in a variety of ways. This short presentation will look at the ways in which the use of AI in content creation and content moderation affects these rights. Let's first consider content creation or personalization.

It entails the use of AI to tailor what type of content is visible to the individual user, typically on a given online platform or search engine. The algorithms used for these purposes use massive data sets, including browsing histories, user activity and user demographics that feed into increasingly personalized algorithmic models to rank and curate information. Paid or sponsored content is often promoted to the exclusion or demotion of other content. Content curation thus dictates what appears on social media feeds and as search engine results. Although this can bring the benefit of a user experience that respects the personal preferences of an individual, it may also have negative effects: First, content creation raises concerns with respect to the right to privacy of individuals, particularly because of the amount of personal data harvested, its storage, use and dissemination to third parties, often without the individual's knowledge or consent. Second, there is a risk that curation of content contributes to the creation of echo chambers or filter bubbles, where individuals are exposed to information that reinforces their existing beliefs, even when those beliefs are based on demonstrably false or harmful information. It may also prevent them from accessing ideas and information that can challenge those beliefs, such as critical and credible news information. In electoral contexts and during the COVID-19 pandemic, we have for example witnessed coordinated disinformation campaigns, the effects of which have reportedly been amplified by content display algorithms, as well as the spread of false information about the pandemic.

Third, the promotion or demotion of content based on user engagement and paid promotions can also have negative impacts on media diversity, favouring more powerful media outlets to the detriment of smaller ones in the increasingly important online media landscape.

Because of the pervasive effects that content creation has in shaping the thoughts and opinions of individuals, the UN Special Rapporteur on the Freedom of Opinion and Expression has cautioned that some forms of content creation could be contrary to the right to freedom of opinion. Let us now turn to content moderation. That is the process of monitoring user-generated content online and of removing or limiting access to content. It is notably carried out by social media companies on their platforms, with increasing dependence on AI tools. Where the moderation of content results in restrictions on user accounts or on content, their removal or other

restrictions. Content moderation can constitute a restriction to the freedom of expression. The use of AI in content moderation responds to a real need to moderate speech in the online sphere occurring in an unprecedented speed and magnitude. It helps restrict speech that must be prohibited, such as incitement to violence, discrimination and hostility, The use of AI to conduct such moderation, however, raises important challenges. Automated content moderation, such as by social media companies, is often carried out on the basis of the companies' own rules or community standards or through domestic laws which are not directly aligned with international human rights law.

Therefore, although content is protected under human rights law, it may be subjected to takedowns or other restrictions as a breach of private standards or of domestic law. Furthermore, the AI tools used are limited in their ability to reliably detect harmful content. They often falsely identify content as violating laws or company rules, leading to a considerable level of restrictions of legitimate, protected expression, such as take-downs or de-amplification. At the same time, AI tools frequently overlook expression, which should be restricted, such as incitement to violence. These failures have various reasons. Among these are the fact that AI cannot, at this stage, truly understand language and human communication. Studies have also shown that various AI tools widely used for moderating content are biased.

For example, speech of certain ethnic groups or by people of certain backgrounds is more likely to be flagged as harmful then content posted by others. Finally, the reliability of AI detection and classification tools largely depends on the language concerned. For many languages, in particular non-Western languages, available AI tools display increased error rates, if they exist at all. These challenges are compounded by concerns regarding the efficiency of remedies afforded to individuals to vindicate their rights, either in the form of internal grievance mechanisms developed by the company itself or through domestic courts. One of the important concerns with respect to the use of AI in the online sphere, particularly by private companies, is the lack of transparency surrounding the AI systems themselves and their use. It is often very difficult for individual users to understand how the use of AI impacts on their rights, but also on how they are likely to impact on the enjoyment of human rights in general. UN Special Procedures and others have therefore called on greater transparency, as well as on stakeholders' engagement and human rights due diligence in the development and use of AI tools.

Relevant Readings / Activities:

- Commissioner for Human Rights (2018). *Safeguarding human rights in the era of artificial intelligence.* Available at: <https://www.coe.int/en/web/commissioner/-/safeguarding-human-rights-in-the-era-of-artificial-intelligence>.
- CAHAI, (2020). Feasibility study on a legal framework on AI design, development and application. Council of Europe. Available at: <https://www.coe.int/en/web/artificial-intelligence/-/the-feasibility-study-on-ai-legal-standards-adopted-by-cahai>.
- Access Now, (2018). *Human Rights in the Age of Artificial Intelligence.* Available at: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.
- UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment, UN docs A/73/348
- UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on content regulation, UN docs A/HRC/38/35
- UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on disinformation, UN docs A/HRC/47/25
- OSCE Policy paper on freedom of the media and artificial intelligence [accessible here: https://www.osce.org/representative-on-freedom-of-media/472488]
- The Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression: Artificial Intelligence, Content Moderation, and Freedom of Expression [accessible here: https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf]
- Barrie Sander, Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation (2020) Fordham International Law Journal 43(4): 939-1006 [accessible here: https://ir.lawnet.fordham.edu/ilj/vol43/iss4/3/]
- Evelyn Aswad, Losing the Freedom to Be Human (2020) 52 Columbia Human Rights Law Review 306 [accessible here: http://blogs.law.columbia.edu/hrlr/files/2020/11/306_Aswad.pdf]
- Maja Brkan, Freedom of Expression and Artificial Intelligence: On Personalisation, Disinformation and (Lack Of) Horizontal Effect of the Charter [accessible here: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3354180]

Lecture 5.2: What does AI mean for the right to privacy?

This presentation will discuss the right to privacy, 'informational privacy' and the subsequent challenges caused by using AI whilst introducing the right to privacy, data regulation and concepts of mass surveillance and facial recognition. Additionally, we are witnessing threats of algorithmic bias embedded in AI tools leading to discrimination, threats of data persistence as data exists beyond the existence of the human subject, data repurposing as data collected for a specific purpose can be used for a wide variety of other reasons, and data spillovers as data is being collected on people who are not the original target. In the end, the presentation will highlight measures to mitigate these threats.

Lecture Script (WebVTT)

What does AI mean for the right to privacy? Hi, I will be discussing what does AI mean for the right to privacy? I have divided this presentation into four parts. First, looking at what is privacy, the right to privacy and technologies, AI and the threats to data privacy and mitigating the threats that come with AI. What is privacy?

The concept of privacy is indeed a very broad one and has been identified to include physical, psychological, decisional and informational privacy. Even though the notion of privacy has been seen to be an inherent part of our existence as human beings, it has always been a very difficult legal concept to grasp as a result of its very subjective nature. Today, the notion of privacy, as widely discussed, pertains more to informational privacy that is the right to prevent intrusion into one's life and information. Some of our earlier mentions of privacy as a fundamental human right can be seen under the Universal Declaration of Human Rights in 1948. Currently, there are over 150 countries that have enshrined the right to privacy in one form or other within their local legislations. Now let us look at the rights of privacy and technologies. As technologies continue to advance in the processing of information and large data sets, the fears and the threats to the notion of privacy have always been on the increase. Therefore, what we tend to see today across the globe in the development of laws for personal data protection is as a result of this growing threat. These are what has become known as data privacy and data protection laws. The most recent making headlines being the General Data Protection Regulation. Data privacy or data protection laws are legal frameworks that sets out guidelines for the collection, the protection and the processing of information from individuals. It is important to, however, note that in some jurisdictions, such as South Africa, personal data has been defined to include information about legal personalities such as corporate entities. Now are the threats to privacy by AI real? With the potential of collecting large amounts of data sets, including personal information, it is no doubt that AI magnifies the ability to use personal information in ways that can intrude on the privacy interests of individuals. Now what are the forms of threats that such AI technologies are taking? These threats are coming in by way of collection of large personal data sets on the blindside of individuals. So we are in an era where we are witnessing an increase in the deployment of mass surveillance and AI facial recognition tools for policing and other purposes. We are also seeing threats of how algorithmic bias embedded in AI tools lead to racial and minority discriminations, wrong identifications and sometimes false imprisonment and death. We are also seeing threats on data persistence, where data is existing longer and beyond the human subjects that created it. We are also seeing data repurposing that is making it possible for data that was collected for particular use to be used in an unimagined way, sometimes to the threats of lives as well. We are seeing that data also collected on people who are not a target of collection, leading to a lot of data spillovers in the event. How do we therefore mitigate

these threats or risks? Clearly, the technologies are not going anywhere.

And there are good uses to which we can put them. For which reason our objectives must be to look at the measures that can minimise the potential threats of such technologies. These measures must include the following: ethical considerations that look at human centric approach to the implementation of such technologies, data governance rules that prescribe the appointment of privacy officers, conduct of impact assessments and privacy by design mechanisms when it comes to the deployment and use of AI algorithms. We need to also look at data stewardship requirements that put in place duties of fairness or loyalty. To mitigate against users of personal information that have adverse and unfair effects on individuals. We also have to look at putting in place data transparency and disclosure information that determines or lets individuals know what information is being collected on them, how it is being used. Last but not the least, we need to put in place rules around data security, data collection and sharing, in order to reduce the impact and the negative interferences out of AI systems that are being developed and deployed.

Relevant Readings / Activities:

- Cavoukian, A., (2019). *Privacy by Design: Implementation and Mapping of Fair Information Practices.* Available at: <https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf>.
- Janeček, V. (2020). *Data Protection, the Value of Privacy and Compensable Damage.* [online] Cambridge University Press. Available at: <https://www.cambridge.org/core/journals/cambridge-law-journal/article/abs/data-protection-the-value-of-privacy-and-compensable-damage/81996C4CBB4E624ED81410AA9946F20D>

Lecture 5.3: What does AI mean for the right to equality?

This presentation will explore challenges faced by societies worldwide as a result of the increasing development and adoption of AI technologies. It will showcase how AI can perpetuate biases based on gender, religion, and ethnicity, and curtail the freedom of societies, specifically the freedom of expression. It addresses the need for a global discussion to ensure equality and to promote non-biased technology. This presentation proposes a transparent process for the development and governance of AI, as well as oversight mechanisms to ensure protection of the right to equality.

Lecture Script (WebVTT)

What does AI mean for the right to equality? As a judge of the Inter-American Court of Human Rights, of the Regional Court, which in America has a jurisdiction over 500 million people, I would like to discuss the topic of artificial intelligence and its link to human rights. Artificial intelligence is undoubtedly a transformative technology that allows to automate processes, increase the effectiveness and efficiency of the institutions, and yield business opportunities across the board. This means that it is one of the fundamental aspects of technological evolution, which will entail a real change on the planet, and it must be said that its benefits will be enormous. For the time being, due to the pandemic, in which the possibilities of monitoring technology has made it possible to predict and implement public health policy, and undoubtedly these benefits come at the cost of some risks which I will try to analyze within the framework of the Rule of Law and respect for human rights.

Undoubtedly, an artificial intelligence managed exclusively by algorithms, difficult or impossible to explain for the common people, constitutes in itself a risk, to the point that it is uncontrollable for common beings to verify and control this type of mechanism. It can also lead to an overabundance of surveillance. It can lead to situations in which people's privacy disappears or is compromised.

Therefore, thinking about the reproduction through artificial intelligence of biases that have to do with gender discrimination, discrimination based on ethnicity or nationality, and even with the exercise of freedom of expression itself. Whenever it restricts people's freedom to express themselves freely according to their opinions. I understand that all these risks can be mitigated or managed as long as the technological development goes hand in hand or in parity, with the development and analysis in perspective of the Rule of Law of the way this intelligence, above all, is applied and affects people.

I must emphasize that although this dialogue may seem to be a first world debate, where there are many resources to make this control possible, and from where this development fundamentally originates, even in this same first world there are serious difficulties to implement adequate formulas to ensure that artificial intelligence respects the rights of individuals. That is why I understand that being in a region of the planet, the southern hemisphere, with serious difficulties from the technological development standpoint, and from the point of view of people's economic capacity, believe that in the countries of the south we should be seriously concerned about this issue and get involved in technological development, but also in the development of adequate guarantees in terms of the Rule of Law. So that artificial intelligence becomes an element that enables transcending the limitations of the human being and does not generate an element that increases control over this human being and restricts his or her freedom.

I believe that this is the great challenge for humanity. I understand that, as stated in the 2030 Sustainable Development Goals, it is essential, in accordance with Goal 16, to develop inclusive societies that are accountable and participatory, where technological development is also framed within the framework of the Rule of Law and the respect for the rights of the individual. This Sustainable Development Goal 16 also points out how essential free access to public information is.

In other words, there cannot be technological mechanisms, such as artificial intelligence, which cannot be accessed by means of proper public information. Therefore, we have to ensure that States are active in their role of being involved in technological development, of following it closely and being able to manage the necessary information so that this development does not turn into an attack on people's freedom.

To conclude, I think it is very important to point out that there are examples in the world where artificial intelligence is being used to rate people on their behavior. depending on whether or not they respect traffic rules, how they go to the supermarket and whether they pay their bills. I think this is a major risk and we cannot reach these levels of control over the private lives of the population. That is why the Inter-American Court, which I joined in 2020, on April 9, 2020, noted that surveillance technologies could be beneficially used to combat the Covid-19 pandemic, only to the extent strictly necessary to combat the pandemic.

From this point on, the use of these technologies constitutes intelligence in private life and is arbitrary on the part of the State. Moreover, why not, from the big technology players, who are the ones who actually manage these technologies. In summary, The Rule of Law demands the democratic control of technological development and requires the commitment of the States and from the people within the State, so that artificial intelligence may be an element that allows improving people's lives and not a factor that ends up harming people's lives and privacy and infringes on their freedom. I believe that this balance between technological development and freedoms is essential, and it is the great challenge that humanity faces nowadays, when it is at the gates of this formidable instrument, which is artificial intelligence.

Relevant Readings / Activities:

- Pizzi, M., Romanoff, M., Engelhardt, T. (2021). *AI for Humanitarian Action: Human Rights and Ethics.* International Review of Red Cross. Available at: <https://international-review.icrc.org/articles/ai-humanitarian-action-human-rights-ethics-913>

8.4.6 Module Assessment

1. How does AI help advance human rights and UN Sustainable Development Goals?
    a. AI does not help
    b. AI makes human rights more accessible
    c. AI can lead to provisions in health, education, water, energy, etc.
    d. AI can reduce the need for education

2. What key recommendation did the UN High Commissioner of Human Rights make regarding AI adoption?
    a. Ban of applications that cannot be operated in accordance with human rights law
    b. Ban of all AI applications
    c. Ban of AI applications that involve AI's adoption in the pharmaceutical industry
    d. No action

3. What is content curation in the context of AI's impact on the freedom of opinion and expression?
    a. Categorization of data online
    b. Use of AI to tailor what type of content is visible to the individual user online
    c. AI adoption to curate content from various social media platforms
    d. Use of digital technologies to curate content relevant to human rights

4. How can AI's risks towards the right to privacy be mitigated?
    a. Through a combination of ethics, governance, transparency, stewardship, and rules
    b. Only through individual consent
    c. Only through laws and regulation
    d. None of the above

5. How can AI potentially affect the right to equality?
    a. Widen inequality gaps
    b. Shrink inequality gaps
    c. Not affect equality
    d. All of the above

8.5.5 Use cases and Examples

Netherlands - SYRI Case: Civil society groups v The Netherlands (2020)
On the 5th of February 2020, the District Court of The Hague in the Netherlands ordered the immediate halt to a digital benefit fraud detection tool (SyRI) targeted at poor neighborhoods in the Netherlands. It violated Article 8 of the European Convention on Human Rights.[40] In its judgment, the court expressed concern about the significant effect of 'risk indications' on privacy. According to the Court, the SyRI legislation contained insufficient safeguards against invasions of privacy including a serious lack of transparency about its functioning. The Court ruled that the legislation regulating the use of SyRI violated higher law and did not comply with Article 8 of the European Convention on Human Rights which protects the right to respect for private and family life, home, and correspondence.[41] Even though tools such as SyRI are not unique, it is part of a global trend toward the introduction and expansion of digital technologies in welfare states, and often, the consequences of these systems for the human rights of the poorest and most marginalized are completely overlooked. This decision sets a strong legal precedent for other courts to follow as it was the very first time a court stopped the use of digital technologies and abundant digital information by welfare authorities on human rights grounds.

ECtHR* - Bulk data collection & Human rights - *Big Brother Watch and Others v the United Kingdom (2018)
On 13 September 2018, the European Court of Human Rights (ECtHR) ruled that the United Kingdom's bulk data-collection program violated human rights law by failing to incorporate adequate privacy safeguards and oversight. It was the first ruling against Britain's mass-surveillance programs since Edward Snowden's 2013 revelations. The ruling relied heavily on the right to respect for private and family life, enshrined in Article 8 and Article 10 on the right to freedom of expression of the European Convention for the Protection of Human Rights and Fundamental Freedoms.[42]

UK - Facial recognition in public & the right to privacy - *Ed Bridges v South Wales Police (2020)*
Ed Bridges challenged South Wales Police's use of live facial recognition in public. He argued the force was breaching rights to privacy, data protection laws, and equality laws. According to Bridges, the South Wales Police has used the FRT tech on more than 60 occasions since May 2017 and would have taken sensitive facial biometric data from 500,000 people without their consent. However, in September 2019, the High Court decided that while facial recognition does interfere with the privacy rights of everyone scanned, the current legal framework provides sufficient safeguards.[43] Liberty, acting as solicitor for Bridges, submitted an appeal. On August 11, 2020, The Court of Appeal agreed with Liberty's submissions, on behalf of Bridges, 37, and found that South Wales Police's use of facial recognition technology breached privacy rights, data protection laws and equality laws. The Court held that there were "fundamental deficiencies" in the legal framework and that Ed Bridges' rights were breached as a result. This meant that the police force had to stop using facial recognition on UK streets.[44]

US - Facial Recognition and Privacy law - *Flores et al v Motorola Solutions Inc. et al.* (2020)
Plaintiffs Flores, Gomez, and Lewis filed a class action complaint on the 14th of February 2020, in the Illinois Northern District Court, against Motorola and Vigilant Solutions for violating the Illinois Biometric Information Privacy Act (BIPA).[45] Motorola and Vigilant allegedly "*collected, captured, obtained, disclosed, redisclosed, disseminated and profited*" from facial scans of Illinois citizens, and created a "'gallery' of over 18 million booking

[40] SYRI case, *NJCM et al. and FNV v The State of the Netherlands*, the Hague District Court, C-09-550982-HA ZA 18-388, 5 February 2020.
[41] SYRI case, *NJCM et al. and FNV v The State of the Netherlands*, the Hague District Court, C-09-550982-HA ZA 18-388, 5 February 2020, §6.7, 6.20, 6.26
[42] *Big Brother Watch And Others v The United Kingdom*, *Applications nos.* 58170/13, 62322/14 and 24960/15), ECTHR, STRASBOURG, 25 May 2021, §522, 528
[43] *R (Bridges) v- Chief Constable of South Wales Police and Secretary of State for the Home Department*, [2019] EWHC 2341 (Admin), Case No: CO/4085/2018, High Court of Justice Queen's Bench Division, Divisional Court Sitting at Cardiff Civil Justice Centre, 4 September 2019
[44] *R (Bridges) v- Chief Constable of South Wales Police & Ors*, [2020] EWCA Civ 1058, Case No: C1/2019/2670, Court of Appeal (Civil Division) On Appeal from The High Court of Justice Queen's Bench Division (Administrative Court) Cardiff District Registry, 11 August 2020. Available at <https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf> jo. Kurth Andrews; *UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in Bridges v South Wales Police*, Hunton, 12 August 2020. Available at <https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/>.
[45] *Class Action Complaint Flores et al v Motorola Solutions Inc. et al,* 1:2020cv01128, US District Court for the Northern District of Illinois, 14 February 2020.

photos or 'mugshots' which is expanding all the time".[46] In addition, these companies profited from individuals' biometrics as the database was being used as a "facial search engine" by various law enforcement agencies for a fee. The database is used in "other facial recognition products thereby allowing the identification and tracking in real time and near-real time of millions of people…wherever they may go." In addition, Motorola and Vigilant stored the biometric scans in a database, without publicly available information about the policies for the database, which is required in BIPA. Defendants also did not inform individuals that their biometrics were being collected and for what purpose, nor did they obtain consent, which are all required by BIPA.

Brazil- Facial recognition and the right to privacy - *IDEC v ViaQuatro* (2021)

The present case is a Public Civil Action in which the plaintiffs and defendants are, respectively, the Brazilian Institute of Consumer Protection (IDEC) and the concessionaire of the São Paulo S.A. subway line 4 (ViaQuatro).[47] IDEC filed a public civil action against ViaQuatro regarding the installation and use of an AI crowd analytics from Ad Mobilize that claims to predict the emotion, age, and gender of metro passengers without processing personal data. The purpose of this case was to terminate the use of these digital interactive door systems along the subway, as well as to convict ViaQuatro for collective damages, related to the violation of consumer rights and personal data legislation of those that used the subway network's system. In the end, the 37th Civil Court of São Paulo condemned ViaQuatro for the irregular collection of data of passengers by means of facial recognition.[48]

[46] *Class Action Complaint Flores et al v Motorola Solutions Inc. et al,* 1:2020cv01128, US District Court for the Northern District of Illinois, 14 February 2020, §7

[47] Idec (Instituto Brasileiro de Defesa do Consumidor) v Via Quatro (Concessionaria da Linha 4 do Metro de Sao Paulo S.a.), Civil Public Class Action - Railroad Transport, No 1090663-42.2018.8.26.0100, State of São Paulo Court of Appeals, District Courts Of São Paulo Civil Central Venue, 37th Civil Court, 10 May 2021.

[48] IDEC (Instituto Brasileiro de Defesa do Consumidor) v Via Quatro (Concessionaria da Linha 4 do Metro de Sao Paulo S.a.), Civil Public Class Action - Railroad Transport, No 1090663-42.2018.8.26.0100, State of São Paulo Court of Appeals, District Courts of São Paulo Civil Central Venue, 37th Civil Court, 10 May 2021, p. 12

8.6 MODULE 6: AI ETHICS & GOVERNANCE CONCERNING JUDICIAL OPERATORS

8.6.1 Module Focus

This module will provide a review of the Responsible AI landscape, including key ethical and governance issues, through the lens of justice. First providing an overview of key issues in the 'AI ethics' global dialogue, it will introduce AI-specific principles created to guide society on how AI should be developed and deployed and suggest how judicial stakeholders can operationalize these principles. It will also provide an overview of AI governance initiatives that have recently been launched and explore the role of justice in these initiatives. Presenting the example of Africa, it will look at how justice systems across a specific region are being impacted by the Responsible AI landscape. It will conclude by highlighting the role of methodologies and measurements for helping stakeholders, including judicial operators, adequately understand the adoption and implications of AI-based systems.

8.6.2 Learning Objectives

From this module, participants will:

- Familiarize themselves with concepts and principles related to AI ethics
- Build an understanding of AI governance approaches and initiatives, including the UNESCO Recommendation on the Ethics of AI
- Understand the relevance of metrics and methodologies to measure the impact of AI in the judiciary

8.6.3 Lecturers

Nicolas Economou | The Future Society, Chair of the Law and Society Initiative; Chair, Law Committee, IEEE Global Initiative| USA
Peter-Paul Verbeek | University of Twente, Professor; Chair of UNESCO's COMEST; Vice Chair of UNESCO's Ad Hoc Expert Group on AI Ethics | The Netherlands
Hon. Jean Aloise Ndiaye | Supreme Court of Senegal | Senegal
Tatiana Jereissati | CETIC.br | NIC.br, Coordinator of UNESCO Projects | Brazil

8.6.4 Structure

Lecture 6.1: What do 'AI Ethics' and 'Responsible AI' mean for justice?

This presentation will introduce module 6 and its objectives. How can we know when to trust AI to advance justice, access to justice, and through the institution of the law, our shared humanist ideas? This presentation will try to answer these questions on the basis of three elements: (1) the codification of shared values and ideas, (2) AI-specific principles providing guidance on how to design and deploy AI systems, and (3) ensure the assessment of trustworthiness in a real-world context. In questioning the trustworthiness of AI-enabled legal fact finding to achieve just outcomes, the lecturer will use the IEEE EAD Law Chapter framework as an example to provide practical suggestions for judicial operators on how to embed the principles of effectiveness, competence, accountability, and transparency in their mandates to uphold the rule of law.

Lecture script (WebVTT)

AI Ethics and Governance Concerning Judicial Operators, Lecture one, what do AI ethics and responsible AI mean for justice? Hello. It is an honor to present to you this introductory section to the module on AI ethics and governance. You, the judges of the world represent an essential institution when it comes to protecting and advancing human rights in the digital age. This module has four learning objectives. First, to introduce you to responsible, ethical, and trustworthy AI. Two, to go over certain salient AI governance initiatives. Three, to introduce you to methodologies and measurement for understanding the impact of AI and, four, to share some thoughts on the impact of AI on judicial systems. Ethical and responsible AI in the law can be summarized in one salient question. How can we know when to trust and when to mistrust AI, to advance justice, access to justice and through the institution of the law, our shared values and ideals?

Answering this question requires three elements.

First, a codification of the high values and ideals that we seek to advance in society. Two, AI-specific principles to provide guidance on how to design and deploy AI systems that actually protect and advance those higher-level values and ideals. And three, a way to ensure that we can assess, in the real world, the extent of which we can trust or mistrust these AI systems. So let us look at each one of those individually. First, with respect our shared values and ideals, while those have been codified in documents such as the Universal Declaration of Human Rights or in the case of Europe, the European Convention on Human Rights. But these documents don't provide with any guidance that is AI specific. And this is why, coming to our second element, organizations such as UNESCO, The European Commission, The Council of Europe, and its Commission on the Efficiency of Justice, the OECD, and the Global Standard-Setting Body, IEEE, have actually promulgated certain AI-specific principles of which you see here certain examples. Take privacy, we all know that AI systems rely on a lot of data in order to make decisions. This data often involves personal data so it is essential that these systems be designed and deployed in a way that respects privacy. And, finally, coming to our third element, we need an operative definition of trustworthiness. So again, we can assess the real world to the extent we can trust or mistrust these AI systems.

The Global Standard-Setting Body IEEE promulgated such a definition, which is very simple and robust and

relies on four constituent components or as we've come to call them trust conditions. First, effectiveness, is AI effective? That means meeting its intended objectives. So, for example, in avoiding bias in facial recognition. Two, competence, are those who procure and operate AI, for example, competent at the job that they do? We have standards of competence in so many domains, from medicine to aviation. What we need to know is that operators of AI, those who procure it, those who design it, are competent to do their work. Three, accountability. When things go wrong and AI causes harm, we need to be able to apportion responsibility to the human agents involved. And finally transparency, do we have the information we need to understand how AI systems make decisions? Let me illustrate this framework through a case study of AI in legal fact-finding. As something quite common in the United States, where it is known as Technology Assisted Review, or TAR for short. AI legal fact finding is used to identify information to prove or disprove certain claims in, for example, litigation or investigations. AI, as by way of example, is used in view of very often dozens of investigators or dozens of lawyers to analyze, to review, to look through often millions of corporate emails or millions of corporate documents to try to identify evidence of lets say corrupt activities in international bidding. Did certain corporate executives have conversations that were inappropriate with local officials?

So what values are implicated when we deploy AI legal fact finding? Well, I believe that essential values are implicated, such as, for example, achieving just outcomes. If AI can really be trusted to identify all the relevant information and facts that US judges need to adjudicate the matters in front of you, then unquestionably justice will be served. And there are other essential values implicated, but we come back to the central question then, which is, can AI legal fact-finding, can TAR really be trusted to advance those values?

Thankfully, in this domain, we have about two decades of practice and also some scientific evidence to help us assess AI legal fact-finding against IEEE trust conditions. Let me take those one by one. First of all, is TAR effective? Is AI legal fact finding effective in finding the information and the facts are being solved. The United States National Institute of Standards and Technology conducted a series of studies trying to answer that question. And without going into detail, you see here three years of results of these studies plotted.

The systems that did well, that identified most of the information that was being sought and the US judges would want to be aware of, for example, or where investigators might probably be aware of are plotted on the upper right of this chart. Those systems that did not perform well, that missed most of the information that they were intended to find, are plotted on the lower left of this chart. And as you can see, many systems are somewhere in the middle. So based on that, I think we can conclude that we can't generally trust AI systems to perform this task and therefore that we need to proceed with caution.

This study is also provided some interesting information that, in my view, suggests something important about the competence of the operators of these AI systems. What this chart plots is, without going into detail, is the self-reported accuracy of the participant in some of those studies, and that self-reported accuracy is connected through a red light that you see here to the actual accuracy that a list determined that these entrants had achieved. So, as you can see, sometimes the red line is short, which means participants were quite good and therefore quite competent at measuring how well they did, how they created their own AI systems. But sometimes you can see the line is very long, meaning that they were not really able to properly measure their own accuracy. And so, if what we can infer from those studies is that many of the operators of these AI systems

were not quite comfortable with measuring their own accuracy. Can we truly deem them to be competent to use these AI and operate these AI systems in the first place? The answer is that we probably here as well need to proceed with caution. It's the same answer with transparency. I don't have the time to go into great detail but sometimes transparency is well understood and well applied. Sometimes less so. So, we have to proceed with caution. And if we need to proceed with caution with respect to this first three trust conditions, I think it's fair to say that we know it would be hard to apportion responsibility and hold human agents accountable. And therefore, I think here as well, we need to proceed with caution.

So, what can we learn from this case study? First is that AI can actually advance the functions of the law and the values that animate the law. Some systems did well. Some operators are competent. And so, we can aspire when we achieve general trustworthiness of the system to their ethical adoption in the legal system and the judicial systems, we will be able to trust that they're actually advancing our shared humanist values and ideals as codified again, for example, in the Universal Declaration of Human Rights. But we know that today the real-world practices are inadequate, so what needs to be done to go from today's date of generally inadequate or uncertain practices to trustworthiness. We need, I think, two things which you can all trust. First, we need more data, such as the studies that showed you and two, we need standards and certifications, standards for technologies, certifications for the operators of these technologies as we have, for example, in aviation where both technologies meets with the standards. And certainly, pilots and mechanics and so on are certified to be competent through their work. So, what can you do in the meantime, as judges, to support the ethical adoption of AI or to support the avoidance of AI when needed?

Well, it is to ask for some evidence of the extent to which AI indicated in matters in front of you, or AI that might be asked to consider using your judicial systems. It is to again ask for sound evidence of the extent to which these systems meet the IEEE four trustworthy conditions. Thank you very much again for the opportunity to present this introductory section to the AI ethics and governance module to you.

Relevant Readings / Activities:

- IEEE EAD V1 Law Chapter: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_law.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined
- Morley, J., Floridi, L., Kinsey, L. et al. (2019) *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices.* Sci Eng Ethics 26, 2141–2168 (2020). Available at: <https://doi.org/10.1007/s11948-019-00165-5>.
- Jobin, A., Ienca, M. & Vayena, E. (2019). *The global landscape of AI ethics guidelines.* Nat Mach Intell 1, 389–399 (2019). Available at: <https://doi.org/10.1038/s42256-019-0088-2>.
- The Future Society (2020). *Areas for Future Action in the Responsible AI Ecosystem.* Available at: <https://thefuturesociety.org/wp-content/uploads/2021/02/Areas-for-Future-Action-in-the-Responsible-AI-Ecosystem.pdf>.
- Fjeld, J., Nagy, A., (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI.* Available at: <https://cyber.harvard.edu/publication/2020/principled-ai>.

Lecture 6.2: What are emerging AI governance initiatives and how do they impact justice?

This presentation will provide an overview of key governance approaches and initiatives that have sprung to address key AI ethical issues. Governance approaches will include legislation, standardization, and soft governance mechanisms. The recently adopted UNESCO Recommendation on the ethics of AI will be presented as a concrete example to explore the role of judicial operators in the AI governance landscape.

Lecture script (WebVTT)

What are emerging AI governance initiatives and how do they impact justice? How to govern AI in an ethical way? That's a question with which many companies, nations, and societal organizations have been struggling over the past few years. The year 2019 has already gone into history already as the year of AI ethics. In many places all over the world, people have tried to develop codes to deal ethically, to govern AI in an ethical way. In these codes typically three words return, three key words, which are sometimes called the

F.A.T. Ideas to the F.A.T principles. Fairness, accountability and transparency. So, the impact of AI on society needs to be fair and just and people dealing with AI systems need to be able to be held accountable for what they do. The systems themselves, need to be somehow open about how they arrive at their analyses and also about the datasets with which they were trained. So, this has resulted in a huge set of ethical codes, but also, in fact, in a growing discomfort with these codes because however interesting they are and however helpful they are to assist people in analyzing the situation. At the same time, they are quite remote and abstract. They didn't always give enough guidance to companies and to people to make choices in practices. At the same time, also the ethical discussion worldwide became bigger and bigger. Especially within the European Union, there was a lot of attention for AI ethics, which resulted in a framework that was called 'Trustworthy AI.' The EU wanted to make sure that all AI systems can be trusted by people, which meant that there needs to be ethics designed into the systems. And we also need an ethical form of somehow implementing AI in society and regulating this and using it in an ethical way. Interesting discussion, a big framework that's quite influential. At the same time, it was a discussion that also worked as some kind of a polarization within the world because it sometimes was framed as the "third way" with the old language of the Cold War, an alternative between ethical frameworks from the Western part of the world and the Eastern part of the world. I think at this point in the discussion, UNESCO entered the discussion with an attempt to arrive at the framework for the entire planet. An ethical framework that could help all member states of UNESCO to guide and govern AI systems ethically. The first thing that UNESCO did was to ask COMEST, its World Commission for the Ethics of Science and Technology, that I am chairing at the moment, to write a preliminary study on the ethics of AI.

And that's what we did. We analysed, in fact, how AI is affecting the mind. Because the mind is the key area of attention of UNESCO. UNESCO is all about education, science, culture, and communication. It's all faculties of the mind for a very good reason, you could say. In the constitution of UNESCO, it actually says that since wars start in the minds of men and women, the foundations for peace can be constructed there as well. So we studied how AI is affecting education and how we need to teach children to think somehow critically about AI systems. Also, how it's affecting scientific explanation or cultural diversity, the future of smaller languages and of artistic creativity, and the freedom of press, for instance.

And on the basis of this, within UNESCO, it was decided that it would be an important idea to work towards a recommendation and normative instrument for the entire world to deal ethically with AI systems. So, an ad-hoc expert group was formed, and that group developed a framework for AI ethics and hopefully in November 2021 (I'm recording this video before that date), it will be adopted by all member states of UNESCO. So in this recommendation, UNESCO tries to formulate not only values, principles to base policies on, but especially also to

make them concrete and specific policy areas and I think that's the most important thing that this recommendation does. A wide array of areas ranging from ethical impact assessment and ethical governance, stewardship, to data policy development and international cooperation, but also issues of environment, ecosystems, gender, culture, education, communication and economy, labor, health, wellbeing a huge set of policy areas, then all of which UNESCO gives recommendations. And the recommendations come from a global ethics point of view. An ethical point of view that aims to work for the entire world, you could say.

Let me briefly highlight three important elements of it. First, there is this ethical impact assessment. This is maybe the key element that we need to translate the abstract ethical values, principles and codes to practices so against a unified framework, typically also a legal framework, the Human Rights Declaration, for instance, we need also local ethical forms of reflection to really let ethics guide the development of implementation and use of AI systems. Very important element of the recommendation. Second, gender. Justice fairness has always been a central issue in AI ethics. Gender inequality is, of course, a huge problem around the world, and this is where we can now make a difference with AI ethics, and the same holds true for ecosystems and sustainability. AI systems have a lot of impact on the planet. They consume a lot of energy. At the same time, they can help scientific research to deal in a more responsible way with these issues. There's much more to be said about the recommendation, but what it does show is that the future of humankind is, in a sense, in our own hands if we can deal in a responsible way with AI systems and all these policy areas. I really hope that you can find the time to read it, and I want to thank you very much for your attention.

Relevant Readings / Activities:

- IEEE Global Initiative on Ethics of A/IS - Ethically Aligned Design
- UNESCO (n.d.). *Elaboration of a Recommendation on the Ethics of Artificial Intelligence.* Available at: <https://en.unesco.org/artificial-intelligence/ethics>.
- Council of Europe (n.d.). *CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment.* Available at: <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>.
- OECD National AI Policies & Strategies

Lecture 6.3: What are the implications of AI governance frameworks in African justice systems?

This presentation will explore the impact of AI on society and justice systems, and AI governance initiatives across Africa. Recently, the government of Senegal launched the judicial digitization movement, and the presentation will indicate how other African countries, from Côte d'Ivoire, Benin to Kenya can benefit from the Senegalese example. The video will discuss both benefits and threats aligned with using AI in African judicial systems from privacy issues, profiling, ethnic and racial discrimination to mass-surveillance. Finally, it will consider the need and ways to govern the use of AI in the judicial context, in Africa, through cooperation between countries.

Lecture script (WebVTT)

What are the implications of AI governance frameworks of AI in African justice systems?

Ladies and gentlemen, good morning, I am particularly pleased to share this presentation with you. Welcome to the module on artificial intelligence and the governance of judicial actors in Africa. As you know, and this is just a reminder, artificial intelligence was created in the 1950s and aims to mimic the behavior of the human brain or at least its logic when it comes to making certain decisions. Artificial intelligence uses the power of algorithms, i.e., the mathematics, with their computing power that allow, with a mass of information that is available, to be able to process and classify them and to solve certain solutions. Artificial intelligence today is in all sectors of life. When you drive your car, in your home, with connected objects, at the state level, with cyber-surveillance, at the security level especially and at the level of health. But what we see is that artificial intelligence has been slow to penetrate the health sector and justice. Because we also know that justice has always been resistant to the use of information and communication technologies. It is only recently, that the movement of use of Information and Communication Technologies (ICT) has generated an interest in the use of artificial intelligence in the justice and we note this also at the African level. We note for example that recently, the Association of the High Courts of Cassation of the countries sharing the use of the French language called AHJUCAF, organized two years ago a colloquium international conference in Lebanon on open data for judicial decisions. This means that the member countries of this organization have a particular interest in the use of artificial intelligence in justice. Also, more recently, this same organization, with the collaboration of the Supreme Court of Senegal, has organized a workshop in Dakar that focused on the need to the need to harmonize the governance of information systems of the high jurisdictions through the use of open data on judicial decisions. And recently, the government of Senegal, on July 5, aware of the importance of intelligence and the digitization of procedures, officially launched the proceeded to the official launch of the digitization of judicial procedures at the level of Senegal. So, we see that this same movement can be observed in other African countries, whether it be in Ivory Coast, but also in Benin, which has come to learn from the Senegalese experience, but also in Kenya and Ghana. So, there is a real interest in using artificial intelligence in the governance of our countries and African judicial systems. There is ample justification for this. Because artificial intelligence offers some positive points to our positive points to our judicial systems faced with like slowness, like heaviness, but also the lack of predictability. So, the use of artificial intelligence today can be a factor that allows to get around these difficulties. And we will see it through the manifestations of artificial intelligence. I will give just one example: the open data of judicial decisions. Today, almost all of the Supreme Courts in Africa are creating databases for all the decisions that are made. The advantage: thanks to the calculator, thanks to the algorithms, we can make more accurate searches to refine searches in a given area and make it available to the actors. But on the other hand, citizens can also access to this judicial open data. That is, the accessibility to judicial decisions and to be able to control the decisions that are made by the courts and tribunals in the country.

This creates what can be called citizen control through the use of this open data. But also, artificial intelligence can be put at the service of judges and other actors through what is called decision support. Decision support is a process that helps to the unpredictability of judicial decisions. So, there is a mass of information that is available, whether it's court decisions, which is digitized upstream, into platforms that judges use through the search engine with the search engine with a system of keywords. And this mass information can be selected quickly and made available to the actors of judicial decisions to aid unpredictability of court decisions. Finally, another example is predictive justice. It hasn't penetrated our judicial systems yet, but it can't be ignored because the process of the use of artificial intelligence is a process that is inescapable. We are not yet at the stage of robots that are going to make decisions, but we see the example of the United States and in France recently, where today it is possible that when the disputes do not reach certain values, they are resolved by algorithms.

In an African context where our courts of justice lack of means, have budgetary difficulties, human resources difficulties, this can be used. So these benefits are important for African justice. But we must not hide the fact that there are also disadvantages that we must be aware of. These are the risks of artificial intelligence in the the justice system that we know: invasion of privacy, invasion of personal data, risks of personal data, risks of discrimination ethnic, racial, social with a profiling system. But also, risk of mass surveillance, especially in police investigations. That's why you have to go with tact and measure. There is a need to regulate the use of the use of artificial intelligence by the judiciary and this is still insufficient.

At national level there is no specific text. And at the community level we have just started some texts through the African Union and especially the last resolution of the African Committee on Human Rights, which Committee on Human Rights, which took a resolution to ask the states to take measures to regulate the use of artificial intelligence. Specifically, there is no text, but in general, we have all the texts at the community level, concerning the protection of personal data, whether it's at the level of the International Union of Lawyers (UIA), whether at the level of the Economic Community of West African States (ECOWAS), but also at the level of our states. In the end, what we need to remember is that for human use of artificial intelligence in the justice system, we need to create an African regulatory framework. And for that, there is a need to strengthen cooperation between the States in this field. So, this is what I wanted to share with you in a few minutes on this important theme of the use of artificial intelligence in the governance of judicial actors in Africa.

Lecture 6.4: What is the role of measurements and methodologies for understanding AI adoption?

This presentation will discuss how AI solutions are increasingly being adopted by different segments of society, including the judiciary, and how their implications are felt in social and economic spheres. It is important that the gap between research and adoption of AI is bridged, and that reliable and robust data are produced through commonly agreed methodologies. Monitoring and evaluation mechanisms should be in place to ensure that AI is being adopted ethically in the judiciary. Measuring the adoption of AI is a complex task but of great relevance to inform stakeholders and guide AI policies in the future.

Lecture script (WebVTT)

Lecture four, what is the role of measurements and methodologies for understanding AI adoption? Recent developments in digital technologies and AI in particular, are generating important social and economic changes in organizations, governments and industries. As discussed in this course module, AI-based systems offer a myriad of possibilities for the Rule of Law, but they also pose new ethical challenges that need to be addressed. In this context, it is crucial that; sound, timely and reliable data are systematically produced to provide a solid understanding of how AI is being used by the judiciary. This information is essential to maximize benefits and mitigate risks in the use of such technologies. Examples of relevant questions are, which AI tools are currently being used in the judiciary? What are the intended and unintended impacts of such use? Are good practices being followed? Which key challenges and opportunities should be addressed by policies in the field? And finally, which metrics should be defined for measuring the impact of AI in the judiciary?

AI measurement efforts are essential for informing stakeholders and guiding the development, implementation and monitoring of AI policies. However, as noted by the OECD, producing official statistics on AI usage through national surveys may be a complex task. Here are a couple of reasons for this. AI technologies are developing at a fast pace, so working on a single conceptual definition on AI for the purpose of statistical production is still a challenge for researchers, National Statistics Office, policy makers and other stakeholders. The measurability of AI is also an issue because AI is not a stand-alone technology. Rather, it co-exists and is embedded in other technologies. Therefore, AI-based systems are not all directly measurable, which means that the presence of AI may need to be inferred. It may also be the case that AI is embedded to the extent that it becomes invisible. Despite these challenges, it is important that an AI measurement agenda is developed both nationally and internationally to strengthen data collection on this topic.

Following commonly agreed conceptual definitions and methodologies, for example, will allow international comparability. Monitoring and evaluation mechanisms are also fundamental for understanding how AI is advancing the judicial sphere. In this regard, I would like to mention a recent study on the current use of AI in the Brazilian judiciary. This study was carried out by an international group of researchers from the School of International Public Administration and Columbia University, and it was conducted in partnership with the ITSREAL and the National Council of Justice in Brazil. Drawing from the findings of this study, the research document presents types of indicators that could be used in monitoring and evaluation framework for AI tools in the country. So these include, for example, indicators on AI use for specific tools. These comprise testing algorithm accuracy, reliability, security, robustness and outcome fairness. Also suggested are indicators for evaluating the governance system, which includes, for example, looking at the number of courts developing or using AI tools. It should be noted that countries are under increasing pressure to produce high quality data and in a timely manner. So for this reason, it is important that the wide range of data providers that compose the ICT data production ecosystem all work in collaboration to produce relevant data on the use of digital technologies.

So, in this regard, the Regional Center for Studies on the Development of the Information Society, Cetic.Br, which is a department of Nic.BR, has developed a new layered model for ICT data production. The top layer consists of data users and their very specific needs. The middle layer consists of traditional national statistical systems, as well as administrative data from the government and also a range of new private sector data sources. And the very bottom layer represents new data sharing opportunities through partnerships that are supported by trusted data platforms and agreements. These platforms, they aim to mitigate challenges faced by data producers who are using privately owned data sets, enabling a secure use in a trustworthy manner in compliance with personal data protection regulations. So different stakeholders from the public and the private sector will require the development of technical, legal and institutional frameworks for this objective. So in conclusion, AI solutions are increasingly being adopted by different segments of society, including the judiciary, and implications may be felt in social, economic, cultural and political spheres. In this scenario, it is important that the gap between research and adoption of AI is bridged and that reliable and robust data are produced with commonly agreed methodologies.

Monitoring and evaluation mechanisms should also be in place to ensure that AI are being adopted ethically in the judiciary. Measuring the adoption of AI is a complex task, but of great relevance to inform stakeholders and to guide AI policies in the future.

Relevant Readings / Activities:

- Montagnier, P. and I. Ek (2021), "AI measurement in ICT usage surveys: A review", OECD Digital Economy Papers, No. 308, OECD Publishing, Paris. Available at: <https://doi.org/10.1787/72cce754-en>.
- Brehm, K., Hirabayashi, M., Langevin, C., Munozcano, B., Sekizawa, K., Zhu, J., (2020). The Future of AI in The Brazilian Judicial System. Available at: <https://itsrio.org/en/publicacoes/the-future-of-ai-in-the-brazilian-judicial-system/>.
- Cetic.br Annual Report 2020. CETIC.BR|NIC.BR. Available at: <https://cetic.br/en/publicacao/cetic-br-annual-report-2020/>.

8.4.6 Module Assessment

1. According to ethical principles on how AI should be developed and deployed, AI shall:
    a. Respect privacy
    b. Ensure fairness
    c. Be transparent
    d. All of the above

2. What are the conditions of trustworthy AI promulgated by IEEE?
    a. Intelligence, robustness, fairness, and privacy
    b. Effectiveness, competence, accountability, and transparency
    c. Equality, diversity, inclusivity, and transparency
    d. Expressive, creative, robustness, and inclusivity

3. How can judges support the ethical adoption of AI?
    a. Collaborate with AI developers to build AI tools for law enforcement
    b. Apply an AI algorithm to assess how ethical another AI tool is or is not
    c. Ask for evidence of the extent to which ethical principles are met
    d. There is little they can do

4. What is a challenge for AI measurability in the judiciary?
    a. Lack of expertise
    b. No single definition of AI
    c. Restricting data protection regulation
    d. None of the above

5. What is not a key element of the UNESCO Recommendation on the Ethics of AI?
    a. *Focuses only on the Global South*
    b. Addresses gender inequality issues related to AI
    c. Promotes the use of ethical impact assessments
    d. Addresses sustainability and environmental issues related to AI

8.6.5 Use Cases and Examples

UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)
In 2021, the UNESCO Recommendation on the Ethics of AI was adopted as the first global standard setting instrument on AI by 193 Member States of UNESCO.[49]

The Recommendation identifies guiding values such as the respect, protection and promotion of human rights and fundamental freedoms and human dignity, environment, and ecosystem flourishing, ensuring diversity and inclusiveness, and living in peaceful, just and interconnected societies. Guiding principles consist of proportionality and do no harm, safety and security, fairness and non-discrimination, sustainability, right to privacy and data protection, human oversight and determination, transparency and explainability, responsibility and accountability, awareness and literacy, and multi-stakeholder and adaptive governance and collaboration. To translate these values and principles into action, the Recommendation identifies 11 policy areas in UNESCO's fields of competence. These areas cover ethical impact assessments, governance and stewardship, data policy, development and international cooperation, environment and ecosystem, gender, culture, education and research, communication and information, economy and labor, and health and social well-being.

EU Artificial Intelligence Act Draft (2021)
On 21 April 2021, the European Commission presented the Artificial Intelligence Act. The draft AI act is the first ever attempt to enact a horizontal regulation of AI, where the Commission proposes to establish a technology-neutral definition and to classify AI systems with different requirements and obligations tailored on a 'risk-based approach' based on the pyramid of criticality, with a modern, layered enforcement mechanism.[50] Meaning, among other things, that a lighter legal regime applies to AI applications with a negligible risk, and that applications with an unacceptable risk are prohibited. Thus, stricter regulations apply as risk increases. These range from non-binding, self-regulatory, soft law, and impact assessments accompanied by codes of conduct to heavy, externally audited compliance requirements throughout the life cycle of the AI system application.

OECD Principles on Artificial Intelligence (2019)
The OECD Principles on AI were adopted in May 2019 by its member countries when they approved the OECD Council Recommendation on AI.[51] They aim to promote Artificial Intelligence that is innovative and trustworthy and that respects human rights and democratic values.

According to these five principles; AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being; AI should be designed to respect the rule of law, human rights, democratic values and diversity; there should be transparency and responsible disclosure to ensure understandability; AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be assessed and managed; and lastly, organizations and individuals developing, deploying or operating AI should be held accountable. Consequently, the document provides five recommendations to governments to: facilitate investment in AI research and development; foster a digital ecosystem for AI data, compute and knowledge; shape an enable policy environment for AI; build human capacity and prepare for labor market transformation; and co-operate internationally in a multi-stakeholder manner for trustworthy AI. In February 2020, the OECD AI Policy Observatory (OECD.AI) was launched to help policy makers and other stakeholders implement the principles.[52]

European Commission for the Efficiency of Justice: European Ethical Charter on the Use of AI in Judicial Systems (2018)

---

[49] UNESCO Recommendation on the Ethics of Artificial Intelligence, SHS/BIO/REC-AIETHICS/202, 2021. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000380455>.
[50] EU Commission; Proposal for a Regulation of The European Parliament and of The Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 21 April 2021. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
[51] OECD; OECD Principles on Artificial Intelligence, Adopted May 2019. Available at <https://oecd.ai/en/ai-principles jo. OECD Council Recommendation>.
Available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
[52] OECD AI Policy Observatory, 2021. Available at <https://oecd.ai/en/>.

Acknowledging the potential of AI to improve the efficiency and quality of justice, the European Commission for the Efficiency of Justice (CEPEJ) formally adopted five fundamental principles entitled 'European Ethical Charter on the use of AI in the judicial systems and their environment' at the Strasbourg plenary meeting in December 2018.[53] These principles are: the respect for fundamental rights; non-discrimination; quality and security; transparency, impartiality and fairness; and the principle "under user control" to ensure that users are informed and in control of their choices.

European High-Level Expert Group on AI: Ethics Guidelines for Trustworthy AI (2019)
On 8 April 2019, the High-Level Expert Group on AI (AI HLEG) - tasked to advise the Commission on the implementation of the Commission's Strategy on Artificial Intelligence - presented Ethics Guidelines for Trustworthy Artificial Intelligence.[54] This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through an open consultation.

These Guidelines emphasize that striving towards trustworthy AI concerns not only the trustworthiness of the AI system but all actors and processes that are part of the system's socio-technical context throughout its entire lifecycle. To this end, trustworthy AI should be lawful, respecting all applicable laws and regulations; ethical, respecting ethical principles and values; and robust, both from a technical perspective while taking into account its social environment. The Guidelines also put forward seven requirements that AI systems should meet to be deemed trustworthy, accompanied by a specific assessment list to verify trustworthiness. These requirements include human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability.

The EU Parliament Resolution with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies (2020)
On 20 October 2020, the European Parliament adopted a Resolution with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies.[55] Main points of the resolution focused on the need for; Human-centric and human-made artificial intelligence; risk assessment; safety features; transparency and accountability; Non-bias and non-discrimination; social responsibility and gender balance; environment and sustainability; privacy and biometric recognition; good governance and so forth.

Guidelines of the Committee of Ministers of the Council of Europe on online dispute resolution mechanisms in civil and administrative court proceedings (2021)
More recently, on 16 June 2021, the Committee of Ministers of the Council of Europe set guidelines on Online Dispute Resolution (ODR) mechanisms in both civil and administrative court proceedings.[56]

The Guidelines are based on the fundamental principles that; member states should seek to ensure trust and confidence in ODR and that it should not create barriers for access to justice; procedural rules which apply to court proceedings in general should also apply to court proceedings involving - ODR unless the specific nature of a particular ODR mechanism requires otherwise; and parties to proceedings involving the use of ODR should be identified using secure mechanisms. The guidelines aim to provide guidance in relation to fair procedure; transparency in the use of ODR and requirements for hearings; special issues related to the ICT nature of ODR

[53] The European Ethical Charter of the CEPEJ on the use of artificial intelligence (AI) in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018), Council of Europe.
Available at <https://www.unodc.org/res/ji/import/regional_standards/ethical_charter/ethical_charter.pdf>.
[54] High level expert group on artificial intelligence (AI HLEG); Ethics Guidelines for trustworthy AI, EU Commission, 8 April 2019. Available at <https://www.ccdcoe.org/uploads/2019/06/EC-190408-AI-HLEG-Guidelines.pdf>.
[55] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)), Adopted 20 October 2020, Brussels. Available at <https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html>.
[56] European Committee on Legal Co-operation (CDCJ); *Guidelines of the Committee of Ministers of the Council of Europe on online dispute resolution mechanisms in civil and administrative court proceedings*, 1407th meeting, 16 June 2021, CM(2021)36add4-final, 16 June 2021. Available at <https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a2cf96>.

techniques and other. They do not cover internal management of electronic case files by the courts or alternative dispute resolution (ADR) mechanisms, such as mediation and conciliation.

Council of Europe Ad Hoc Committee on Artificial Intelligence (2021)

The 2021 report of Strasbourg entitled "Possible elements of a legal framework on artificial intelligence" contains the outcomes of the work of the Council of Europe Ad Hoc Committee on AI (CAHAI) on the potential elements of a legal framework for the development, design, and application of AI, based on the Council of Europe's standards on human rights, democracy, and the rule of law.[57]

The CAHAI observes that the application of AI has the potential to promote human prosperity and individual and social well- being by enhancing progress and innovation, yet at the same time recognizes the potential risks to human rights, democracy, and the rule of law. To effectively prevent and/or mitigate these risks, the CAHAI considers that an appropriate legal framework on AI, based on the Council of Europe's standards on human rights, democracy, and the rule of law, should take the form of a legally binding transversal instrument. Such a binding instrument should focus on preventing and/or mitigating risks emanating from AI with the potential to interfere with the enjoyment of human rights, the functioning of democracy in the observance of the rule of law all while promoting socially beneficial AI applications. According to the Committee, this instrument should be underpinned by a risk-based approach meaning that the legal requirements to the design, development and use of AI should be proportional to the nature of the risk they pose to human rights, democracy, and the rule of law.

8.7 Path Forward

Concluding videos by UNESCO and The Future Society, thanking participants for their engagement, and congratulating them on the successful completion of the course. The videos also share further pathways for engagement in both the UNESCO Global Judges Initiative and the Athens Roundtable on AI and the Rule of Law.

9. Course Code of Conduct

The course engages diverse stakeholders and addresses a multicultural and multilingual audience. It is dedicated to providing a harassment-free learning experience for everyone; hence participants are asked to commit to the following acceptable and expected behavior:

- Treating one another with respect and dignity
- Respecting the differences in people, their ideas & opinions
- Demonstrating honesty and integrity
- Resolving conflicts and disagreements constructively and respectfully
- Being active and engaged participants during the course

[57] CAHAI; *Possible elements of a legal framework on Artificial Intelligence*, 6th and final plenary meeting on 30 November to 2nd of December 2021.

10. Course Evaluation

The course's intended outcome is to train at least 2000 judicial operators across at least 20 countries, with a particular focus on gender diversity and stakeholders from the Global South. Ultimately, the course should contribute to the achievement of Sustainable Development Goals 16.3, 16.10, 16.a, 16.b and 17.

Upon registration, participants will be asked to submit data regarding their demographics and background relative to the topic. Upon completion, participants will be asked to complete a survey to evaluate their satisfaction with the course. The information will be analyzed to develop an evaluation report.

11. Acknowledgments

The course is a product of collective intelligence and would not be possible without:

- the collaboration between UNESCO, The Future Society, Cetic.br/NIC.br, the National Judicial College, and IEEE.
- the generous support of the UNESCO Multi Donor Programme on Freedom of Expression and Safety of Journalists (MDP) and the Open Society Foundations.
- the invaluable contributions and presentations from the 21 module lecturers representing 14 different countries.
- the thoughtful input and feedback into the course curriculum from a diverse group of over 30 experts worldwide.
- the guidance of the UNESCO team. A special thanks to Vanessa Dreier, Prateek Sibal, Charline d'Oultremont, and Jacinth Chia.
- the dedication and spirit of The Future Society team. A special thanks to Niki Iliadis, Katerina Panagiotidou, Nicolas Miailhe, and the TFS Affiliates who contributed.
- the creativity of Beyond Borders Media to produce the visual design and recording for each lecture. A special thanks to Sophie Green and Reinier van Oorsouw.