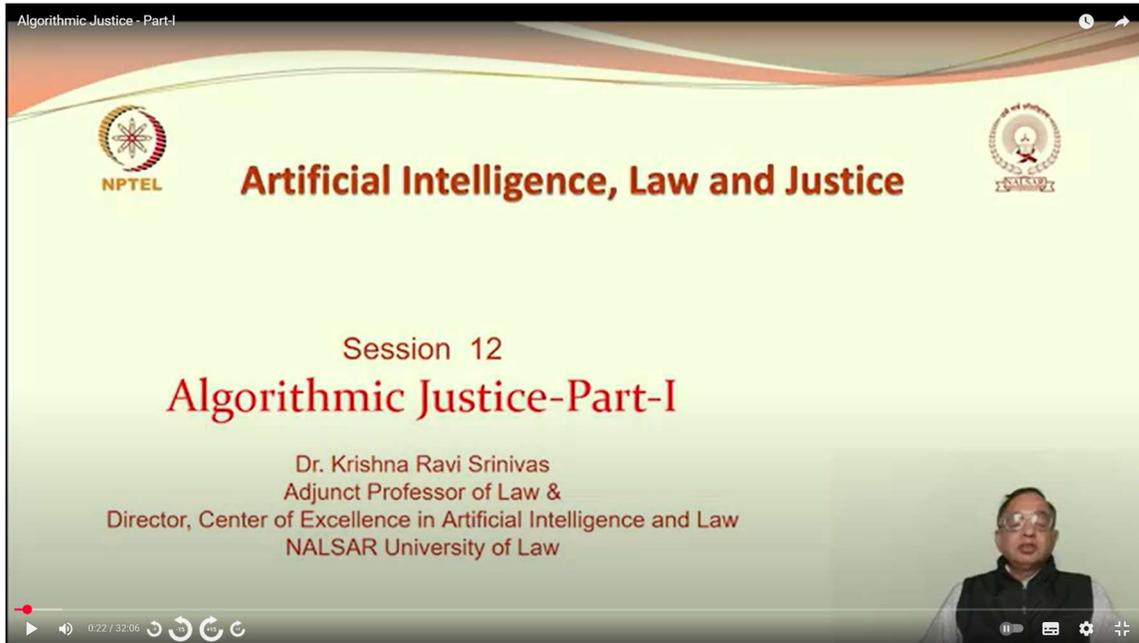


Course Name – Artificial Intelligence, Law and Justice
Professor Name – Dr. Krishna Ravi Srinivas
Department Name – Center of Excellence in Artificial Intelligence and Law
Institute Name – NALSAR University of Law
Week – 03
Lecture – 12



Algorithmic Justice - Part-I

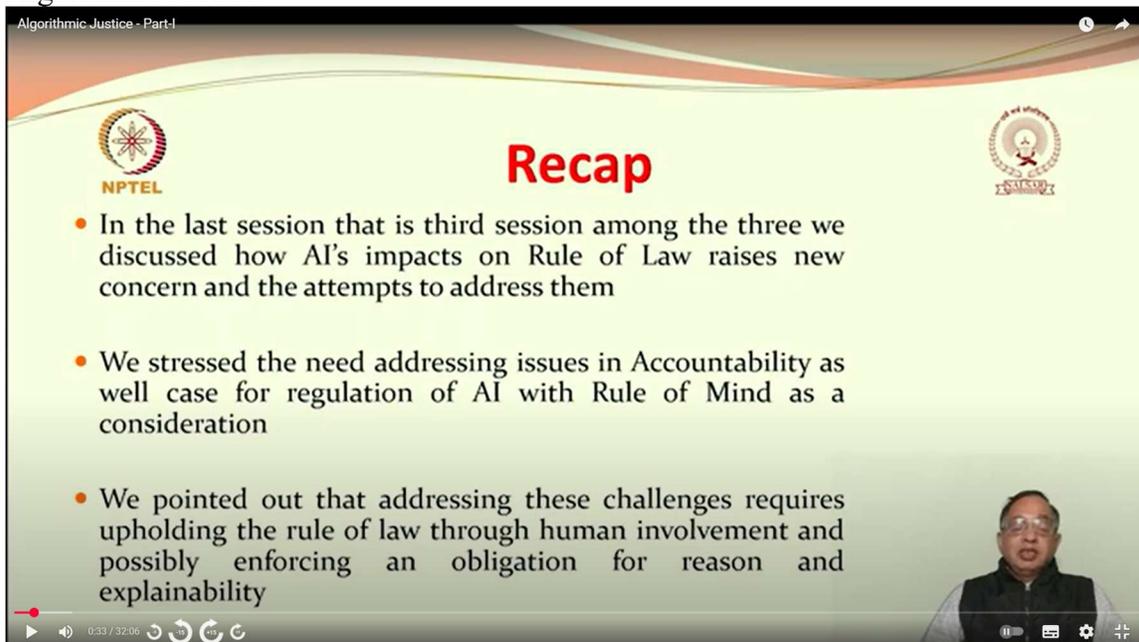
 **Artificial Intelligence, Law and Justice** 

Session 12
Algorithmic Justice-Part-I

Dr. Krishna Ravi Srinivas
Adjunct Professor of Law &
Director, Center of Excellence in Artificial Intelligence and Law
NALSAR University of Law

0:22 / 32:06

Artificial Intelligence Law and Justice Course, Session 12, Part One of Two on the Topic Algorithmic Justice.



Algorithmic Justice - Part-I

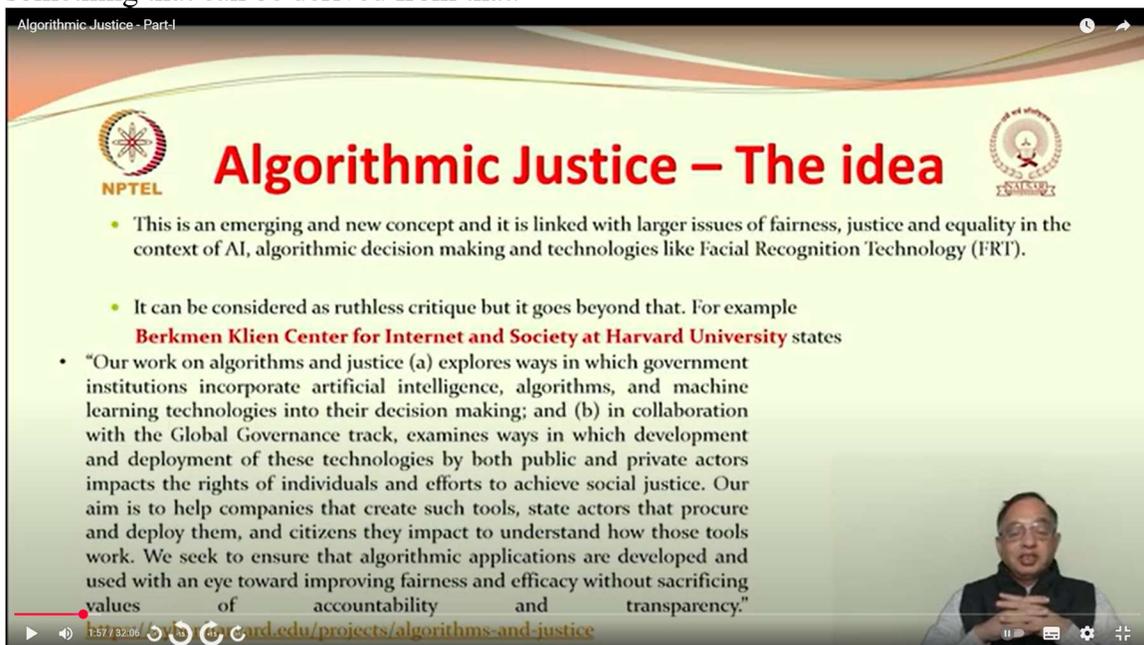
 **Recap** 

- In the last session that is third session among the three we discussed how AI's impacts on Rule of Law raises new concern and the attempts to address them
- We stressed the need addressing issues in Accountability as well case for regulation of AI with Rule of Mind as a consideration
- We pointed out that addressing these challenges requires upholding the rule of law through human involvement and possibly enforcing an obligation for reason and explainability

0:33 / 32:06

Recap, in the last session, which was the third session we had among the three, we discussed AI's impact on the rule of law and how it raises many concerns, as well as the ways to address them. In particular, we stressed the need to address the issues of

accountability as well as care for the regulation of AI with the rule of law as a consideration. We pointed out that addressing these challenges requires upholding the rule of law through human involvement and possibly enforcing an obligation for reason and explainability. In the sense that we had a look at how AI can impact the idea of the rule of law in both theory and practice and what the implications of that are. In particular, when there are issues, such as AI can adversely affect the basic principles of the rule of law when applied to daily practices. In that, we looked more from a rule-of-law perspective, particularly at how AI systems can affect it. In this session, we are going to look at another concept, which is a very interesting and evolving concept called algorithmic justice. We all know about justice, but we have also discussed algorithms, particularly algorithmic decision-making. So, we can connect with what we have already read or taught in algorithmic decision-making. Algorithmic justice is a sequel to that or is something that can be derived from that.



The screenshot shows a video player interface for a lecture titled "Algorithmic Justice - Part I". The slide content includes the NPTEL logo on the left and the Harvard University logo on the right. The main title is "Algorithmic Justice – The idea" in red. The slide contains the following text:

- This is an emerging and new concept and it is linked with larger issues of fairness, justice and equality in the context of AI, algorithmic decision making and technologies like Facial Recognition Technology (FRT).
- It can be considered as ruthless critique but it goes beyond that. For example **Berkman Klein Center for Internet and Society at Harvard University** states
 - “Our work on algorithms and justice (a) explores ways in which government institutions incorporate artificial intelligence, algorithms, and machine learning technologies into their decision making; and (b) in collaboration with the Global Governance track, examines ways in which development and deployment of these technologies by both public and private actors impacts the rights of individuals and efforts to achieve social justice. Our aim is to help companies that create such tools, state actors that procure and deploy them, and citizens they impact to understand how those tools work. We seek to ensure that algorithmic applications are developed and used with an eye toward improving fairness and efficacy without sacrificing values of accountability and transparency.”

At the bottom of the slide, there is a video player control bar showing a timestamp of 1:57 / 92:06 and a URL: harvard.edu/projects/algorithms-and-justice. A small video inset in the bottom right corner shows a man speaking.

So, we need to define what algorithmic justice is. Algorithmic justice is a new and emerging concept, and it is linked to a larger issue of fairness, justice, and equality in the context of AI. So algorithmic decision-making and technologies like facial recognition technology. We should keep in mind that when we talk about algorithmic justice, we are not talking about mere algorithmic decision-making in some sectors, specific examples, or applications. We are talking about a broader issue wherever algorithms are being used, including facial recognition technology, and how that decision-making can adversely impact people's rights and their entitlements. It can be considered a ruthless critique, but it goes beyond that. For example, the Berkman Klein Center for Internet and Society at Harvard University is also working on algorithmic justice, but the way of working is not just critique. They also work with companies and with other stakeholders to ensure that algorithmic justice is not something that remains a mere negative critique, but they do a lot of work on making things better. So, the criticisms from algorithmic statistics are not there merely for the sake of criticism. Many of them are, in fact, expressed with a lot of concern about the rule of law, fairness, due process, and what every legal person will tell is that the fundamental issues which the law has to deal with or the fundamental issues

where the legal principles have to be applied without any exception. So, having said that, let us also look at how it has emerged.

Algorithmic Justice - Part I

Origins and Definition

- The idea of Algorithmic Justice (AJ) was popularized by Dr. Joy Buolamwini whose experiences with FRT's showed that the technology applications were biased and her face was recognized only when she wore a white mask. Based on her experience, research and others experiences with AI (including FRTs) she wrote a book 'Unmasking AI' and founded Algorithmic Justice League (AJL). Her work was featured in a documentary 'Coded Bias' <https://sanford.duke.edu/story/dr-joy-buolamwini-algorithmic-bias-and-ai-justice/>
- Defining AJ Algorithmic Justice: "The application of principles of social justice and applied ethics to the design, deployment, regulation, and ongoing use of algorithmic systems so that the potential for harm is reduced" (Head, Fister, and MacMillan, 2024).
- Algorithmic Reparations: Drawing from intersectionality critical race theory, Davis, Williams, and Yang argue that we must "name, unmask, and undo allocative and representational harms as they materialize in sociotechnical form." They suggest algorithmic reparations as "a foundation for building, evaluating, and when necessary, omitting and era <https://library.highline.edu/c.php?g=1401364&p=10368769>

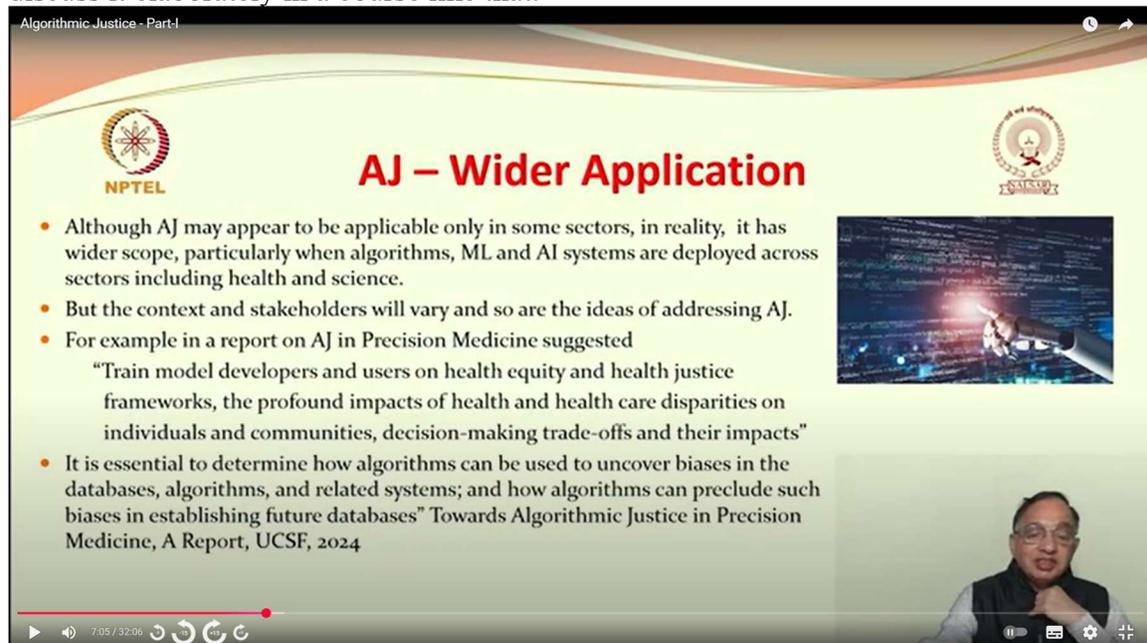
3:33 / 32:06

This idea of algorithmic justice is not very old. It is quite new. In fact, it is not even a decade old. This was popularized by Dr. Joy Buolamwini, whose experiences with facial recognition technology showed that technology applications were biased and that her face was not recognized unless she wore a white mask. See, the problem was that she was a scientist; she was educated at Georgia Tech, but then she faced a unique problem. The facial recognition technologies did not recognize her face at all, but they only recognized her when she wore a white mask. So, she did in-depth research and then found out that the problem was the AI algorithms or the algorithms that were used to train the facial recognition technologies were simply biased. So, she wrote a fascinating book called *Unmasking AI* and also founded an Algorithmic Justice League. It is an NGO. And then her work can be viewed in *Coded Bias*, which is also available on Netflix. So, she started describing algorithmic justice in the sense that she made it very popular. But then the discussions about algorithms, unfairness, accountability, and reasonableness have been present for quite some time now, at least for a decade in the legal literature, administrative law, and elsewhere.

We can have a crisp definition of algorithmic justice. The application of principles of social justice and applied ethics in the design, deployment, regulation, and ongoing use of algorithmic systems is intended to reduce the potential for harm. So algorithmic justice looks at the potential for harm and tries to mitigate it or minimize it. At the same time, algorithmic justice is a very broad concept. It is not focused on any discrimination or bias against a single group, nor is it focused on any single technology. So, it is technology-neutral. In that sense, algorithmic justice can be expanded to or applied wherever, irrespective of the application or systems, algorithms are used for decision-making, particularly in the criminal justice system and everywhere else. So, this concept is an emerging one that has huge applications due to algorithmic decision-making becoming more and more popular and widespread. When we talk about algorithmic justice, there should be a counter concept called algorithmic reparations. So, drawing from

intersectionality and critical social theory, Davis, Williams, and Yang argue that we must name, unmask, and undo allocative and representational harms, lest they materialize in the sociotechnical form.

So, they suggest that algorithmic representation serves as a foundation for building, evaluating, and, when necessary, permitting the error. But the problem here is that algorithmic reparation is not a well-known concept. It is not well accepted in the sense that not many people have worked on it. But it is a very interesting concept. But we are going to focus only on algorithmic justice here for the simple reason that discussing algorithmic reparation is not part of the syllabus. Again, algorithmic reparation is not discussed very much in the literature. And then the third most important point is that the very idea of algorithmic reparations is not fully developed in the sense that we can now discuss it elaborately in a course like that.

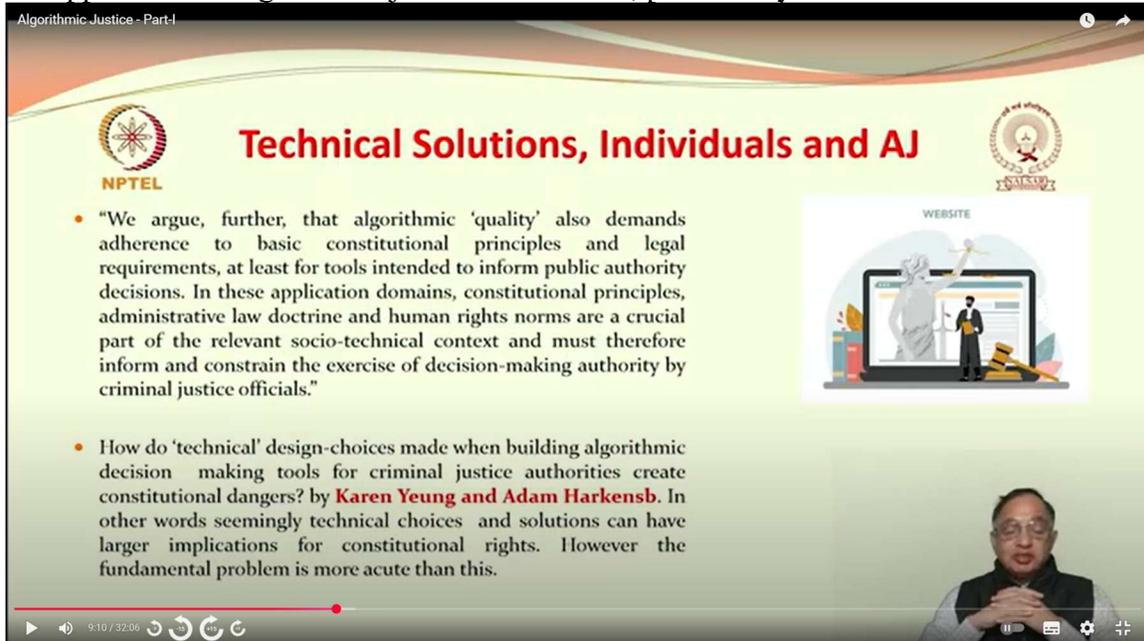


The screenshot shows a video player interface. At the top left, it says "Algorithmic Justice - Part I". The slide content includes:

- NPTEL** logo on the left and a university logo on the right.
- AJ – Wider Application** in red text.
- A list of bullet points:
 - Although AJ may appear to be applicable only in some sectors, in reality, it has wider scope, particularly when algorithms, ML and AI systems are deployed across sectors including health and science.
 - But the context and stakeholders will vary and so are the ideas of addressing AJ.
 - For example in a report on AJ in Precision Medicine suggested
“Train model developers and users on health equity and health justice frameworks, the profound impacts of health and health care disparities on individuals and communities, decision-making trade-offs and their impacts”
 - It is essential to determine how algorithms can be used to uncover biases in the databases, algorithms, and related systems; and how algorithms can preclude such biases in establishing future databases” Towards Algorithmic Justice in Precision Medicine, A Report, UCSF, 2024
- A small image of a hand pointing at a digital interface.
- A video thumbnail of a man speaking in the bottom right corner.
- Video player controls at the bottom showing a progress bar at 7:05 / 32:06.

So having seen the origins and definitions, we should see why we are looking at it. Are we talking about algorithmic justice only in the context of law and justice? No, because algorithms are being used in many disciplines, including AI in healthcare, AI in medicine, AI in diagnostics, AI in education, and more. So, we may think that the idea of algorithmic justice or AIJ is applicable only in some sectors. It is not so. So, the context and the stakeholders will vary, but the idea of addressing algorithmic justice will remain the same. For example, in a report on algorithmic justice in precision medicine, it was suggested to train model developers and users on health equity and health justice frameworks. The profound impacts of health and health care disparities on individuals, communities, decision-making, trade-offs, and their impacts. You might remember that in my earlier classes, I also mentioned how the problematic assumptions in algorithmic training, which is based merely on healthcare applications or diagnostics for selected populations, lead to the problem of leaving out certain populations in clinical trials. So, in a broader context, even in health, we need to look at the scope for algorithmic justice. It is essential to determine how algorithms can be used to uncover biases in the databases, algorithms, and related systems, and how algorithms can preclude such biases in establishing featured databases. So, this idea of algorithmic justice being used involves

using algorithms to identify what is going wrong with algorithms and then using algorithms to make things better. So, this report on algorithmic justice and precision medicine from the University of California, San Francisco, was published in 2024. It was based on the proceedings of a meeting, but it does have a lot of interesting information on the application of algorithmic justice in medicine, particularly in Christian medicine.



The screenshot shows a video player interface for an NPTEL lecture. The title is "Technical Solutions, Individuals and AJ". The NPTEL logo is on the left, and the VIT Vellore logo is on the right. The main content area contains two bullet points. The first bullet point discusses the need for algorithmic quality to adhere to constitutional principles and legal requirements. The second bullet point asks how technical design choices for criminal justice tools create constitutional dangers, citing Karen Yeung and Adam Harkensb. An illustration of a person at a computer with a scale of justice is on the right. A video feed of the lecturer is in the bottom right corner. The video player controls at the bottom show a timestamp of 9:10 / 32:06.

Algorithmic Justice - Part I

Technical Solutions, Individuals and AJ

NPTEL

- "We argue, further, that algorithmic 'quality' also demands adherence to basic constitutional principles and legal requirements, at least for tools intended to inform public authority decisions. In these application domains, constitutional principles, administrative law doctrine and human rights norms are a crucial part of the relevant socio-technical context and must therefore inform and constrain the exercise of decision-making authority by criminal justice officials."
- How do 'technical' design-choices made when building algorithmic decision making tools for criminal justice authorities create constitutional dangers? by **Karen Yeung and Adam Harkensb**. In other words seemingly technical choices and solutions can have larger implications for constitutional rights. However the fundamental problem is more acute than this.

WEBSITE

9:10 / 32:06

A key theoretician in this field is Professor Karen Young. She has written extensively on algorithmic justice, algorithms in administrative law, and more importantly, on technology and regulation. So, we will be citing her work often in this course, in this session, and the next one for the simple reason that she is a pioneer, and more importantly, her observations have been very acute to the point and have stirred a lot of debates in the sense that how we address this question. So, to begin with, her idea is that we argue that algorithmic quality also demands adherence to constitutional principles and legal requirements, at least for tools intended to inform public authority decisions. In these application domains, constitutional principles, administrative doctrines, and human rights norms are a crucial part of the relevant social technical context and must therefore inform and constrain the exercise of decision-making authority by criminal justice officials, which means that algorithmic justice is also concerned about how decision-making by the authorities will be based on algorithms and algorithmic decision-making. Her point is much more than that.

Her point is that whatever we think of as technical design choices when building algorithms and making tools for criminal justice creates constitutional dangers. So, in other words, what you think of as a technical choice and a technical solution, built by the experts, algorithm developers, and others can have larger implications on constitutional rights. However, the fundamental problem is more acute than this because we cannot simply address algorithmic justice as a technical problem. The moment we acknowledge the root cause of this and say, "Oh, the problem is that those who design those technical systems were not aware of it," and then they made some technical choices that are not in the right spirit and not in the right sense, these could be rectified. When we can move towards a technically perfect solution that also enforces constitutional rights, that is one

way to look at it. How feasible this is, again, does not have a very general answer because it depends on the context, the application, and the target group. So, the potential for such a remedy is still there, but the point we need to move on to is that the reason we are talking about this particularly in the context of the criminal justice system is a very simple one.

The screenshot shows a video player interface for a lecture titled "Algorithmic Justice - Part I". The slide content includes the NPTEL logo, the title "Technical Solutions, Individuals and AJ", and two bullet points. The first bullet point states that while constitutions recognize individuals, algorithms do not. The second bullet point discusses how algorithmic representations fail to capture biological, psychological, and narrative properties of human identity, leading to a statistical individual. The speaker is identified as Jens Christian Bjerring and Jacob Busch. A small image of a hand cursor over a scale of justice is also visible. The video player controls at the bottom show a progress bar at 11:45 / 32:06.

Algorithmic Justice - Part I

NPTEL

Technical Solutions, Individuals and AJ

- While Constitutions recognize us as individuals algorithms do not.
- “Algorithmic representations give no special attention to biological, psychological, and narrative properties, and, as such, they fail to capture central aspects of our ordinary representation of human identity. Indeed, we have argued, algorithms make predictions about us by relying on properties that do not directly relate to nor reflect the individuals we understand and represent ourselves as being” Artificial intelligence and identity: the rise of the statistical individual - **Jens Christian Bjerring, Jacob Busch**

11:45 / 32:06

There are technical solutions, and then there are individuals like us. Jens Christian Bjerring and Jacob Busch make a very interesting point: the constitution does not recognize us as numerically derived persons or data, or as an agglomeration or a consolidation of some data picked up from here and there; in other words, my body is not a combination or a bundle of all the tissues, bones, and everything; it is much more than that. So, when we talk about bodily integrity, when we talk about a body in the context of health, we refer to the whole. We are not talking about a bundle; we are not talking about a consolidation like that. So constitutional rights recognize that individuals are not merely a representation of XYZ data put in XYZ medium. Algorithmic representation, on the other hand, gives no special attention to our biological, psychological, and narrative properties, and as such, it fails to capture the central aspects of our ordinary representation as human identity. Put in other words: Can my identity be reduced to some set of data, which is then certified or analysed by certain parameters and pushed further, or can I be reduced to a tiny speck in big data? Or can my identity be defined and understood as something that is totally based on data? No, we cannot. Indeed, we have argued that algorithms make predictions about us, relying on properties that do not directly lead to or reflect the individuals we understand and represent ourselves as being. So, when I stand in a court as a lawyer, or as a witness, or in any other capacity I do not represent myself in terms of my data. I do not represent myself in terms of my bank account number, my Aadhar number, or anything else. I represent myself; I stand before the court, any authority, or in any legal sense as a person, as an individual. Whereas the problem with the algorithm solutions and most AI systems is that they reduce all this to data analytics, probabilities, conjectures, calculations, and then arrive at a decision, I am not merely a combination of certain parameters and data to be decided by an algorithm.

So, if I can consider myself an individual and not a statistical individual, then I will also look at what exactly the technical solutions are looking at in me. I look at algorithms as a way to do certain things, but algorithms look at me not as an individual; they look at me as a statistical individual. So, when they look at me as a statistical individual, the idea that I am a human being, that I do have some rights; in other words, we don't represent ourselves as statistical information but as persons.

Algorithmic Justice - Part I



Old Biases, New Modes?

- While human biases are too well known and have been dealt with for ages and with mechanisms that curb the arbitrary exercise of power based on biases and prejudices. But with algorithms and Machine Learning "Human bias gets re-packaged in complex layers of mathematical code and computations, and gains the facade of objectivity, rendering it particularly difficult (not less so) to spot and challenge – especially in the case of complex, black-box algorithms"
- Opacity and technical complexity add to the information asymmetries. Further a model that performs well on training data and testing data could make many mistakes in dealing with real life circumstances when they are different from what it was trained on. Moreover over reliance on technical black box models reduces the chances for meaningful human involvement. AI algorithmic oversight: new frontiers in Regulation *Madalina Busuioc*



14:53 / 32:06

In other words, we can also say that nobody can be reduced to mere statistical data or statistical inferences like that. A person may be sick, but that doesn't mean you should reduce that person to some health parameters and declare that the person is equivalent to these health parameters and health data. We can't do that. But let us also understand one more thing. Some of the biases we are talking about or accusing algorithmic systems of are not new. They were there earlier. Humans are prone to having some biases, and some of them are well known. So, we have found some ways and means to address them with mechanisms to curb the arbitrary exercise of power based on bias and prejudice. That is why the legal system has an appeal process, due process, and the procedures to establish certain things where the court or the judicial process itself will tell you what to do and what not to do. But the processes are done in such a way that equality is built in, biases are attempted to be eliminated, or more importantly, they cannot be an arbitrary exercise of power. The problem here is that biases can be contained; biases can be tackled by making transparent, systematic, open systems where people would know exactly what they are being represented as. So that bias angle can be tackled, and by denying excessive authority or arbitrary exercise of power to anyone, including the judicial system, we can try to curtail, if not fully eliminate, the biases. So, we know how to deal with the "old biases," but we don't know how to deal with the "new biases" arising from algorithms. So, the problem here is that machine learning algorithms and human bias, which is already present, are put in a different shade; it is also added to that, and then it gets layered into the mathematical code and computation. But it is often considered objective, as it is not based on bias, and then it is very difficult to really spot and challenge, in the sense that unless you know how the algorithms work, unless you have the expertise to

understand the inbuilt algorithm, which could be explicit or the implicit algorithm, where the bias is not implicit but is shown in the outputs. This is a question of opacity and technical complexity. Opacity is something equivalent to what we have already discussed regarding black boxes. Technical complexity arises as layers of code are added and algorithms perform multiple iterations, particularly when processing data from different sources for various purposes and objectives, resulting in increased technical complexity of the system.

Now you need to take into account the information asymmetries. So, the information asymmetries are something like this: they occur particularly when a software solution or a software system knows more about me or has a better idea about me than I do in certain circumstances. So, the information asymmetry between the algorithmic AI system, which has enormous data on me or anyone, versus me, who doesn't even know or won't even know that so much data is available for an algorithm to process and then assess me. So, this is an information asymmetry. So, we are seeing three factors here: opacity, technical complexity, and information asymmetry.

So, when these things combine with algorithmic decision-making, we need to be really worried. Further, an important fact is that models are not tested on human beings; models are tested on data, models learn from data, models perform analysis, and then come up with the output. They may make mistakes, they may get corrected, better algorithms can be built, and better data can be fed. And then the whole process can be iterated until the quality control norms are met or the potential for error is also reduced. But the reality is a little different. When you test a model while running it, it is not something that is equivalent to real-life circumstances. So, the real-life circumstances where they are applied, where people are present in the flesh, where the living persons are, are totally different. Although you can say that everyone is a criminal and every criminal is a criminal, it is not so. Each criminal cannot be reduced to a bundle of "oh, they are all criminals" and then labelled as if everybody in this group is a criminal.

Law does not look like this. Law does not classify people in such a way that they are labelled as criminals. It gives an opportunity for an individual to defend himself or herself. But the problem here is that when we try to overlay these technical black boxes, we are also trying to reduce meaningful human involvement and intervention, and this particularly happens when the models that are tested on test data are made to deal with real-life circumstances where the people with whom they are dealing don't even know that the systems are dealing with them in a way that is opaque to them. So, the irony is that the system is something like a black box or it is something like a camera that can always watch you or that can always detect you, but you will be clueless about how the system works.

Algorithmic Justice - Part-I



Criminal Justice – Tools and Errors



- **Prof. Karen** points out that it is presumed that one is innocent till proved guilty in Criminal Justice System
 - Falsely convicting the innocent is an error (Type I error)
 - This is more serious than letting the guilty go (Type II error)
 - Public and government officials must be accountable
- Right to Due Process
 - Can increase the risk of Type II error but
 - Can minimize Type I error
 - But often developers of ADM tools do not take this into account
- Implications
 - The tools cannot be substitutes for due process based decision making
 - They can impact depending upon the context of use and decisions given




20:46 / 32:06

Professor Karen again points out that you are presumed innocent until you are proved guilty in criminal justice; in other words, unless and until a person is finally convicted after exhausting all the appeals, that person will not be considered a criminal. So, until the final verdict comes, you are claimed to be or you are deemed to be innocent. But then the criminal justice system, when it is using algorithmic decision-making, can experience type 1 errors. In that sense, it can falsely convict innocent people. The type 2 error is when it let the guilty go free.

So, there are two possibilities. One, an innocent person may be wrongfully convicted. A person who is really guilty can be let go, freed, or acquitted. Now the problem here is that Type 2 errors are less severe in the sense that we are letting the guilty go. But punishing the innocent is a much more serious mistake. So, a type 1 error is much more serious than a type 2 error, although we can find ways and means to ensure that a type 2 error doesn't reoccur.

So, those who do this should they be held accountable? Yes, they must be accountable. This means that there should be a right to due process, and then due process should be applied. Now the problem here is that the risk of type 2 error can undermine typical due process because often courts acquit a criminal based on a lack of evidence or not being fully satisfied by the court, although in reality that person would have been the one who is guilty. But it can minimize type 1 error, although it increases the risk of type 2 error.

So that's a trade-off. Still, we go by the criminal justice system because it is better not to punish an innocent person than to do it the other way around, and an innocent person should not be punished under any circumstances, even if there is a possibility that a guilty person can get acquitted. The problem that she and many others have pointed out is that the algorithmic decision-making tools do not take into account this type 1 error versus type 2 error, punishing the innocent versus letting the guilty go in the criminal justice system. So the implications are that these systems cannot be a substitute for due process-based decision-making because the idea is not at all there, and then depending on the context of use and the decision, their impact will vary in the sense that if used on a mass scale, it is possible that a lot of innocents will get punished when we do ADM, where

these problems exist but are not being taken into account. So, they could exaggerate more of a type 1 error, and then they may not give that much of a type 2 error. So, we should improve the criminal justice system through its adherence to due process and by opening up many options for anyone to prove he or she is innocent, ensuring that innocent people are not convicted at all.

The screenshot shows a video player interface for an NPTEL lecture. The title is "Algorithmic Decision Making and Context". The NPTEL logo is in the top left, and the University of California, San Diego logo is in the top right. The main content is a list of bullet points:

- Thoughtless application of ADM can result in
 - It being used in contexts where it is not the right solution
 - Rights being violated directly or indirectly
- Recommender Vs. Judge
 - Recommending to me what I should watch or buy
 - Is not the same as Judging me by ADM
- Should ADM have a place in Criminal Justice System
 - 'Efficiency' Vs. rights and transparency

There is a small inset image of a laptop with a scale of justice on it. At the bottom right, there is a small video window showing a man speaking. The video player controls at the bottom show a progress bar at 24:08 / 52:06.

But here comes the catch. When we thoughtlessly apply ADM in all these things, whether it is the right solution or not, rights will be violated directly or indirectly. It can happen in facial recognition technology, it can happen in public service, it can happen in the criminal justice system, it can happen in bail, and it can happen for those who are in jail. But we cannot simply equate that and say that all ADMs are equally bad or all ADM applications are equally terrible. No, see, when Netflix or Amazon or any recommender system tells me to watch something, buy something, or consume something, it does weigh upon my browsing data, my visits to various websites, my previous purchases, or YouTube for all I have seen on YouTube, and then other things. Since my browsing data is being collected, my purchase data is being collected, and then the recommender system comes up with some options for you: "Okay, buy this, have a look at it.

Since you have bought this, you can also try this, maybe recommend it." Recommending by Netflix, Amazon, or Google is not the same as someone judging through algorithmic decision-making. So, the former, however problematic we may feel or however irritating it could be, is not the same as the latter because the latter has a much more severe potential for harm than anything else. So, the question here is whether algorithmic decision-making should have a place in the criminal justice system at all, or should we say that efficiency versus rights and transparency is a trade-off that we have to live with because algorithmic decision-making systems, despite all the problems, are better solutions in the criminal justice system. So, who would decide what is an efficient solution, the right solution, and then what the level of efficiency is? Can we say that an ADM which has an efficiency level of 70% is good enough to be applied across the entire criminal justice system? Or will we say that we need a system that is more humane, or a system where the innocent will not get punished, or if they are punished, they will have

an opportunity to prove themselves by going to a court of law and then trying to produce evidence to that effect? So decision-making algorithms, also in context, purpose, and who is impacted, make a huge difference.

Algorithmic Justice - Part-I



Fallacies and Grounded Truth



- "Algorithms don't harm people fallacy"
- (Gun's don't harm people, people harm people..)
- Even if a tool does not, in and of itself, interfere with or engage human rights, it does not follow that we can ignore other kinds of harm: like
 - Troubling tendency of 'digital enchantment' in contemporary policy discussions,

to ignore that ADM systems are powerful technologies which are capable of producing dangerous decisions (even if tempered by a team of benevolent developers, or the 'perfect' exercise of discretion by a human decision-maker), because of the capacity to operate automatically, at scale, and to trigger action that is remote in both time and space from the location at which the action is triggered.

E.g. 'government by database' much more threatening to human rights and democracy than a policeman with a pen and notebook"



26:43 / 32:06

So according to Professor Karan, the idea that algorithms do not harm people is a fallacy. This is something like guns do not harm people; people harm people. For the agency, this argument is about people, not about the guns. But the problem is that there is a huge difference between guns and ADMs. That is very obvious. So even if the tool does not engage with human rights in itself, it does not follow that we can ignore other forms, like the troubling tendency of the digital enchantment of contemporary policy discussions. To ignore ADM systems or powerful technologies that are capable of producing dangerous decisions. So, the problem is not that you need better decision-making in the sense of better designs, benevolent developers, or perfect exercise of human discretion by human decision-makers because the problem is that they have the capacity to automatically scale and trigger action that is remote both in time and space from the location where the action is triggered. So, the ground truth or the ground reality is that no matter how much we say that ADMs are very efficient and that they could be designed by people who are very sensitive to such biases, we can try to eliminate as much as possible all the biases in the data and training capacity to come to the decision when it is ADM. So, her point is that we may think that government databases, the accumulation of databases, efficiency, linking databases, and the increasing use of data in decision-making, particularly in criminal decision-making, is a good thing and that it will enhance efficiency. That is the normal tendency to argue. But in her view, government with databases is much more threatening to human rights and democracy than a policeman with a pen and a notebook. So don't underestimate the power of algorithms. Don't underestimate the scope for misuse when algorithms use databases. So, her argument is that we need to pay attention to the policies, and we cannot take the stand that guns per se are okay, but the problem is people. Here, her argument is exactly the other way around. Guns, per se, are problematic.

Algorithmic Justice - Part-I



Algorithmic Accountability



- Scholars like **Prof. Karen** have argued that ADMs applications in Criminal Justice System has serious consequences.
- How to counter that and what are possible interventions
- She suggests that solutions through Human Rights Law.
- Administrative Law including use of Judicial Review Principle, and, Anti discrimination law





29:02 / 32:06

One of the ways these things have been discussed in the literature is through algorithmic accountability. So, as persons like Karen have suggested, they have also proposed some measures, such as how we can counter that, solutions through human rights law, administrative law, the use of the judicial review principle, and the anti-discrimination principle. Some people have gone ahead and said there should be data protection laws, and some of the other legal principles can be used. But the idea here is that algorithmic accountability should be brought in, should be practiced, and more importantly, algorithmic accountability is not something that can be reduced to mere numbers. In the sense that we cannot say an algorithmic system that is 90% error-free, which has errors only in 10%, is good because 90% error-free is perfect. Because we need to pay attention to where the 10% margin of error occurs, whom it will impact, and then how it will impact. So, the algorithmic accountability is something that needs to look beyond the number-crunching systems and the numbers themselves.

Algorithmic Justice - Part-I

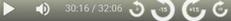


Algorithmic Accountability

Can lawyers working with developers of ADM 'fix' some of these

- Should there be institutional mechanisms that build safeguards when ADM is used in Criminal Justice System, Employment, Public Welfare
- Is it possible to use ADM in a positive way and for social good
- Given the widespread use of ADM in public sector and by private sector who will assess them and monitor their impacts

In the next class we will discuss these



One of the ways people in the literature have been working on this is that, you all talk about all the problems in the ADM; you all talk about the algorithms deciding with bias and discrimination, and then you all talk about these systems as dangerous monsters. So, can't we make things better? One of the arguments is that by having lawyers work with developers of ADM, some of these can be addressed. In the sense that such systems are developed for specific applications in criminal justice, employment, and public welfare. There should be institutional mechanisms that build safeguards and guardrails into such systems. In the sense that the institutions that commission these ADMs or that procure these ADMs should have some idea of how they can bring in safeguards and guardrails in the systems themselves, particularly when they are applied in critical systems like criminal justice, employment, and public welfare. Is it possible to think of algorithmic decision-making in a positive way and for social good? And then, more importantly, given the widespread use of algorithmic decision-making in the public sector and private sector, who is going to assess them? Who is going to control them? Who is going to take a look at them and then decide? We have had enough, as the numbers are becoming too much to handle, so let us put an end to it. We will limit ADM only to some sectors, and in the next class, we will discuss some of those. We will also look into some of the recent debates on ADM, as well as some of the solutions. Thank you.