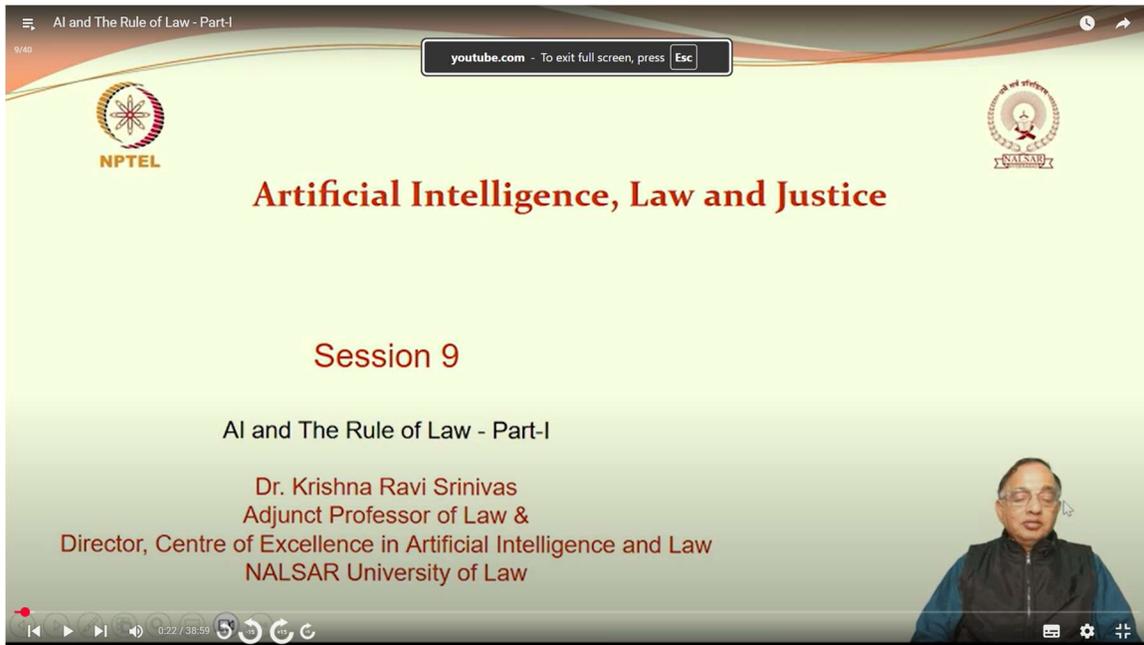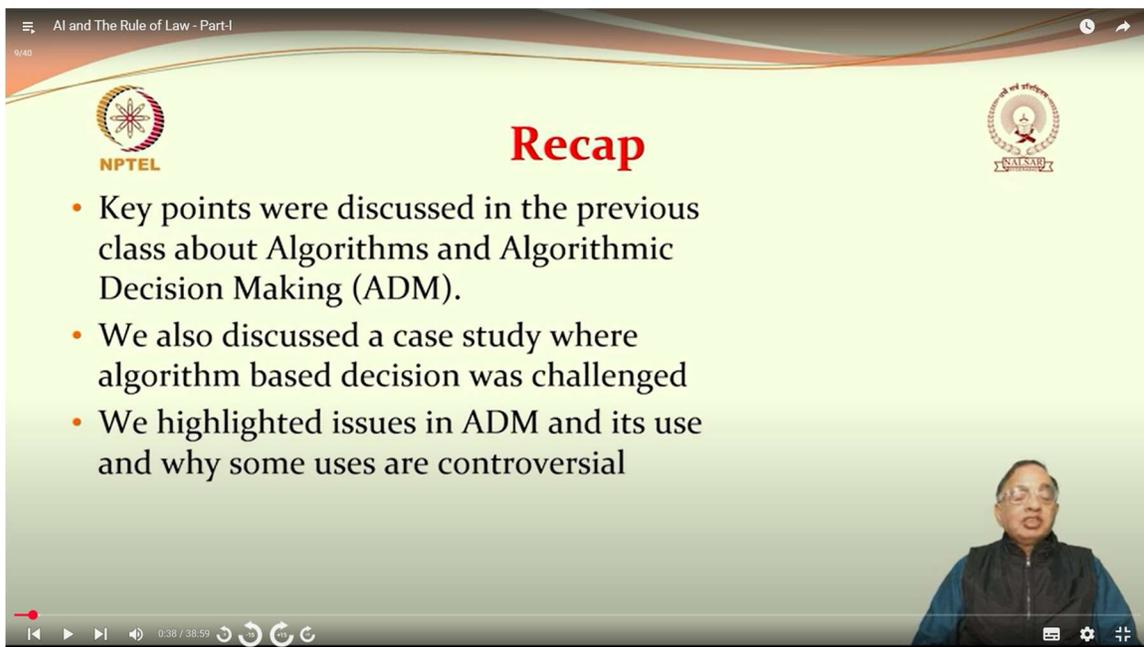**Course Name – Artificial Intelligence, Law and Justice**
Professor Name – Dr. Krishna Ravi Srinivas
Department Name – Center of Excellence in Artificial Intelligence and Law
Institute Name – NALSAR University of Law
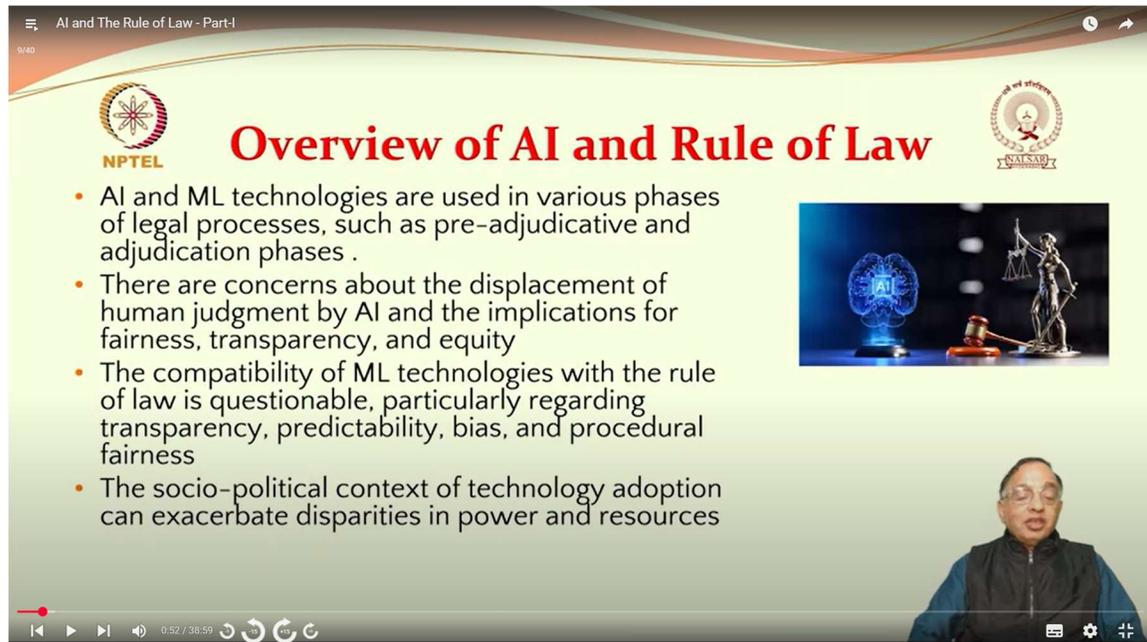**Week – 02**
Lecture – 09

AI and The Rule of Law - Part-I. Artificial Intelligence and Rule of Law - Part I This is session 9. This is the first class of the three classes that we will have on this topic.

Let us recap the previous session. Key points about algorithms and algorithmic decision-making were discussed in the previous classes. We also discussed a case study where algorithm-based decision-making was challenged. And then we highlighted issues in algorithmic decision-making, its use, and why some uses are controversial.

In this class, we will start with an overview of AI and the rule of law. We have already discussed the rule of law in the previous classes, where we emphasized the importance of the rule of law in a democratic society and its constitutional basis. So, in these three sessions, we will look in depth at how AI will and can impact the rule of law and then how people who are safeguarding or who want to support the rule of law, or who are considering that the rule of law should be the cardinal principle to govern society as well as the legal institutions, are responding to that. So, we need to start with some fundamental ideas about why so much debate is happening on the role of AI in the rule of law. Basically, artificial intelligence and ML technologies are used in various phases of the legal process. We have seen many cases and instances and provided numerous examples of AI being used in different legal processes, including the algorithmic decision-making process, as well as the increasing role of ML, NLP, and AI-related tools in law, particularly in the judicial system. So, when this happens, AI is becoming increasingly entangled in the pre-adjudicative and adjudicative spaces in the law. And so, there are a couple of concerns about the displacement of human judgment by AI. And when we talk about displacement, we are not referring to the immediate takeover of human judgment by AI, with AI replacing human judges.

We are not talking about that. We are saying that AI is trying to displace human judgment instead of simply taking it over. In the sense that some sort of human judgment

in some areas, in some sectors of law is either being displaced by human judgments or there is a good potential for AI to do that. So how does this impact the rule of law, particularly the principles of equality before the law, transparency, and equity? Moreover, there is one view that ML technologies are not 100% compatible with the rule of law for the simple reason that there are many doubts and questions about their transparency, predictability, bias, and procedural fairness. So, if we measure the ML technologies and the processes in which they are inscribed through the lens of the rule of law, then we may find that this lens will be able to tell us or does tell us that there are critical issues in this.

And then ML technologies and the processes in which they are imbibed may not fully meet the standards of transparency, predictability, non-bias, and procedural fairness. When we say procedural fairness, we mean something like the procedure should be open, it should be transparent, and, more importantly, it should be fair. It should be fair in the sense that a person who challenges the procedure and is affected should have the facility or the option to challenge their decision and findings and to go for an appeal. The procedural fairness also includes that equal opportunity should be given for a person to defend himself or herself, and more importantly, nobody can be asked by default to implicate themselves, whether they are silent or not. So procedural fairness is a legal concept that is very critical when we talk about any due process in law.

It is all the more important in light of the constitutional principles that are present, such as equality before the law; no one should be deprived of any property, and no one should be punished on any grounds unless there is a court of law that has made the decision, followed by a procedure that is fair, equitable, and meets the criteria for justice. So, when we look at these parameters or when we look at these standards, to what extent AI-based decision-making, to what extent ML-based decisions, and the systems that are run on AI systems based on these will fulfil this criterion is a big question. So, we also need to look into the socio-political context of the technology adoption. And when we say that we have discussed the digital divide, we also know that technology per se may appear to be neutral, but when it is embedded in sociotechnical systems, whether it is the huge infrastructure projects or any other system, there are and there could be disparities in power and resources, and AI and ML have the potential to enhance the disparities rather than reduce them. So, these are some of the concerns about AI and the rule of law.

Overview of AI and Rule of Law

- **Pre-Adjudicative Phase**
  - ML and AI used to select targets for tax and regulatory investigations
- **Adjudication Phase**
  - ML and AI guide determinations of individual violence risk during pretrial bail
- **Human Judgment vs. Code-Driven Counterparts**
  - Predictions of displacement of human judgment by AI
  - Resistance due to concerns about fairness, transparency, and equity
- **Normative Concerns**
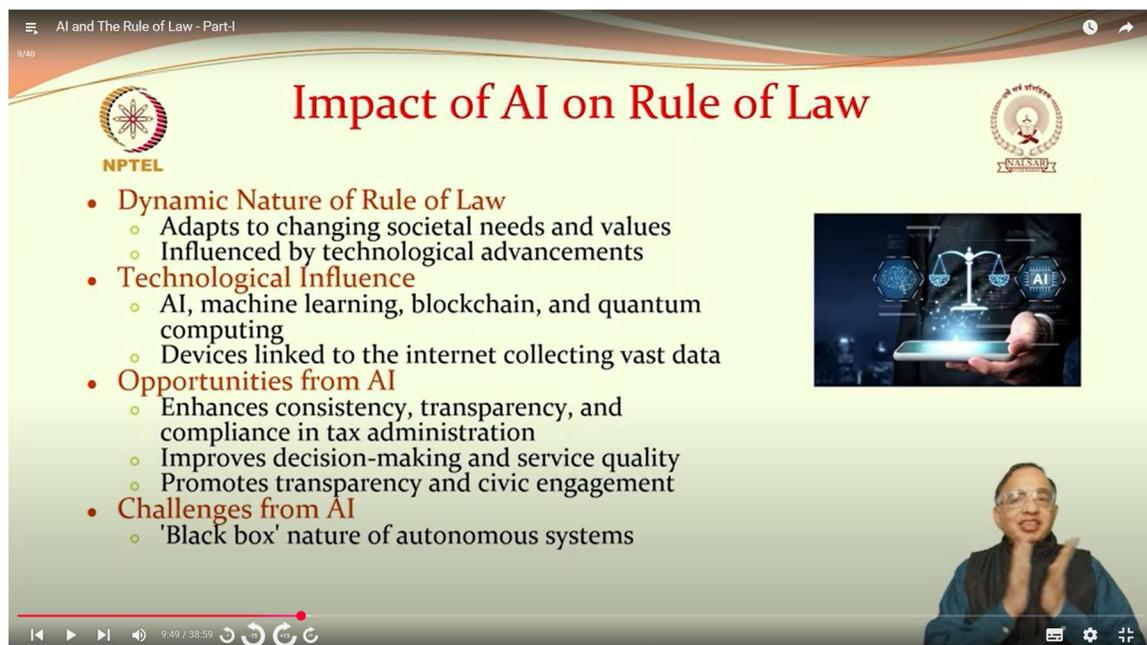 - Criticisms often overlap with rule of law concerns

In the pre-adjudicative phase, typically machine language systems and AI are used to select targets for tax and regulatory investigations in the sense that tax administration and various regular administrative processes, like commercial administration, use AI and ML for a lot of purposes, including identifying potential tax violators, identifying people who should be in the tax net but are not, and also identifying, based on the previous year's filings and records, how some people might have responded by disclosing their income in part but not in full or how some people could have evaded the tax system by pretending that they are not entitled to be taxed, although they may or should very well be subject to the tax system. To do that, the tax authorities need a whole lot of access to a lot of data, which also includes the previous tax returns, various other financial transactions including bank transactions, investments, purchases, and dealings in the stock market, and all. So, when authorities conduct a data-intensive search analysis and then find out that some people have escaped the tax net or have not filed their returns in an adequate way, we see that they are using ML and AI in the pre-adjudicative phase. This is more like a preliminary investigation. But in criminal law, the determination of individual violent risk during pre-trial bail is also a part of it.

We will see some examples of this in this session or in the subsequent sessions. So, when a person applies for bail in a pre-trial, particularly in a criminal case, the court will consider the potential of that person to commit a violent crime or whether the person should be granted bail given his antecedents and precedents, along with other factors. So, in the adjudication phase itself, AI and ML tools have been used either to assess a person's eligibility and potential to be granted bail or to deny bail. And then the problem here comes from human judgments when human judges look at things versus the court-driven counterparts; there is a wide variety of perception. One way you can argue that

humans have inbuilt biases is that they have a lot of biases, whether explicitly stated or implicitly taken.

But systems and machines are neutral. They cannot differentiate between persons on the basis of gender or other factors. They look at facts, they look at bare facts, they look at data, they look at some parameters, and based on the training and the parameters they have imbibed, they make the decisions. So human judgment could be argued to be inherently biased, or its potential to be non-biased is much less if we compare it with the way machines perceive and then make judgments or come up with orders. So, when human judgment is being displaced by AI, there are real concerns about fairness, transparency, and equity, and are we replacing the human sense of being fair, being transparent, and being equitable with an opaque, machine language-driven sense of the way the AI systems make decisions? So, these are some of the concerns that we will repeatedly see in the course in different contexts, in different circumstances, with some case laws.

And then there are also normative concerns because criticisms often overlap with rule of law concerns. Because the rule of law concerns that we saw are also normative concerns. Because the rule of law ultimately is also a question of normative principles and then abiding by them. So, these are some of the real concerns that are being expressed time and again about the use of AI, particularly in decision-making applications.

But let us also understand one thing: the rule of law, as we emphasized earlier, is not a fixed doctrine; it is an evolving doctrine, and there are multiple interpretations. In fact, even the core components of the rule of law are not the same across countries or across constitutional courts. So, the rule of law, if we take some of them as core principles and

then some as binding principles, we cannot argue that it will remain the same. It will always remain the same, but the way we interpret it will also change. And then the rule of law has to adapt itself to societal needs and values, and then society itself undergoes change in the sense that from kings, queens, and monarchs to a constitutional scheme of things, we have traversed a long distance in terms of ideas, practices, and more importantly, in terms of normative values and principles that guide us. So, the idea of the rule of law itself will undergo change.

In a technology-driven society, what changes should the rule of law undergo, or should we simply say that the rule of law is something like a basic structure? It cannot be diminished, it cannot be abrogated either in part or in full, and it cannot be negated, nor can it be simply thrown away, as it is something that will always be the cornerstone and the foundation. So one can take this view irrespective of technological advancements and the changing needs of society and its values, the principles of the rule of law should remain the same, because if they themselves are changing, then the rule of law cannot become something that is too mutable; it becomes something that we will never be able to identify with the core concepts of the rule of law, which have evolved over a period of centuries. And then comes the question of what sort of influence technologies will have. Seeing technology's impact on law is not something new. Law has been facing technology since time immemorial.

But the question here is what sort of impact digitisation technologies, particularly technologies like AI, machine learning, and algorithmic decision-making, will bring. The question here is that the devices related to AI, the internet, communication, media, and digitisation are also linked with data collection, data classification, data sharing, monitoring, surveillance, and, more importantly, they enable diffused control and diffused surveillance. In the sense that when you see CCTVs across cities and thousands of places, you know that CCTV cameras capture images; they are live; they are there 24/7. So, they also collect a vast amount of data, as we saw from the data governance perspective. Questions also arise about what happens to this data, who has access to it, who deals with it, who stores it, and who will have access to it. So, when these devices that are part of the AI ecosystem collect a lot of data, share that, and then when everything is fed into a whole lot of big data, what exactly are we talking about? We are talking about data becoming more mutable, data becoming part and parcel of the AI-based ecosystem, which again is, in one sense, posing a great threat to the rule of law according to some; but for some other people, some other experts, the rule of law is such a concept that it can adjust, it can reinvent itself in the context of technological changes without substantially compromising its core principles and normative values. If we take a very positive perspective, we can say that AI brings consistency; it is transparent and compliant in tax administration in the sense that applying AI tools in tax administration enhances efficiency and makes the system more transparent. It is also a system that

cannot be biased on account of human biases in the sense that it goes simply by data, simply by what is available, and then simply goes by record. It does come up with judgments; it does assess, but that is unbiased and based purely on data without even knowing whether the person is male or female or whether that person's skin colour is this or that, or whether that person is of this height or that height. So those things do not matter when AI systems are working.

More importantly, AI systems can play a great role in decision-making because they are quicker; they are not only faster and more efficient, but they also take care of many mundane tasks that a person would normally engage in. As we saw in legal research, as we are seeing in various other options or tasks, AI can take over and do many things that we normally do with tremendous effort, wasting a lot of time and effort on mundane tasks. And more importantly, an AI-based system can be a transparent one that can engage with the public with more confidence in the sense that the public can engage with chatbots or AI-based systems in the front end by making them more transparent. People will understand that they are not dealing with black boxes. They are dealing with systems that will tell them exactly what they need to know and will also guide them.

By creating user-friendly interfaces for AI-based systems in public administration, including chatbots, virtual assistants, and other tools, they can promote transparency and inspire more confidence than if done the other way around. In the sense that if you have a customer service section where people are not simply inaccessible because they receive countless calls a day or they're continuously being called for one purpose or another from different consumers. So, these are some of the opportunities and plus points, but the thing that haunts, or that is going to haunt, any discussion on the impact of AI not just on the rule of law but on various other things is the very black box nature of autonomous systems, particularly AI ones. So, whether you talk about dynamic nature or the rule of law getting adapted to technologies like AI, ultimately the question that will pop up is: What do you do about the black box nature of AI? How do you deal with that? Can we wish it away and still claim that AI is transparent, not opaque, free from bias, and capable of inspiring public confidence? Very doubtful.

When machine language is used for enforcement and adjudication, the considerations are very different in the sense that the objectives to be met by ML vary in different contexts and applications. When they improve performance with training experience and derive rules from data, it also requires humans to train them again and again, but more importantly, it helps the people on the administrative side understand that systems taken care of by AI are really the ones that mimic what humans would do. So, if they are going to produce results with lesser bias and more accuracy than the traditional regression tools, they are much better. Or, in other words, they should be the ones who should be recommended for any public service use. But the question is that the technological shift, as it creeps in slowly and steadily or as it simply leapfrogs, will create a lot of problems for the rule of law because we are not talking about AI as a technology; we are also talking about AI as a sociotechnical system that interacts with societal concerns and the larger society and economy.

A great feature of the ML tools is that they can interact with different forms of the rule of law, whether they are formal, substantive, or procedural, in various ways depending on the implementation. For example, an ML-based implementation can be based on very formal laws. It can also have a very procedural approach by using algorithms and other methodologies, and it can be very substantive as well since the data tools rely on training data gauge various relationships, and then they can come up with a lot of things on their own. And, more importantly, supervised ones need predefined categories. Unsupervised ones, which can learn on their own, do a lot of classification based on the inherent structure of the data. In other words, humans need not bother about doing so many things.

Leave the task to AI, which will take care of many things that otherwise humans would have to do. In the sense that structuring the data classification and then understanding the data on its own, AI tools can do much better than humans. That is why, when it comes to data, AI is unparalleled in data sorting, management, and making the best use of data.

On the other hand, whether ML tools are beneficial when they interact with the formal rule of law depends upon the applications and implementation. According to Fuller, a rule of law in the AI context, particularly when the question of algorithms and algorithmic decision-making arises, there is a need for a published code for future disputes because the published code will inform people and provide them with an understanding of how the dispute is to be resolved, how the dispute is being addressed, or how AI views the problem and the solution.
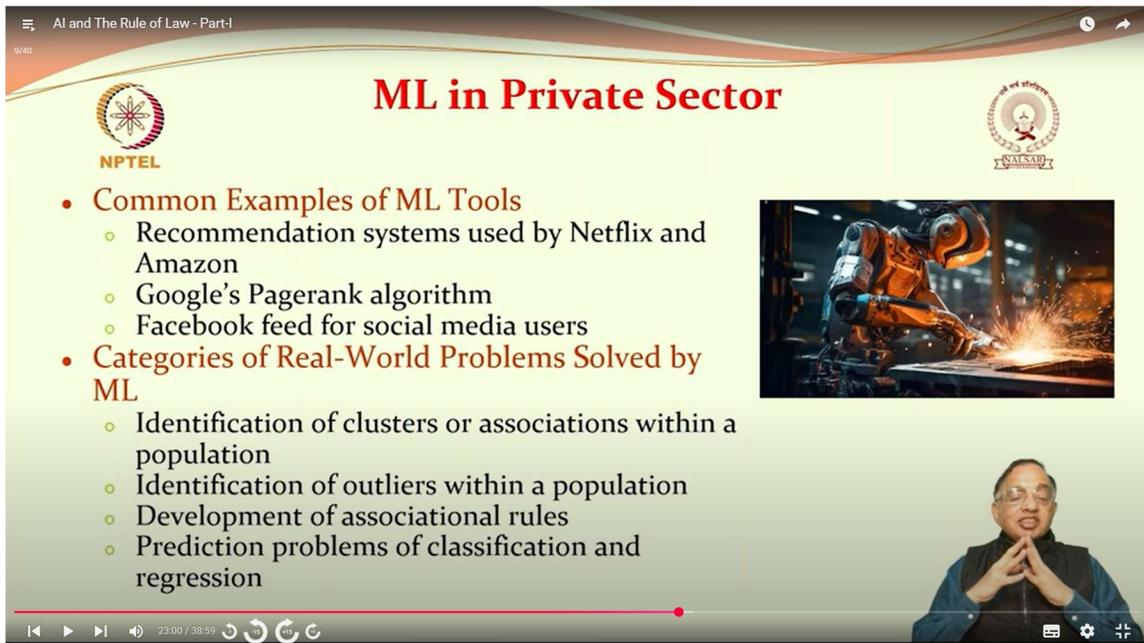
On the other hand, code is something that is written in a way that normal people can't even make sense of. But it can be made understandable to the ordinary citizens by trained lawyers in the sense that the code can be one that is explainable in the sense that you put the computer language on one side and then explain the steps on the other side. Say that if the left hand is the computer code, on the right-hand side you can explain it. This is the input value being taken by the system; then, if the next step is the system assessing the value, it takes the next step. Like that, it can be split into different steps one by one in normal, understandable text language, and then people can make sense of that.

But Fuller also says that there could be some understanding in the sense that even if trained people could not understand it, there could be some abstract sense that should be allowed. In the sense that we need not say that even a trained person should be able to understand what the system does 100%. And computer scientists themselves may not be able to fully help us understand it. So, he says that it is fine. In the sense that some sort of margin for that could be given.

That is Fuller's understanding. But the problem is not the people who have designed it, not the people who want to understand it. The problem is looking at the systems of the

people who are going to be impacted or the stakeholders who are going to be impacted. If you look at the systems from their perspective, and if they are obscured by their classification, then we should be very clear that they need to be categorically informed about the potential harm they could face. And more importantly, when a system learns on its own, if there are going to be reward functions on predicted tools, these are not normally available for examination. So, when the full code is not made available or what is made available only tells us part of the story, then ML systems have to be objected to because you may feel that the system is fair, it is transparent, it is not opaque, and it treats everyone fairly, and then the steps it takes are good.

In the sense that they are understandable, they are nothing that needs to be questioned. But if there are some reward functions that are built-in but not made visible, then that system is not fully transparent. It is hiding something while showing something else. So, this is something we need to look into. And then there are any number of challenges in understanding ML system tools for the simple reason that it's code.

To give examples of ML systems in the private sector, the recommended systems used by Netflix and Amazon are based on ML tools. Similarly, the Google search algorithm, which takes the page rank into account, is based on algorithms. And Facebook also sends across lots of feeds for social media users, and that again is based on machine learning algorithm training. So, these are some of the things where we know that ML tools are in vogue or being put to use in a very active way on a day-to-day basis. But machine learning systems can solve a lot of problems. For example, if we feed the data, they can identify clusters or associations within a population. Machine learning systems can quickly identify clusters in terms of geographic clusters, income classification, or any

other criteria, and associations in the sense of how they are linked. It can associate gender with some other parameters. It can associate your location with some other parameters. It can associate your income with some other parameters and then come up with a map associating these things.

And more importantly, you throw data to the ML system and then ask, "What are the deviants here?" Tell me how many individuals in this population are adults having less than 5 fields, and tell me how many in this population are suffering from a rare disease that we know occurs in one out of 10,000 or one in a lakh. Or it could be an outlier. An outlier could be anyone. An outlier could be a super-rich person. An outlier could be someone who is much below the poverty line.

So, the ML systems can do these identifications effortlessly and easily. More importantly, they can come up with associated tools. They can come up with associational tools. They can also predict classification and regression in the sense that they can perform classification and regression analysis, meaning that having an ML tool is as good as having a statistician at your service. So, when machine learning tools are being used in the private sector for many purposes, they are also using their traditional statistical methods of regression analysis and other ways of looking at things like mean, median, average, etc. So, this is something which we need to be aware of when we talk about the application of ML in different day-to-day practices.

State Adoption of ML

NPTEL

- Governments use ML tools for investigation, targeting, and as substitutes for human judges .
- ML tools are also used in defense technologies, diagnostic tools, and work facilitation
- ML tools are used by various government bodies to analyze data and identify legal violations .
- Governments use ML for allocating resources and predicting crime location
- ML tools are used to predict pretrial violence and guide bail determinations .

And how do the states look at ML, and how do they adopt it? As we saw, they are often used for investigation, targeting, and as substitutes for human judges in the criminal system in some countries, but not in all countries. But they are also extensively used in different technologies for diagnostic tools. Today, medicine is transformed beyond

recognition by both AI and ML. And there are a lot of diagnostic tools that are based on ML and large language models.

And more importantly, AI is getting integrated into digital technologies, like how the data from your smartwatch can be captured, fed into, and then read by an ML tool, which can then provide an analysis of your health or health parameters if all the information fits into it. So, identifying legal violations, identifying people who have been doing well or who are law-abiding citizens, and then identifying outliers, identifying the normal population—everything is possible through ML tools. We want to know which ML tools can be used for allocating resources. For example, an ML tool can easily identify based on the population data how many people live in specific parts of the city or country, where most poor people reside, where most women with anaemic conditions live, and where people suffering from a specific disease are scattered or clustered across some cities or spread evenly across the country. And here, we can also do one more thing: if we tell the ML system, or machine learning system, "I am going to give 100 crores," for which I need to identify who the most eligible beneficiaries are based on the income criteria, other criteria, or criteria that we can choose.

For example, I can tell the machine language system that, look, I am going to allocate 100 crores for some areas that are not well developed. Now you pick from this data, from this geographical location and then, with geospatial mapping, identify which areas lack the infrastructure that is mostly needed but do not have it. An ML system can do such a job in a wonderful way without any bias, provided the data that is fed to it is right; it can then correlate that with geospatial data and really map whether the infrastructure here is either non-existent or weak, and then the 100 crores can be better spent here rather than in an area that already has a lot of infrastructure built. So, for such decision making, we can use ML.

This is fine. This is understandable. But when we use ML systems to predict crime, it becomes problematic. Particularly when police use ML systems for predictive policing, it becomes very controversial. What is predictive policing? Police or the criminal system, or whatever you say, the police have a lot of data available to them in terms of geographies, in terms of crimes, in terms of areas where crimes are more, where crimes are less, and then which part of the day, which week of the day, and then which period. So, by doing a statistical analysis, police can predict or come to the understanding that a lot of robberies take place in these areas.

In this area, a lot of pickpockets take place. In this area, a lot of other crimes, including violent crimes such as domestic violence, rape, and violence against women and children, take place. So, based on this data, they can predict that in this area, this month, this many murders have been committed. In this area, there have been many riots. They can easily predict, or at least try to identify, where the next problem will arise in terms of law and

order. So in the broader idea of law and order, predictive policing will help the police department identify the potential for harm, the need to take preventive action, and also the need to ensure that violent eruptions, riots, and unforeseen mass murders do not happen.

But using ML in this has been controversial for the simple reason that machine learning systems can be fed data that may have biases, be incomplete, or predict something that may not happen. In a sense, the prediction could be seen as a simple prediction. Where nothing else would happen, the police would have based their decision on that prediction and put a lot of police in a particular locality, anticipating some trouble, but nothing might come out. So, there are issues with it, but the problem here is that this location analysis often happens, and then they match that with the people whom they call "living in crime-infested areas." So predictive policing across the world has been a very problematic application for the simple reason that those who are affected are the ones who are already marginalised or who are already considered a deviant population. Similarly, we discussed pre-trial violence and bail.

Compas Algorithm

NPTEL

NALSAR

- Compas Algorithm Usage
  - Used in many American jurisdictions
  - Generates a risk score from one to ten for defendants
  - Guides judges' bail determinations

- Controversies and Criticisms
  - Accusations of racial bias
  - Higher proportion of factually innocent black defendants detained
 - Compared to factually innocent white defendants

31:19 / 38:59

Coming to that, COMPAS is a very interesting case that we will see later in a greater detail. It is based on algorithm usage. It is given to a judge to decide whether a person can be granted bail or not. So, it takes a whole lot of data: the person's previous criminal record, the person's age, the person's income, and what his/her past record has been in terms of other things, like whether he/she has been charged with a felony. Has he/she been on trial for some other cases? How many cases is he/she facing? And what is his/her overall record in terms of other parameters: income level, employment level? Whether he/she is involved in any family violence cases like that.

Then there is a ranking, and a score from 10 to 1. Based on the score, the judge can decide whether a bail should be given or not. Two, if bail has to be granted, what conditions should be imposed? And three, if the judge decides that bail will not be granted because the score is something that does not favour the granting of bail, the judge can simply say that, as per the analysis, as per the data fed, and as per the COMPAS algorithm that is being used for it, you are not entitled to bail. So, judges taking bail orders based on this rationale enhance the efficiency, but it does not deliver full justice, and it does not do justice to people who really need it. So, the COMPAS algorithm has been a very problematic one, which we shall see later. One fundamental problem in the COMPAS algorithm that repeatedly emerged was that it was based on racial bias.

Then, given the problems or accusations regarding the American criminal system, a higher proportion of factually innocent Black defendants were detained, and those who were detained were not criminals or potential criminals; they were innocent, but on account of bias and other factors, they were detained. And the algorithm that was used to train COMPAS, unfortunately, compared the factually innocent white defendants with the black ones, and then it was wrong. So, the algorithm came up with the wrong decision or the wrong finding for the simple reason that it compared two things which should not be compared at all. In the sense that you cannot compare factually innocent white defendants and factually innocent black defendants just like that and then say that, by default or by this logic, the potential for a black person to commit a crime is greater. So, those sorts of things were the problems with the algorithm.
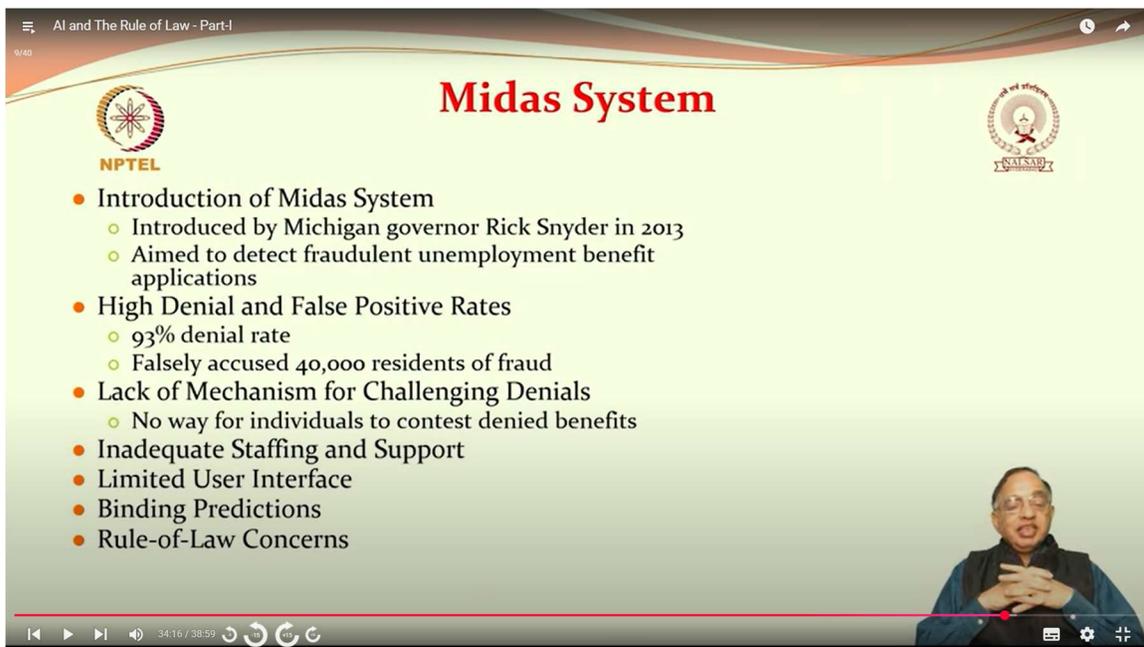
The next case is the Midas Case. This, again, is a case from the US introduced by the Michigan government in 2013. Here, the problem was that in the USA, under social

security, people can get unemployment benefits for a particular period, and then the government gives them support so that they can at least survive. However, when this is not done manually but through a system called MIDAS, there were huge problems. First of all, the denial rate was very high. It was more than 90%. Then, it falsely accused the people who were given this unemployment benefit of fraud, claiming that some 40,000 residents were involved. And then the system was built in such a way that if it is denied to you, there is no appeal procedure, there is no way you can challenge it and say that this is wrong; I am entitled to it because I am unemployed, and I should be able to get the benefit for the simple fact that I am unemployed. The problem with that came in so many ways that there was inadequate staffing, and then the predictions were binding in the sense that humans couldn't override or rule them. Or it was MIDAS, a system that was like a strait jacket. And then, when people are assessed through such a system, even if they are innocent, even if they are entitled, by some strange logic the system decides they will not be entitled to, they would not be given the benefit. So, the Midas system also became very controversial.
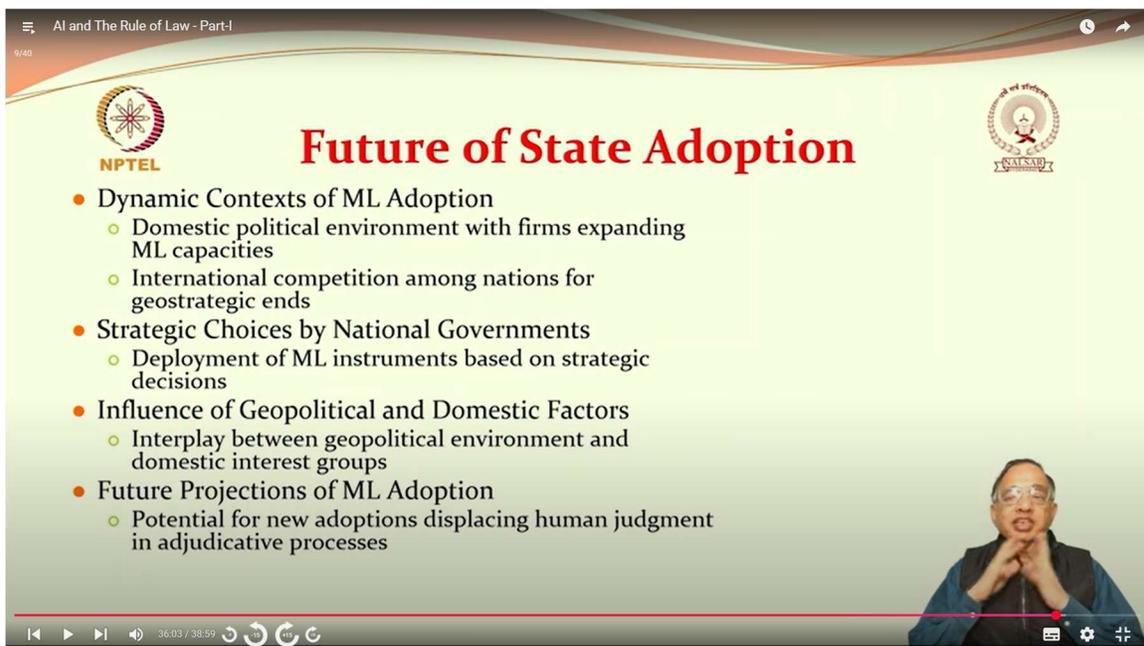
Future of State Adoption

- Dynamic Contexts of ML Adoption
  - Domestic political environment with firms expanding ML capacities
  - International competition among nations for geostrategic ends
- Strategic Choices by National Governments
  - Deployment of ML instruments based on strategic decisions
- Influence of Geopolitical and Domestic Factors
  - Interplay between geopolitical environment and domestic interest groups
- Future Projections of ML Adoption
  - Potential for new adoptions displacing human judgment in adjudicative processes

But what exactly happens when states go for it? States adopt machine language systems and AI systems for many reasons, including efficiency, enhancing quicker public service, and, more importantly, such systems often help the state do a lot of things without relying on a huge battalion or a large army of civil servants and employees. So, the dynamic context is that countries often compete with each other in terms of technological innovations, and now there is a global AI race happening with China, the USA, and the EU competing with each other, so when considering the dynamic context of ML adoption, we need to take that into account as well. And then, on many occasions, governments decide to go for AI-based systems based upon strategic decisions,

particularly in applications like defence and in highly sensitive sectors. So, the interplay between geopolitical and domestic factors also plays a major role in AI adoption as well as the use of ML systems. The problem here is that there is a huge potential for ML systems to be adopted in the future in areas where the adoption can really replace human judgment either partially or fully.

So, we have discussed a couple of things in this session as an introduction to a topic that we are going to explore in three sessions. In the next one, we will take this forward by discussing more about the intricate aspects of the rule of law and AI, particularly why the topic of the rule of law and AI has become very hot and relevant now. And then there are some theoretical perspectives on this topic that we will also look into. If we can recollect what we discussed earlier about the rule of law and then what we discussed in the first session, one thing should be obvious: that the rule of law is being emphasised time and again for the simple reason that it is the cornerstone, or it is the foundation, of any democratic society. It is the foundation for any society based on constitutional values. It is the foundation for any society where the principle of equality reigns. More importantly, nobody can assert himself or herself as above the law in any society and then say that everything else depends upon my whims and fancies. Thank you.