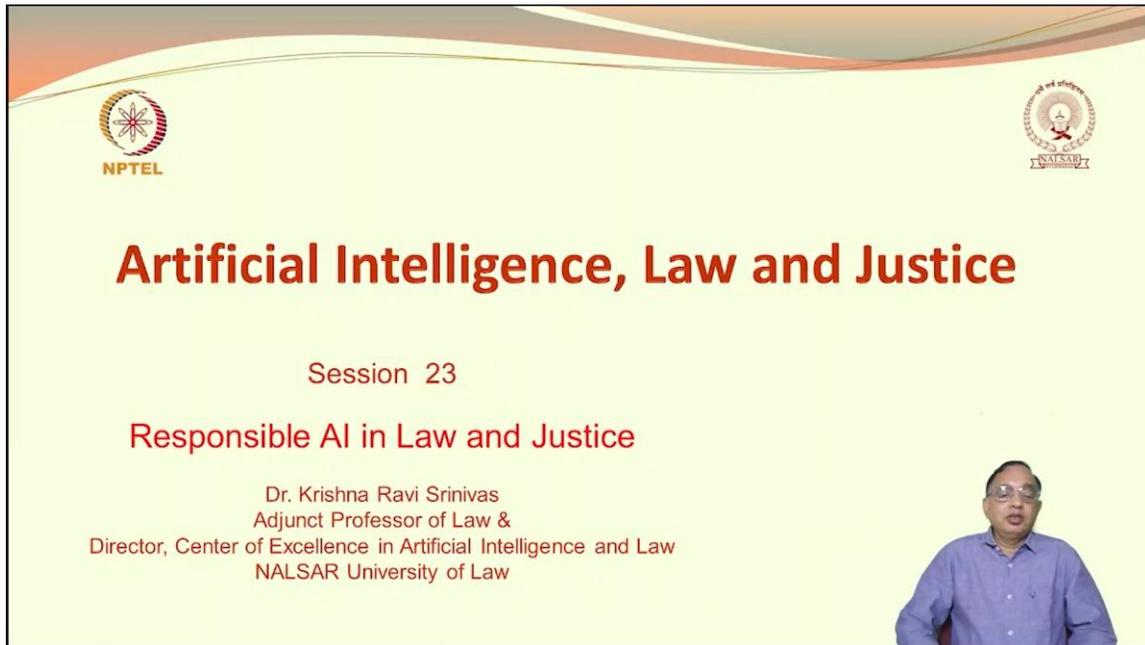Course Name – Artificial Intelligence, Law and Justice
Professor Name – Dr. Krishna Ravi Srinivas
Department Name – Center of Excellence in Artificial Intelligence and Law
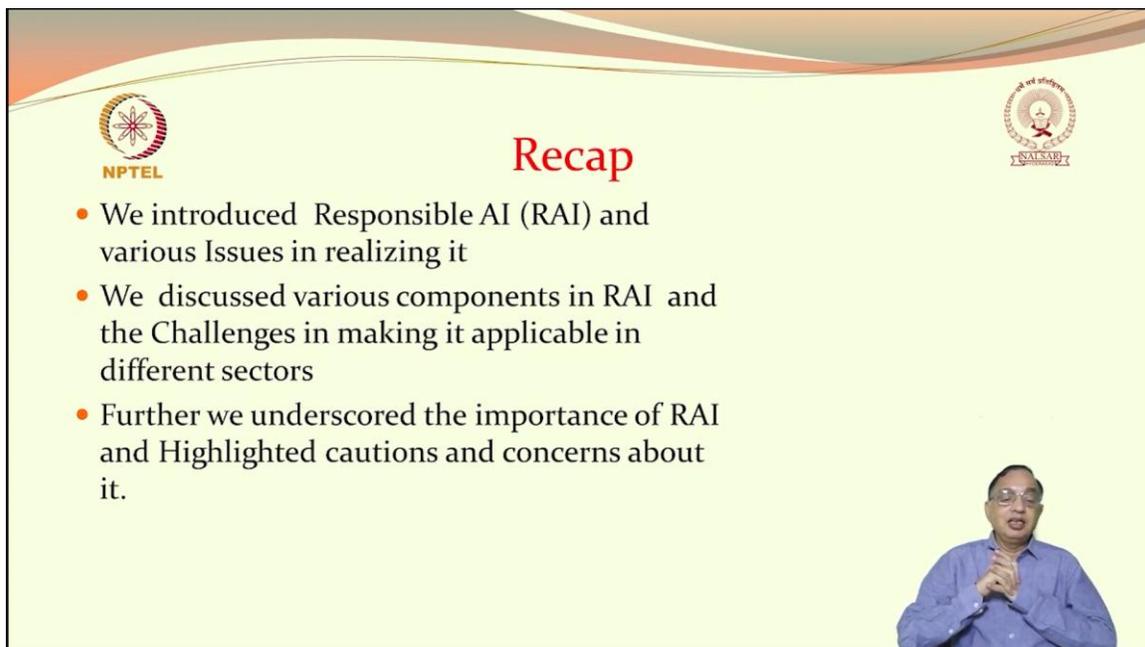Institute Name – NALSAR University of Law
Week – 05
Lecture – 23

Artificial Intelligence, Law and Justice, Session 23. The topic here is responsible AI in law and justice.

In the previous session, we learned about Responsible AI as a concept. We also looked at

the various issues in realizing it, both in practice and across different sectors. We looked at the various components of Responsible AI and then the challenges in making it applicable in different sectors. Moreover, we also underscored why Responsible AI has become very important and why so much effort is being made to promote Responsible AI and translate that into action in different contexts, particularly by different sectors. We also looked into some of the cautions and concerns about responsible AI. Particularly when there is a concern that Responsible AI could become yet another "tick the box" initiative, or that Responsible AI could be something too vague or ambiguous for people to understand, leading to individuals interpreting and implementing it in their own ways. But we need to look into something more because, as a concept, while Responsible AI is very interesting and challenging, bringing that into law and justice becomes all the more challenging and, I won't say difficult, it's becoming all the trickier. Why?
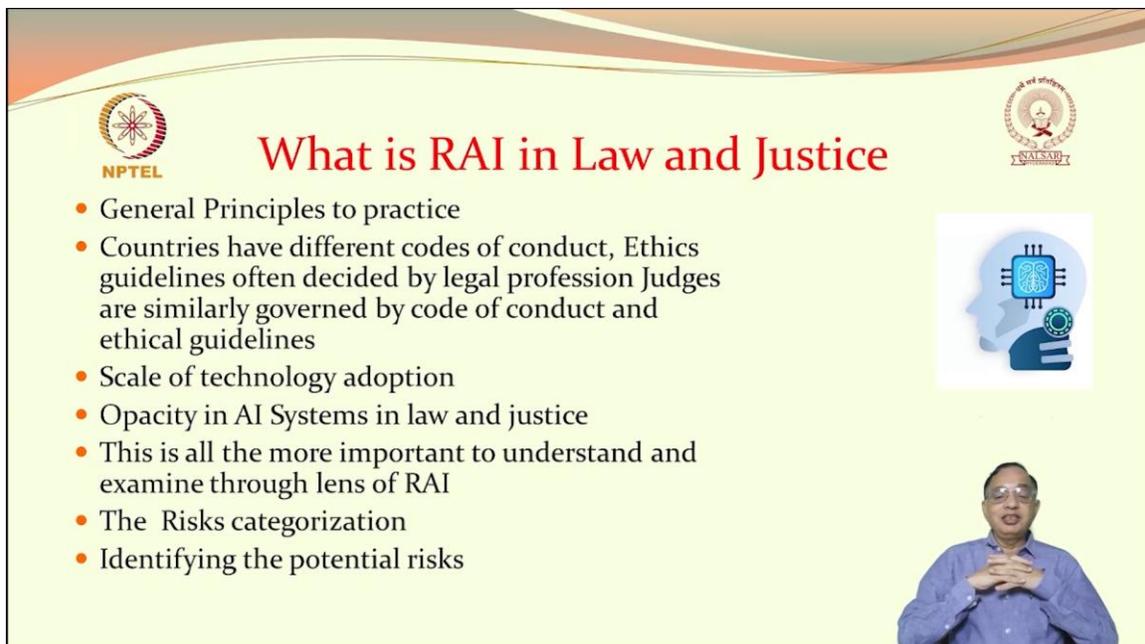
Because what exactly is responsible AI in law and justice is something that is easy to ask but very difficult to define. For the simple reason that, although there are UNESCO guidelines, although there are OECD principles, and although there are any number of guidelines and principles on what constitutes RAI or responsible AI, taking that to law and justice is not all that simple because these are very general principles. But when it comes to law and justice, different countries have different legal systems. We know about the common law system, and different countries have different systems not only based on the principle of law but also based on their historical experience, their constitution, their governance system, and other factors. And based on these parameters, a country's law and justice are also structured or tailored in such a way that they constitute the core of that. In this sense, it is the Constitution of India that drives the law and justice system in India.

In some other countries, it could be something else. But even if we say that democratic constitutions almost look the same, there are significant variations among them. And equally important is that not all democratic constitutions give the same weight or the same importance to all the rights. So, there is a wide variety. And then the wide variance becomes all the more interesting or challenging when we talk about aligning responsible AI with either the rule of law or fundamental rights, or when we say that responsible AI should not have any bias or discrimination. And when we say that Responsible AI, these are the seven or eight principles that are at the core of Responsible AI; translating them into corresponding principles that could be put to use in law and justice is not all that simple. Because the very idea of transparency is also linked to the idea of privacy, there is an obligation to disclose to the public or the public has the right to know. Different countries have different ideas and practices in that. Some countries are very opaque in certain areas, such as private data, but then they are very transparent in other areas, such as public data. In some countries, both are very difficult to access.

So, the systems are such that transparency may mostly be on paper, but data availability or access to data may be very difficult. Similarly, when we say that it is free from bias and discrimination, it could be something that is self-evident. But then different legal systems have different ways of looking at it. Some systems may have affirmative action policies embedded in the law; some may not. Or some may be talking about true gender equality. On the other hand, some systems may not be discussing gender equality, but they may be skewed when it comes to gender equality. In the sense that by default the system may be perceived as gender equal, but in practice, different laws and different practices may not be gender equal or gender neutral in some circumstances. So, the idea here is that it is very difficult to pinpoint exactly where this will fit in and then translate that into actionable points in different circumstances when it comes to law and justice. Then comes the question of civil law, criminal law, and different branches of law. And we have seen that this is why the application of AI in the criminal justice system is all the more controversial and challenging to implement, and why it has become not only controversial and challenging but also worrisome.

So, when we talk about responsible AI in the criminal justice system, certain ideas or principles have to be given more importance, whereas the same may not be true in civil law. And if you talk about labour law, again certain principles like non-discrimination and the right to equality should be given more importance. And then the responsible AI system that deals with labour law should be something that assigns significant weight to them. So, these are some of the fundamental issues when we talk about responsible AI in law and justice. Then comes the scale of technology adoption, such as how we develop technology that fulfils responsible AI across the law and justice. If the system is not going to be so adaptive in the sense that it is going to be selectively adopted, then we should look at where exactly we need to fit in the RAI principles. But different countries

have different codes of conduct. For example, different bar associations have different codes of conduct and ethical guidelines. And then the legal profession itself is a self-regulating one, although there may be other laws that regulate the legal profession. And then judges are governed by a code of conduct and ethical guidelines. But these things, however much they may appear to be the same or identical when looked at, prima facie, could have substantial and numerous differences. For example, what could be permissible in one jurisdiction as acceptable behaviour or a permissible code of conduct for an advocate might be totally prohibited in another. The same thing goes for the judges. So, when these things happen, it becomes very difficult to translate them into actionable points. Then comes the whole question of opacity.

So, in the context of law and justice, we need to really figure out when the system itself is opaque; particularly in law and justice, responsible AI is all the more important. But if the system is such that the judicial and legal system prefers to have an opaque system rather than to ask for or demand a responsible AI-based system. And then the risks of categorization are also present. In the sense that when we talk about responsible AI in law and justice, how do we categorize them? Which are more important? Which are less important among the seven or eight principles? Which should be prioritized? And so, is the potential risk: what would be the potential risk if some norms are not given that much importance? There could be a lot of difference of opinion because, in some jurisdictions, privacy may not be a fully acceptable or fully justiciable fundamental right; it would be left to the government to decide the level of privacy people can enjoy. In some other systems, privacy can be a fundamental right, and the state could be sued if it violates the privacy rules. But the AI systems will have to be factored in, taking this into account. So, can one argue that in a system where privacy is not the fundamental right in the way that it operates in other countries, the responsible AI should be tweaked in such a way that it waters down the level of privacy as a major concern, or should it be based on universal principles? Because if we do that, then the responsible AI system will find a lot of faults with the system, and that could become a contentious issue. If the lawyers agree that we can compromise privacy in the name of national security or we can compromise privacy in the name of something else, but if the responsible AI principles strictly talk about transparency, accountability, and privacy, there could be a conflict. This is not hypothetical; this is a real issue.

Then there are divergence and differentiation. In the last class, we saw how different bar associations and state-level associations are trying to come to grips with ethics in AI by issuing various guidelines. So even within a country like the USA, there is no homogeneous approach or harmonization because each association will have its own precedents, specific factors, and specific rules; taking all that into account, they will come up with specific guidelines. While the American Bar Association's broader guidelines may be applicable to all, if you are practicing in Tennessee, the guidelines tweaked to your local bar requirements would be applicable. And then the same thing would also be applicable in countries where the bar associations themselves have some code of ethics, good conduct, but then they are not fully equipped technically and otherwise to come to grips with AI. Because AI adoption across countries in law and justice is not even, many countries do not have sufficient capacity in AI itself to implement it in all spheres.

So that could be a reason why some countries may still be struggling with the older ones where it doesn't make sense to talk about responsible AI. The UN agencies have actually taken up a lot of interest, and in fact, they are working on many things that are related to law and justice. Many of the UN agencies are part and parcel of the UN system's larger initiative on capacity building, framing guidelines, developing toolkits, and many other things. Two institutions I want to highlight are this. One is UNESCO. We have seen what UNESCO is doing. The other is the United Nations Interregional Crime and Justice Research Institute, or UNICRI. So UNICRI has come up with a responsible AI toolkit for law enforcement. But when we talk of judicial decision-making, we also need to take into account how human decision-making in different jurisdictions works, because the fundamental principles may appear to be the same, but the way they work would be totally different. There could be specific reasons for it, and AI models make decisions based on complex models, and understanding them is not simple; that's the first level of

complexity. Then the complexity in judicial use needs to be understood because often the judicial system itself is a labyrinth were putting that in terms of responsible AI may be difficult to understand. Then, when it comes to different levels of transparency, different levels of accountability, or different levels of privacy and fundamental rights, the question is what exactly responsible AI should be in those contexts? Should it be a mute spectator, or can we say that we can water down responsible AI criteria and other things when we assess systems, or that responsible AI-based systems in these countries will not meet these standards that are applicable elsewhere but not in this country? So, this divergence and differentiation is a major factor that, in fact, also tells us that applying AI in law and justice is much more difficult and much more challenging than one would imagine.
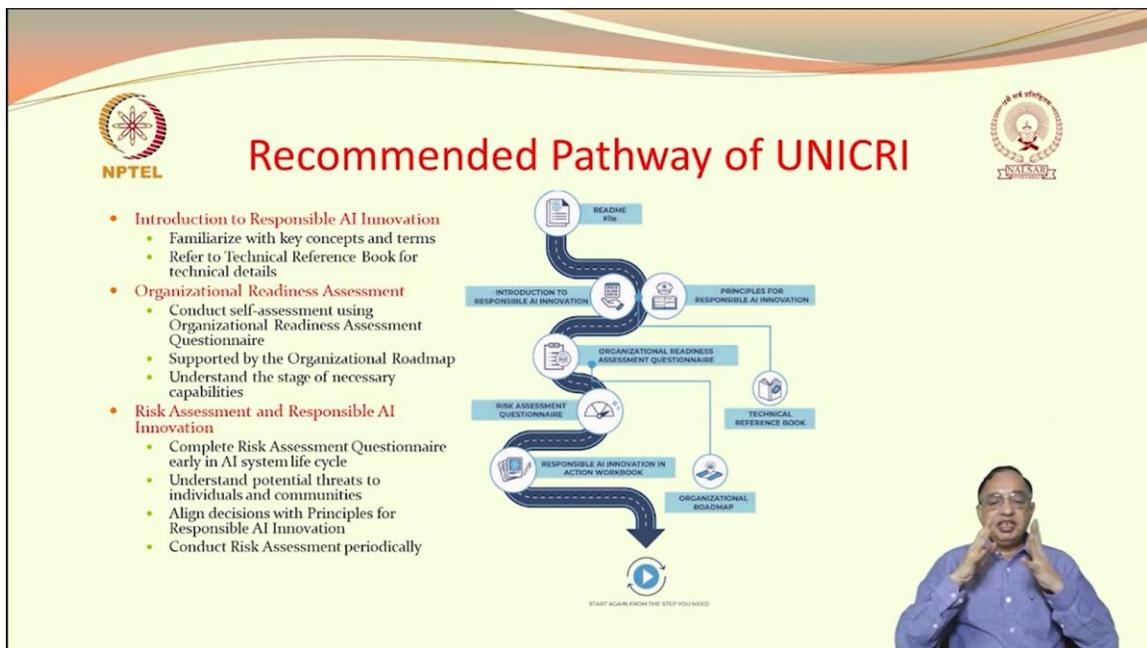
So, the UNICRI has come up with this recommended pathway where it says you start with the README text, then the introduction to responsible AI innovation, and then you go to the organizational readiness assessment. In fact, UNESCO has done a significant amount of work in this area, where they refer to it as rapid assessment methodology. They have developed a methodology for assessing the organization's capacity for responsible AI and adhering to AI principles. Then responsible AI and risk assessment will obviously go hand in hand, so they have provided some guidelines for it. Complete the risk assessment questionnaire in the AI lifecycle itself, then understand the potential threats, align decision-making with principles for responsible AI innovation, and conduct risk assessments periodically. This is the broad diagram they have provided, which is very important and interesting, but this is given in the context of not just law and justice; it is specifically in the context of legal matters related to law enforcement. So, the similar one may not be generally applicable for law and justice.

The question is how do we make them reliable and trustworthy, and then how do we bring in transparency and accountability, so the legal requirements translating them into RAI categories mean that they should be credible, and then the critical decision-making system should again be based upon RAI principles. Therefore, we need to bring in trustworthiness in critical decision-making when it comes to law and justice. Although Explainable AI is important, in my view, it is all the more important and should go hand in hand with developing responsible AI systems.

One way to look at it is that, in addition to the specific responsible AI principles, we need to have something more in the sense that the law and justice AI systems should guarantee
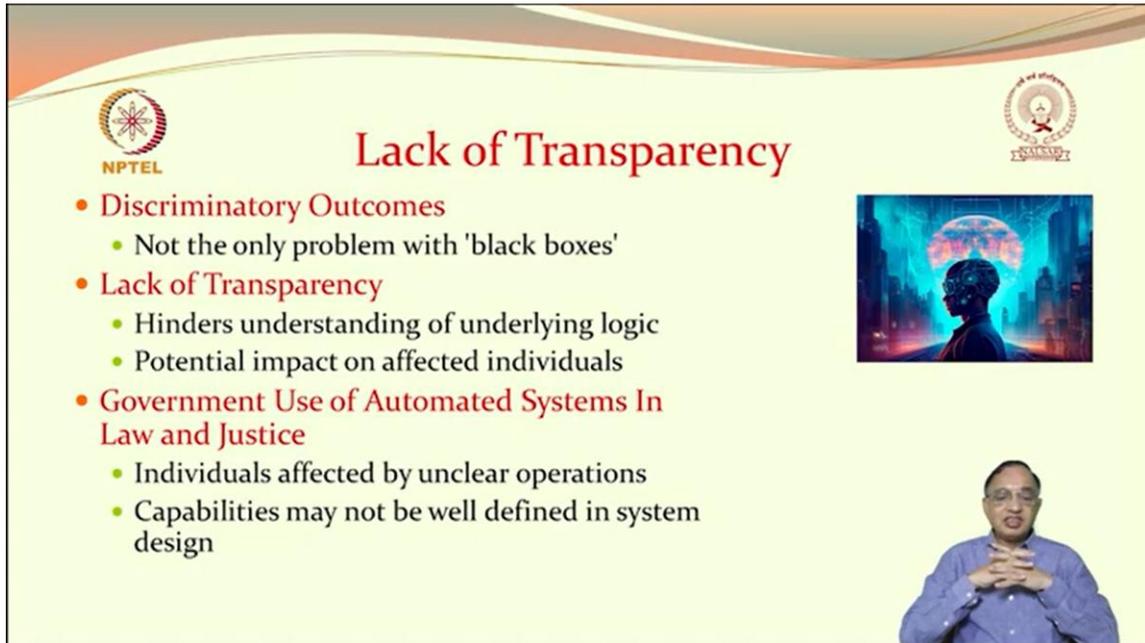
against bias and discriminatory outcomes. This goes one step further than simply being free from bias and discrimination; it should not only be free from them, but it should also guarantee against them in the sense that it goes one step more to ensure that there is no element of bias or discriminatory outcome due to the functioning of the system. But there are issues such as opacity in the AI systems. Again, how do you define bias in legal terms? How do you define discrimination in legal terms? Because these are again very contentious issues. What may appear to be a bias to me may appear to be something normal for somebody else. Because someone else may say that I do not see any gender bias in the way people were treated; they were treated equally. And then X person may say that there is a specific gender bias in this because they were not treated equally. It seemed like they were treated equally, but the decision was not based on that.

So, lawyers and judges themselves may not be very familiar with how to translate this into something that is like a guarantee against bias and discriminatory outcomes. If the system is going to result in discriminatory biases, they themselves may be impacted. But then they really need to understand the system well. They should go back to explainable AI again. They need to work with the developers to ensure that this does not happen. But the challenges of addressing bias in legal terms or even in practical terms are many. Because what happens when decision making is done based on so many parameters and criteria is that what could appear to be a bias to me may not appear to be a bias to someone else, so the decision making can be challenged. Two judges may look at the same decision making differently, and then we see this happen when two judges on a bench come up with totally different judgments; either it goes to a third judge or to a larger bench. So, when this is the real issue in law and justice, how do we translate these sorts of things into responsible AI? It is a problematic one. what sort of bias and discrimination should responsible AI be sensitized to? Then comes the question of understanding the reasons behind AI systems.

So, if we talk about responsible AI, then we should also understand that we should bring responsible AI as a criterion to bring in more transparency. And then, to make that a holistic perspective, we should bring in all seven principles and say that each one of them is applicable and fulfilled. Otherwise, what would happen is that the system may appear to have responsible AI fulfilled, but not all of them will be fulfilled 100%, or it may be more in letter than in spirit. So, these sorts of things need to be factored in. And then the solution should be a systematic algorithmic audit built into law and justice systems because a systematic algorithmic audit is primarily important to bring accountability, enhance transparency, and make the system more credible. But how do you translate systematic algorithmic audit as part of responsible AI, which has a lot of practical things to be looked into? So, the idea of integrating that as part of the larger transparency framework or part of the larger RAI framework is difficult. It is better to test the system for different factual backgrounds and then assess it in comparison with human judgment.

So, when this is done, we are not talking in terms of mere abstract responsible AI principles. We are talking in terms of the system being tested for different factual backgrounds, different case lists, or case laws, and then seeing how the decision-making is done versus how a human would have made a decision. Or to put it the other way around, if the system is fed a well-defined case and then, based on the same factors, same parameters, same cases, how does the system arrive at a conclusion? This will help us understand whether the system is really fulfilling the responsible AI criteria, when we compare it with human judgment, we may get a better idea.

It is also possible that human judgment may have some bias and discrimination, but that could be less or more than that of the AI-based system. So, we need to really do a lot of homework, or we need to really develop a lot of nuanced understanding and then work when we talk about responsible AI in law and practice rather than talking in terms of mere general principles.
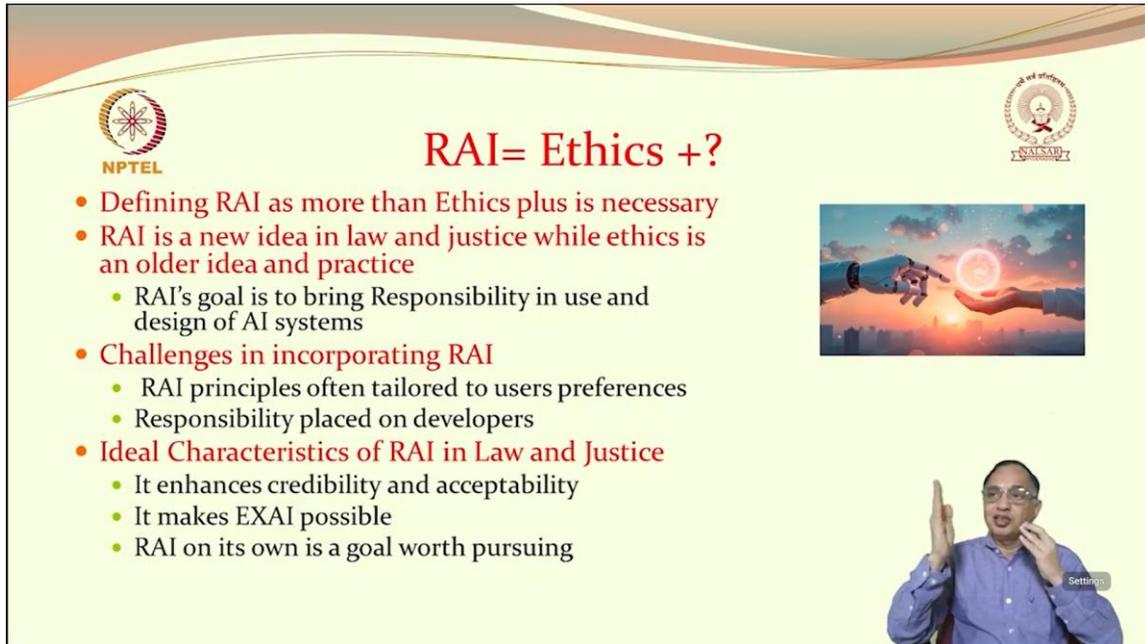

But the discriminatory outcome could occur due to other factors. And then if the lack of transparency becomes the underlying logic we won't know. The impact on affected individuals has to be factored in. So, the responsible AI systems should also flag certain things in the sense that there is a potential for certain things going wrong when it comes to transparency: flag it at that stage, stating that the system appears prima facie to fulfil the criteria of transparency, but on a deeper level, it is not exactly fulfilling the level of transparency that is expected of a legal system. So, when individuals are affected by the unclear operations or by automated systems, there could be many problems. The government may say that this is a responsible AI system; it adheres to it. We have done a quality check, we have conducted an audit, we have met the criteria, and we have been certified by somebody. But the capabilities in the system may not be well designed, or the

system design may have some problematic aspects that could hinder decision-making by the system, leading it to make biased decisions, which could be the reason the system takes wrong decisions, although prima facie it may appear to be right. So, the question here is that since we talk about the law and justice system being very sensitive, there needs to be a much higher level of transparency and a greater realization of RAI.

Translating these things into practice means a lot of work; homework needs to be done, a lot of things have to be tested and retested, and maybe some reverse engineering of the algorithmic thinking should be done to find out where exactly it happened. But the question is that, as we saw, the bar associations will immediately jump to talking about AI ethics because it is something they understand. And then ethical principles are also easy to tinker with in the sense that you apply one ethical principle, tweak it, and put it in another context. So, it is easier to come up with revised AI ethics guidelines than to come up with revised or start fresh responsible AI guidelines in law and justice. But as we saw, responsible AI is not something that is equivalent to ethics alone. So, defining AI is more necessary than having ethical guidelines because right now many of the bar associations are stopping with ethical guidelines in the context of AI. But this needs to be developed further. Ethics is an old idea; people are familiar with it, judges are familiar with it, the legal system is familiar with it, there is a broader code of conduct, good behaviour, and everything is known. But then the idea of RAI itself is so new that people may not be very familiar with it, or their understanding may be constrained by their earlier understanding of ethics.

People may think that this is nothing but translating ethics into AI systems under the guise of responsible AI. So that sort of clarity should be brought in. Then, RAI principles are often tailored to user preferences. For example, the user of a system in the banking

industry would prefer a responsible AI system that meets certain legal characteristics or fulfils certain standards set by the banking authority. But then that responsible AI system may not fulfil 100% of all the criteria of all the seven or eight principles. So, when it comes to law and justice, we may not even know what the user preferences are. If the state says that I would prefer a system that is fast, very efficient, and then gives the judgment very soon, or not something that is very liberal when it comes to giving bails or when it comes to giving remissions, then the whole idea is that if the user's preferences are going to decide what the system will do, then bringing responsible AI into that will be very difficult. So, when responsibility is placed on developers, they need to negotiate; they can at least say that they will not be able to develop a 100% responsible AI system to meet your needs or preferences. So, the ideal characteristics of a responsible AI system in law and justice are something; this, again, you may say, is very utopian and very difficult to practice. It enhances credibility and acceptability. At the next level, it makes responsible AI possible. Even otherwise, responsible AI on its own is a good goal that is worth pursuing irrespective of all the problems that we have highlighted now.
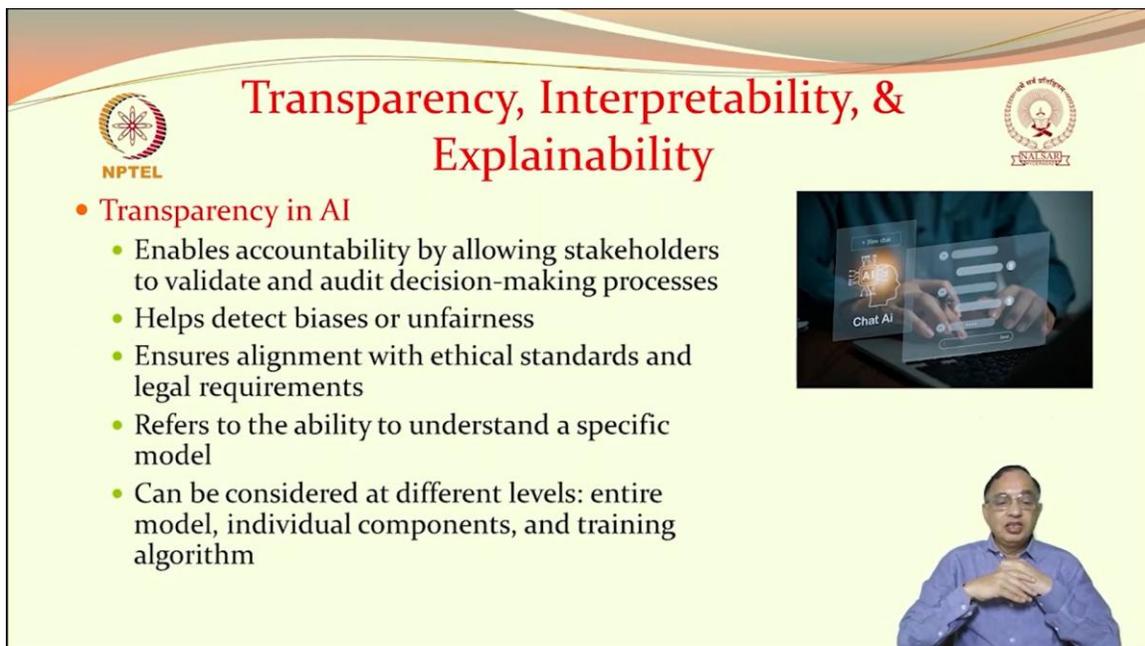
So, we can bring responsible AI into the law and justice system by increasing accountability and allowing stakeholders to validate and audit the decision-making process, but at what stage should it be done? For example, if there is a question of serious punishment in the criminal justice system that exceeds 10 years and if the judge has used AI-based tools to gain an understanding or to help him or her in writing the decision, then the stakeholders or the person who is to be affected should have the right to question the responsibility of the AI system. So, this is something that is equally important. But then, even at the level of building a system, if stakeholders can be involved, some things, like transparency, can be enhanced. Then aligning with ethical standards and legal requirements has to be done with care because it's again contextual. As we said, different
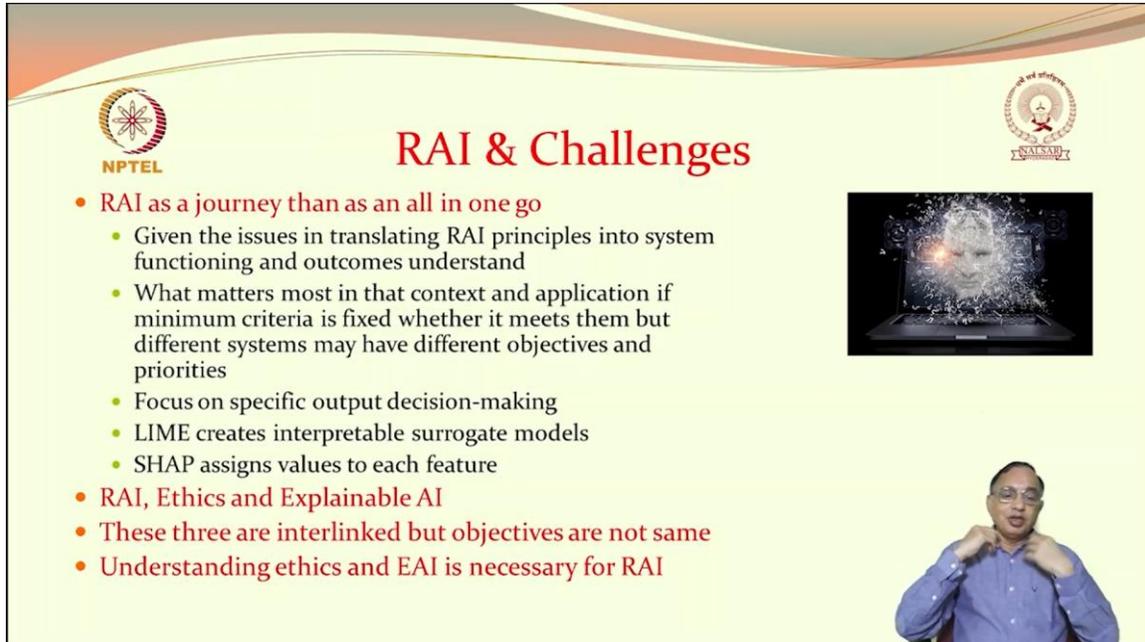
states, different jurisdictions, and different courts have different norms when it comes to ethics. What is 100% permissible, what is considered permissible but not desirable, and what is considered desirable but not abhorrent in the sense that nobody does it, although there is no bar on it, these minute, nuanced understandings are necessary when we talk about responsible AI in law and justice. So, again we need to look at the responsible AI system in law and justice as an entire model, including individual components and the training algorithm. Only then will we be able to understand whether the responsible AI principles are fully embedded and then enmeshed and operate in the system as a whole. Otherwise, what would happen is that we may be looking at certain things in certain places and then ticking the boxes. We may forget that in some other instances, in some other parts of the system, responsible AI may be just there for the namesake.

So, translation challenges exist from abstract to concrete to specific, with different interpretations putting into practice, expectation to realization. So, one way to overcome some of these initial teething troubles is to understand responsible AI as an ideal that cannot be realized in one go, in the sense that you would need to tweak the system; lots of trial and error would be needed to make the responsible AI system workable. Then also understand that the system may score high on one parameter but may not be scoring as high on other parameters. So that sort of thing we really need to factor in because if the system scores high on, say, transparency and accountability but then does not score high on some other parameter, what exactly is the reason? That needs to be looked into; it could be that the designers and the people who wanted the system to be developed had a conversation, and based upon that conversation, the people who wanted the system to be developed assigned low priority to two or three parameters and assigned a higher priority to other parameters. However, they wanted something to be incorporated, but not to a certain level of importance. So, translating that again will vary depending upon the

importance and the priority. As a common principle, it should be ensured that at least all the systems in law and justice where AI is used adhere to the responsible AI principles at least to a minimum level. It is not that certain things, values, and principles can be simply knocked off, and then we can claim that only these are essential for us.

So, as we said, take that as a journey rather than as an all-in-one-go approach. And then start with the minimum criteria to meet them; different systems may have different objectives and priorities. So, focus on specific output decision-making, whether it is criminal justice or legal justice. And then we also need to look into the differences between responsible AI ethics and explainable AI ethics, and then see that these are not competed against or taken as substitutes for one another when it comes to building responsible AI systems. Although the three are interlinked, the objectives are not the same. But understanding the ethics of explainable AI is necessary for responsible AI systems to be developed.

Conclusion

- **Variations in understanding RAI**
  - Translating RAI into AI systems in law and Justice depends on variations in understanding RAI helps meet
  - In many jurisdictions there are no specific guidelines or procedures in incorporating RAI so the paradox is there is much talk but in terms of guidelines that help both developer and users there is not much
- **Challenges of RAI**
  - Risk of misinterpretation and misalignment
- **Responsibility and Oversight by collaboration and incorporating RAI from conception of system is important**

So, the conclusion here is that there are variations in understanding responsible AI. Translating AI into AI systems in law and justice depends on variations in understanding. In many jurisdictions, there are no specific guidelines or procedures for incorporating AI, so the paradox is that there is much talk, but in terms of guidelines, very little is available that would be useful for developers, the legal profession, judges themselves, as well as users. So, the question of the risk of misinterpretation and misalignment will arise because responsible AI could be interpreted differently. We did not delve into that in detail in this lecture because this is something we talked about even earlier in another context. Transparency can be misinterpreted, and then it could be misaligned with something else. And then, responsibility and oversight through collaboration and incorporating AI from the conception system are important. Our responsible AI cannot be an afterthought. It should be something that is taken as a key component right from the way the system is designed and then the system is looked into.

Or right from where the need for the system is decided or the tender for that is decided, and then the government comes up with the tender, it should be made clear that this AI system for law and justice will adhere to responsible AI principles, and these are what they are. But then this is again the broader principle or the broader idea of bringing responsible AI into the public procurement system. That, again, is a very different concept, which we will not go into because it is not something that is directly related to law and justice. But bringing responsible AI into the public procurement system is something that would really help in making the AI systems in law and justice more aligned with responsible AI. So, we need to understand that responsible AI in law and justice is not a simple matter. It is going to be quite complex, challenging, and demanding. At the same time, it is inevitable. Now in this short lecture, we have not gone into too many details. We have not given too many explanations, nor have we looked into

too many different things from different countries, for the simple reason that the field is evolving rapidly, it is too premature at this time to highlight certain things as good examples, the better ones, or to emphasize them as those that could be adopted. It is also equally important that, in many jurisdictions, many bar associations have not categorically developed responsible AI guidelines or responsible AI-specific guidelines. Similarly, although UN organizations like UNESCO and UNICRI have been working on Responsible AI in law and justice, a lot needs to be done.

For example, the Responsible AI toolkit that UNESCO brought for judges is a good example, and one country, Colombia, has adopted it. But then much more needs to be done because that toolkit itself addresses only specific issues and only some specific values in Responsible AI. But incorporating all the elements of Responsible AI is a much broader, much larger, and more importantly, much more demanding task. That is why in this lecture, I have not gone into the level of depth that I have done in some other lectures on the rule of law or the rule of law in AI or algorithmic decision-making, because there is much more literature on that than anything else. There are a lot of legal principles that have been discussed. Whereas in the case of responsible AI in law and justice, much more needs to be written about, discussed, and then we need to develop that field in a much better way than it is now. Thank you.
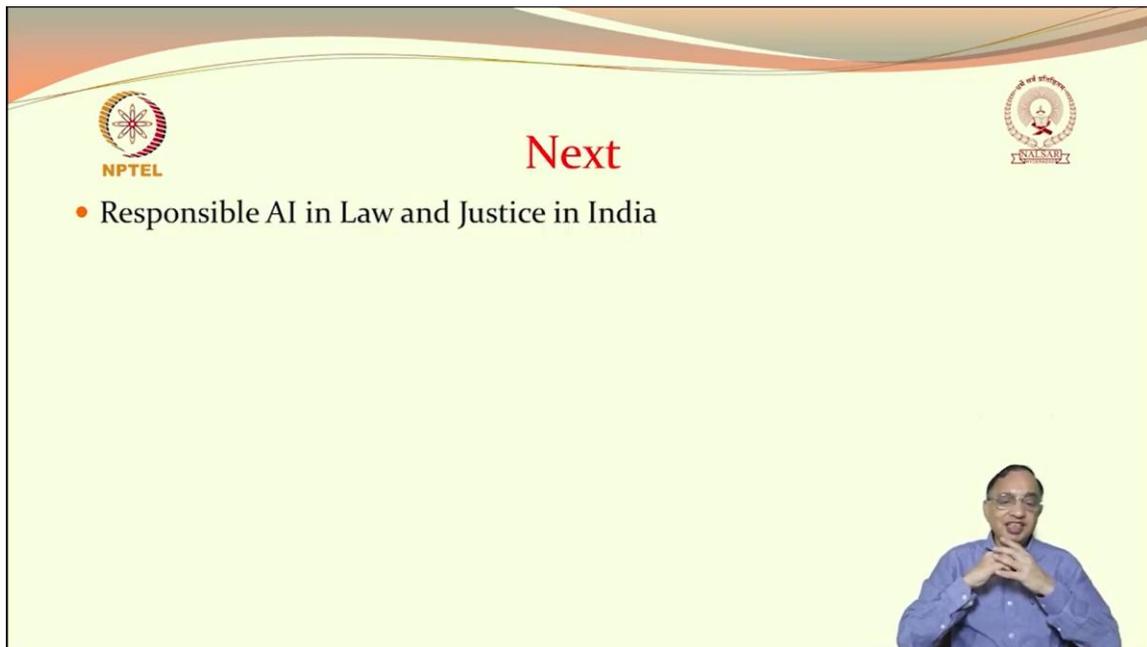
In the next class, we will look into what responsible AI is in law and justice in India. That will again be very contextual. So, the challenge we will be talking about here is how we translate responsible AI in the Indian context, who is working on it, and what that entails.