

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 – 24



Artificial Intelligence, Law and Justice

Session 24

Responsible AI in Law and Justice in India

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



Course on artificial intelligence, law, and justice, session 24. Responsible AI in law and justice in India.



Recap

- We discussed Responsible AI (RAI) in Law and Justice and the issues and challenges in it.
- We pointed out that while there is a consensus on principles of Responsible AI translating that into practice in Law and Justice has seen varied responses.
- Further we mentioned that Responsible AI in Law and Justice might mean different things to different stakeholders and preferences in values in RAI make a difference



Let us recap the previous session. In the previous session, we discussed RAI or responsible AI in law and justice and the issues and challenges related to that. We pointed out that while there is a broad consensus on the principles of AI, translating that into practice in law and justice has seen varied responses. Further, we also mentioned that responsible AI in law and justice might mean different things to different stakeholders, and preferences in values in RAI will make a difference. So, having seen Responsible AI as an idea and how it is being translated into law and justice in different contexts with different stakeholders, let us move on to the current topic, which is RAI in law and justice in India.



RAI and Playbooks

- RAI in India is discussed by inter alia, NASSCOM
- NASSCOM has come out with a play book
- Digital Futures Laboratory has come up with a play book
- On RAI in Law and Justice we have cited Vidhi's report on this topic.



See, responsible AI in India is discussed by many people, including NASSCOM. NASSCOM has come up with a playbook for developers. Digital Futures Laboratory, a civil society organization, has also come up with a playbook. We have earlier cited the work of Vidhi's report on Responsibility in Law and Justice in India, particularly in the context of courts.



RAI and Niti



- In 2021 and 2022 Niti Aayog published documents on Responsible AI for India with
- slogan "AI for ALL". It identified core principles and followed that with case studies.
- NASSCOM has come with a play book on RAI.
- DFL has published a similar one.
- In 2023 NASSCOM published a report State of RAI in India



In 2021-22, Niti Aayog published documents on Responsible AI in India with the slogan "AI for All." So, it identified the core principles and followed that with two case studies. One case study is on facial recognition technology. The second case study was on Digi Yatra. And, as we have said, NASSCOM and DFL have also published on this. In 2023, NASSCOM published a report on the state of Responsible AI in India.



NASSCOM Findings 1



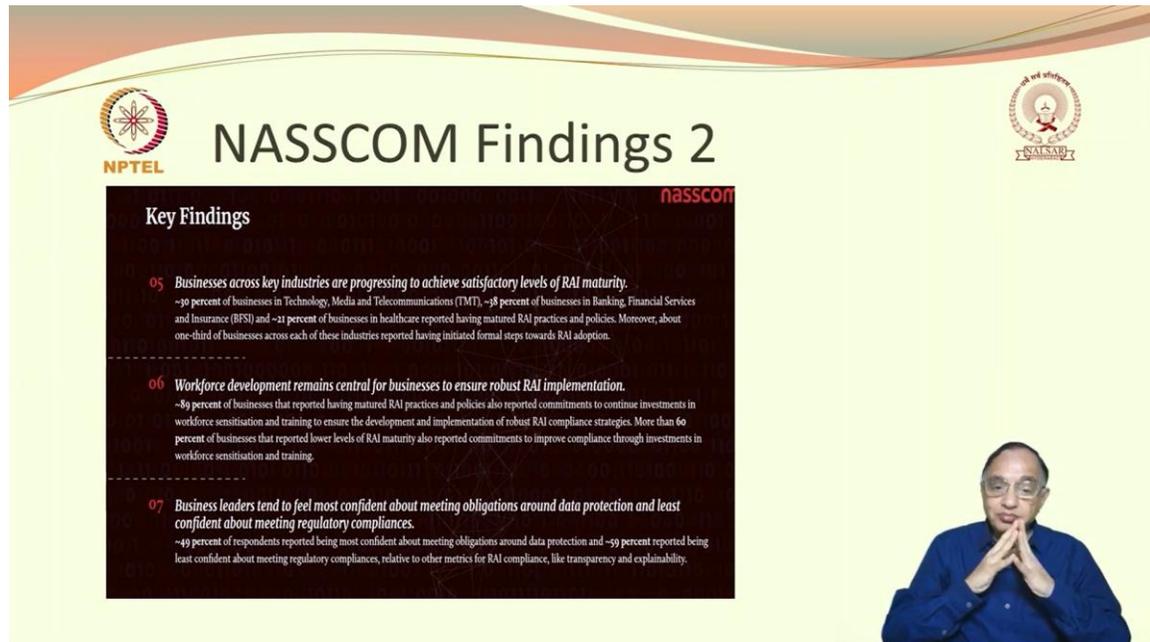
Key Findings

- 01 Indian businesses are gearing up steadily for RAI adoption.**
~90 percent of respondents reported having matured RAI practices and policies, while ~30 percent reported having initiated formal steps towards RAI adoption. ~30 percent reported having basic awareness of RAI imperatives without a formal strategy or framework, and ~10 percent reported having no RAI practices in place.
- 02 Businesses reporting higher AI maturity also tend to report higher RAI maturity.**
More than 60 percent of respondents reporting the highest level of AI maturity also reported having matured RAI practices and policies.
- 03 Large enterprises are 2.3x more likely than startups and 1.5x more likely than SMEs to report matured RAI practices and policies.**
As expected, the majority of businesses that reported having matured RAI practices and policies were large enterprises (~43 percent); ~28 percent were SMEs, and ~19 percent were startups.
- 04 Developers are almost two times more likely than users to report higher levels of RAI maturity.**
~34 percent of AI developers reported having matured RAI practices and policies compared to only ~18 percent of users.



So that report said, by and large, India is well prepared, and Indian businesses are gearing up steadily for adoption. Furthermore, it said that businesses reporting higher AI maturity also tend to have higher responsible AI maturity. It also said that larger enterprises are

much better suited than startups and SMEs for dealing with responsible AI. Furthermore, it stated that developers are better off using responsible AI principles than others.



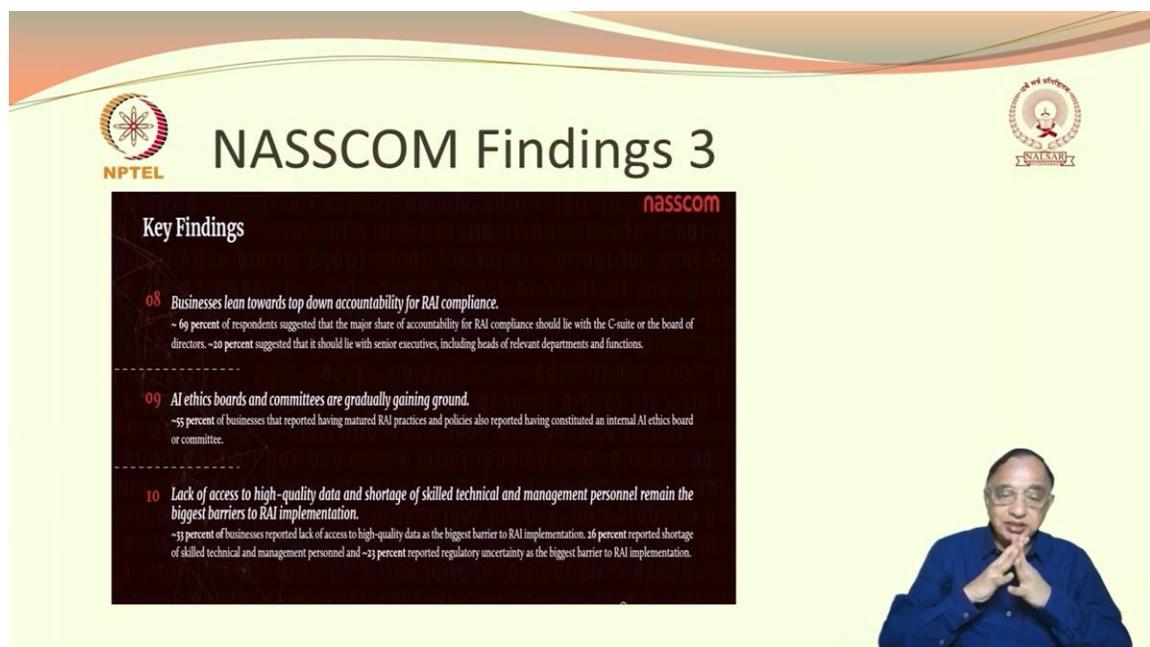
The slide features the NPTEL logo on the left and the NASSCOM logo on the right. The title "NASSCOM Findings 2" is prominently displayed in the center. Below the title, a dark box contains the text "Key Findings" and the NASSCOM logo. Three key findings are listed, each with a sub-bullet of statistics. To the right of the slide, a man in a blue shirt is shown in a video call window, with his hands clasped in a prayer-like gesture.

NASSCOM Findings 2

Key Findings

- 05 Businesses across key industries are progressing to achieve satisfactory levels of RAI maturity.**
~30 percent of businesses in Technology, Media and Telecommunications (TMT), ~98 percent of businesses in Banking, Financial Services and Insurance (BFSI) and ~21 percent of businesses in healthcare reported having matured RAI practices and policies. Moreover, about one-third of businesses across each of these industries reported having initiated formal steps towards RAI adoption.
- 06 Workforce development remains central for businesses to ensure robust RAI implementation.**
~89 percent of businesses that reported having matured RAI practices and policies also reported commitments to continue investments in workforce sensitisation and training to ensure the development and implementation of robust RAI compliance strategies. More than 60 percent of businesses that reported lower levels of RAI maturity also reported commitments to improve compliance through investments in workforce sensitisation and training.
- 07 Business leaders tend to feel most confident about meeting obligations around data protection and least confident about meeting regulatory compliances.**
~49 percent of respondents reported being most confident about meeting obligations around data protection and ~59 percent reported being least confident about meeting regulatory compliances, relative to other metrics for RAI compliance, like transparency and explainability.

And it also pointed out that businesses across key industries are progressing at rather satisfactory levels of RAI maturity. It also found that workforce development remains a significant challenge, and robust RAI implementation could be constrained. Then business leads tend to feel most confident about meeting around data protection, and so that seems to be the constraint; it is said that data protection could become a major issue for RAI.



The slide features the NPTEL logo on the left and the NASSCOM logo on the right. The title "NASSCOM Findings 3" is prominently displayed in the center. Below the title, a dark box contains the text "Key Findings" and the NASSCOM logo. Three key findings are listed, each with a sub-bullet of statistics. To the right of the slide, a man in a blue shirt is shown in a video call window, with his hands clasped in a prayer-like gesture.

NASSCOM Findings 3

Key Findings

- 08 Businesses lean towards top down accountability for RAI compliance.**
~69 percent of respondents suggested that the major share of accountability for RAI compliance should lie with the C-suite or the board of directors. ~20 percent suggested that it should lie with senior executives, including heads of relevant departments and functions.
- 09 AI ethics boards and committees are gradually gaining ground.**
~55 percent of businesses that reported having matured RAI practices and policies also reported having constituted an internal AI ethics board or committee.
- 10 Lack of access to high-quality data and shortage of skilled technical and management personnel remain the biggest barriers to RAI implementation.**
~33 percent of businesses reported lack of access to high-quality data as the biggest barrier to RAI implementation. ~26 percent reported shortage of skilled technical and management personnel and ~33 percent reported regulatory uncertainty as the biggest barrier to RAI implementation.

Then the third set of findings was that businesses are moving towards RAI compliance,

and AI ethics bodies are being set up in different organizations, as they are in the process of establishing them. Then it said that the lack of access to high-quality data, a shortage of skilled, educated labour, and management personnel could be constraints for RAI implementation; in other words, it states that high-quality data and the availability of skilled individuals in responsible AI could be constraints for India. Of course, the NASSCOM study was done on a much smaller sample survey; again, it was based on interviews. Let us also take into account that NASSCOM represents a key industry stakeholder in the software services and IT services industry in India, so its findings are very important.



Developers Playbook on RAI

- Last year NASSCOM came out with a play book on RAI for developers
- It gives guidance on Risk Mitigation for
- Discriminative AI Model
- Gen AI Model , and,
- AI application
- <https://nasscom.in/ai/pdf/the-developer's-playbook-for-responsible-ai-in-india.pdf>
- Although not directly related to applications in
- Law and Justice it can be adapted for developing
- RAI in Law and Justice



And then the NASSCOM Developers Playbook on RAI in India gives guidance on risk mitigation for three types of AI applications: one, discriminative AI applications; two, general AI models; and three, any AI applications. While this playbook is very important from a developer's perspective because it tells what should be looked into, what needs to be gathered, and how to go about risk mitigation, identifying the stakeholders, particularly taking into account what the various things are that need to be factored in if you want to implement RAI. This playbook also relies on the UNESCO ethics principle and other guidelines that are generally accepted. While this is not directly applicable to law and justice, it can be adapted for developing responsible AI in law and justice.



RAI principles in India



- DFL's playbook on RAI refers to UNESCO's AI ethics principles and is meant for social impact organizations.
- ICMR's "Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare" is another documents that is relevant for RAI in India.
- In January 20255 a multistakeholder body under the Ministry of Electronics and Information Technology (MeitY) published a document outlining AI governance principles for mitigating AI harms and promoting RAI.



Whereas the DFL playbook on RAI refers to UNESCO's AI ethics principles, it is meant for social impact organizations. So, while this is useful, it has a very limited purpose in the sense that it is oriented towards social impact organizations. On the other hand, ICMR's Ethical Guidelines for the Application of Artificial Intelligence in Biomedical Research in Healthcare is another document that mentions RAI in India, but again, the focus is more on biomedical research in healthcare. So that is again a totally different area from law and justice. So, we do not have much to pick up from that for developing RAI in India in the context of law and justice. In January 2025, a multi-stakeholder forum of the Ministry of Electronics and Information Technology published a document outlining AI governance principles for mitigating AI harms and then promoting RAI.



RAI Principles



1. **Transparency:** AI systems should be accompanied with meaningful information on their development, processes, capabilities & limitations, and should be interpretable and explainable, as appropriate. Users should know when they are dealing with AI.
2. **Accountability:** Developers and deployers should take responsibility for the functioning and outcomes of AI systems and for the respect of user rights, the rule of law, & the above principles. Mechanisms should be in place to clarify accountability.
3. **Safety, reliability & robustness:** AI systems should be developed, deployed & used in a safe, reliable, and robust way so that they are resilient to risks, errors, or inconsistencies, the scope for misuse and inappropriate use is reduced, and unintended or unexpected adverse outcomes are identified and mitigated. AI systems should be regularly monitored to ensure that they operate in accordance with their specifications and perform their intended functions.
4. **Privacy & security:** AI systems should be developed, deployed & used in compliance with applicable data protection laws and in ways that respect users' privacy. Mechanisms should be in place to data quality, data integrity, and 'security-by-design'.



So that document was again based on a lot of stakeholder consultation. It itself had stakeholder representation identifying some RAI principles for India. This has again been based on the OECD ethical guidelines, and other documents, particularly the Global Partnership on AI's documents. So, it identified transparency, then it defined transparency as an AI system that should be accompanied by meaningful information on its development processes, capabilities, limitations, and should be interpretable and explainable, as appropriate; users should know when they are dealing with AI. So, this is the first principle. Then comes accountability. Developers and deployers should take responsibility for the functioning and outcome of AI systems and for the respect of user rights, the rule of law, and the principles mentioned earlier. So, a mechanism should be in place to clarify accountability. This again is something that we have seen even earlier: when it comes to RAI, accountability is a key parameter. Then, safety, reliability, and robustness: AI systems should be developed, deployed, and used in a safe, reliable, and robust way so that they are resilient to risks, errors, or inconsistencies, and the scope for misuse and inappropriate use is reduced, and the unintended, unexpected adverse outcomes are identified and mitigated. Although it is stated in a single sentence, this is an issue of high importance because they are trying to combine three important things: safety, reliability, and robustness. While safety and reliability can be easily defined, robustness can only be tested over a period of time. In the sense that no system can be certified to be 100% robust at the moment it is developed and started to be deployed.

And then, AI systems should be regularly monitored to ensure that they operate in accordance with their specifications and perform the intended functions. This is again obvious, but the regular monitoring is a key question because, in any system, if the regular monitoring is to be done by a third party for external evaluation and then for vetting, or if it should be done by the internal processes, is again a big question. If it is

going to be done by a third party, then there should be very specific, clear-cut standards so that the third party can assess whether the system is able to meet those high standards. But if it is going to be something internal, then the problem would be that the internal process should be vetted by a third party or should be very close to or identical to the industry standards; otherwise, what will happen is that this could again become yet another matter of ticking the boxes. And then privacy and security: AI systems should be developed, deployed, and used in compliance with applicable data protection laws and in ways that respect users' privacy. Mechanisms should be in place to ensure data quality, data integrity, and security-based design. This again is very important, but the problem here is that to translate this into practice fully, you need a well-designed, well-developed data protection regime in place. And more importantly, when it says data quality, data integrity, and security by design, security by design again has to be standardized or must meet some standards. So, someone will have to certify that security by design is fully applicable here and that all the standards have been met.



RAI Principles

- 5. **Fairness & non-discrimination:** AI systems should be developed, deployed, & used in ways that are fair and inclusive to and for all and that do not discriminate or perpetuate biases or prejudices against, or preferences in favour of, individuals, communities, or groups.
- 6. **Human-centred values & 'do no harm':** AI systems should be subject to human oversight, judgment, and intervention, as appropriate, to prevent undue reliance on AI systems, and address complex ethical dilemmas that such systems may encounter. Mechanisms should be in place to respect the rule of law and mitigate adverse outcomes on society.
- 7. **Inclusive & sustainable innovation:** The development and deployment of AI systems should look to distribute the benefits of innovation equitably. AI systems should be used to pursue beneficial outcomes for all and to deliver on sustainable development goals.
- 8. **Digital by design governance:** The governance of AI systems should leverage digital technologies to rethink and re-engineer systems and processes for governance, regulation, and compliance to adopt appropriate technological and techno-legal measures, as may be necessary, to effectively operationalise these principles and to enable compliance with applicable law.

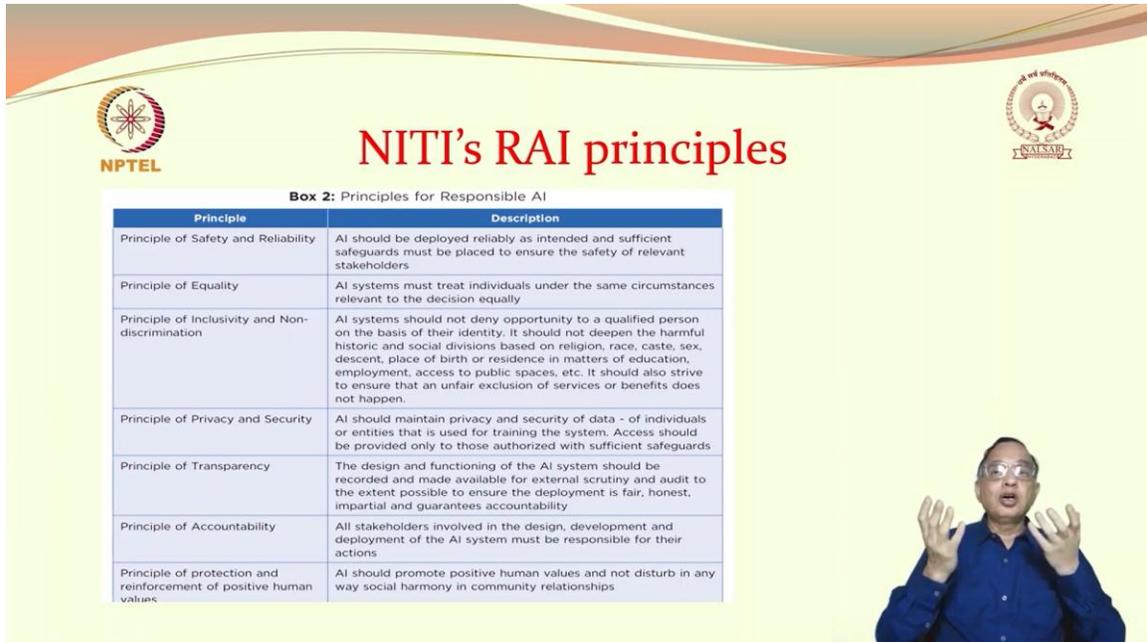


Then fairness and non-discrimination: AI systems should be developed, deployed, and used in a way that is fair and inclusive for all and the one that does not discriminate or perpetuate biases or prejudices against or preferences in Favor of individuals, communities, or groups. This again is something of a high order for the simple reason that it tries to combine fairness, inclusivity, and accessibility for all, and that does not discriminate on the basis of this. This, again, is as we said, something of a high order, but the problem with implementation would be that if you are going to apply non-discrimination fully, you will find it difficult to bring in certain things where you can accommodate certain groups to have representativeness built in here. So, in practice, it could also mean that whom to bring in, whom should not be discriminated against, and how to make it fully representative may not be something that meets 100% of the criteria

of non-discrimination. This is something very similar to reservations in education and employment. Reservation in education and employment is called affirmative action, and it is done for a specific purpose so that there are access and availability for people who would otherwise be left out. But combining that with the principle of equality, the government and the courts have devised so many systems that they can balance what otherwise appears to be mutually inconsistent rules. So similar rules will need to be applied when we talk about RAI principles, particularly law and justice. So, the fairness and non-discrimination translate into specific guidelines and rules; we will need to keep in mind that India also has policies for affirmative action, which are also known as reservations, and then a lot of things can be learned from that to strike a balance between fairness and non-discrimination and the provision for representation of those who are underrepresented. Then, of course, the human-centred, do-no-harm value; all these things are obvious. It should be subject to human oversight, judgment, and intervention as appropriate to prevent undue reliance on AI systems and to address complex ethical dilemmas that such systems may encounter.

Mechanisms should be in place to respect the rule of law and mitigate adverse outcomes on society. For this, we have also seen that there are many ways to handle it, and we have noted that RAI can be viewed from a technical perspective and again from a socio-technical perspective as well. So here we need to bring in the technical perspective and the non-technical broader societal perspective so that both can be balanced together. And, of course, inclusive and sustainable innovation: the deployment and development should be aimed at distributing the benefits of innovation equitably. This needs lots and lots of work to be done, for the simple reason that we do not attribute this inclusive and sustainable innovation to AI systems. Whereas we talk of inclusion, the broader use of the data should be inclusive, the algorithm should not discriminate, and the training model should not discriminate. But when we talk of inclusive innovation and sustainable innovation, it means that the benefits of the innovation should be distributed equitably. How we will do it in an AI system is a matter of debate because AI systems are not meant to bring about equitable distribution per se. That is not the objective. Often, AI systems are meant to achieve some goals, including business goals. So, all AI systems should be used to pursue beneficial outcomes for all, and delivering on sustainable goals, again, is a very challenging, if not impossible, goal to meet. Why? The problem is that AI's role in sustainable development goals is well recognized, and there is an increasing amount of literature on it. At the same time, we have seen that AI is also posing a huge issue in terms of resource utilization, resource allocation, and particularly in the context of the huge demand for energy, water, and other resources. So, whether AI is really a sustainable innovation is a hotly debated topic. So, if we take that into account, how to make AI systems as inclusive and sustainable is a matter of debate, and maybe here we would need lots of standards and a lot of deliberation to understand how to translate these things into practice. Then digital by design governance: So here the idea is to leverage

digital technologies to rethink and re-engineer systems and processes for governance regulation compliance to adopt appropriate techno-legal measures as may be necessary to effectively operationalize these principles and enable compliance with applicable law. So, the eighth criterion is something that sets the bar very high in the sense that it says that Digital by Design governance should take into account appropriate technical, technological, and techno-legal measures as may be necessary. So, if these eight principles are met in full, we can claim that the system is something that can meet the goal or can be classified as a responsible AI system.

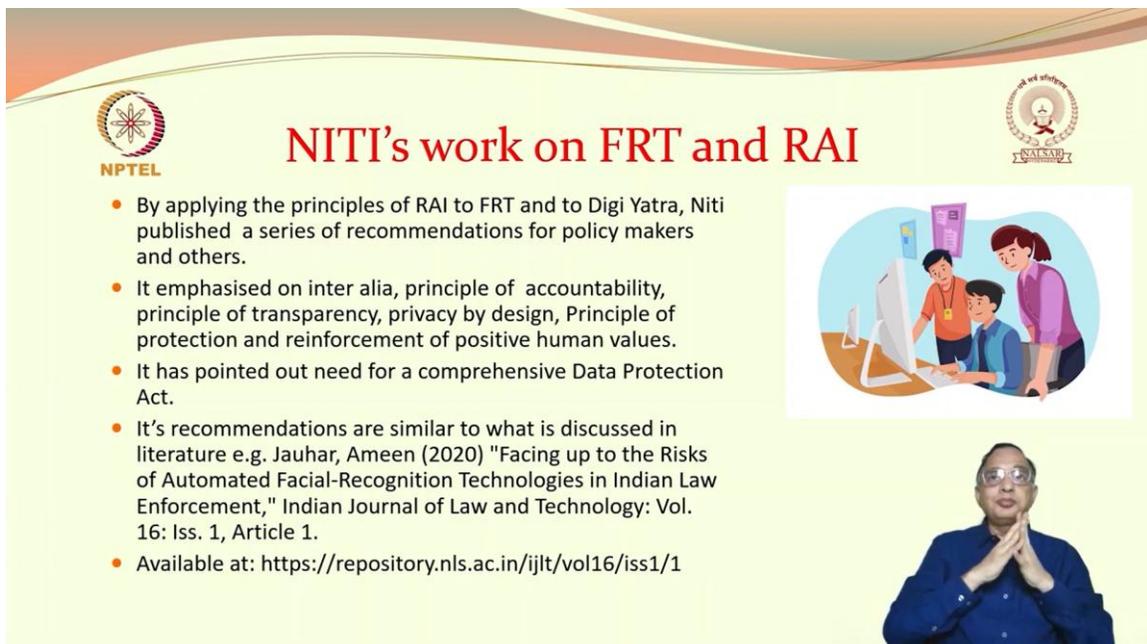


The slide features the NPTEL logo on the left and the NITI logo on the right. The title "NITI's RAI principles" is centered in red. Below the title is a table titled "Box 2: Principles for Responsible AI". To the right of the table is a video inset showing a man in a blue shirt speaking with his hands raised.

Principle	Description
Principle of Safety and Reliability	AI should be deployed reliably as intended and sufficient safeguards must be placed to ensure the safety of relevant stakeholders
Principle of Equality	AI systems must treat individuals under the same circumstances relevant to the decision equally
Principle of Inclusivity and Non-discrimination	AI systems should not deny opportunity to a qualified person on the basis of their identity. It should not deepen the harmful historic and social divisions based on religion, race, caste, sex, descent, place of birth or residence in matters of education, employment, access to public spaces, etc. It should also strive to ensure that an unfair exclusion of services or benefits does not happen.
Principle of Privacy and Security	AI should maintain privacy and security of data - of individuals or entities that is used for training the system. Access should be provided only to those authorized with sufficient safeguards
Principle of Transparency	The design and functioning of the AI system should be recorded and made available for external scrutiny and audit to the extent possible to ensure the deployment is fair, honest, impartial and guarantees accountability
Principle of Accountability	All stakeholders involved in the design, development and deployment of the AI system must be responsible for their actions
Principle of protection and reinforcement of positive human values	AI should promote positive human values and not disturb in any way social harmony in community relationships

NITI came up earlier with its own broad guidelines on RAI principles, NITI's work was much earlier. See, when NITI started working in this field in 2021 and 2022, the idea of RAI was developed, but it was not as developed then as it is now. On the contrary, at that time, RAI was more or less in the shaping stage or, in a sense, it was in the evolutionary phase. So, NITI took into account the developments at that time, identified the OECD principles, the AI ethics principles of UNESCO, and various other guidelines, and then identified a set of principles: principles of safety and reliability, principles of equality, principles of inclusivity and non-discrimination, principles of privacy and security, principles of transparency, principles of accountability, and principles of protection and reinforcement of positive human values. So, the seventh and last principle for the protection and enforcement of positive human values means AI should promote positive human values and not disturb in any way the social harmony in community relationships. This, again, is something that is very difficult to expect from AI because, for the simple reason, most systems developed in AI are not the ones that will be directly dealing with social harmony or with the idea of promoting positive human values. So, NITI's RAI principles are well maintained and done with a lot of interest and care.

But if you look at both the committee's recommendations and NITI's RAI principles, there is a bit of overlap, but they also differ. Here, particularly, if we compare them side by side; we would realize that the same terms may be used in different ways. For example, equality is mentioned here; there, equality is not mentioned explicitly by the word "equality," but through non-discrimination, inclusion, etc. So, the idea here is again privacy and security; it is also mentioned there. So, what happens in RAI or responsible AI literature or lingo, particularly, is that certain things are mentioned in a very positive way; for example, the principle of equality or non-discrimination. Certain things are mentioned in a way that is not exactly positive, but then to say that a system should be non-discriminatory, non-biased, etc. So, what is exactly covered by equality will also be mapped by non-discriminatory, non-biased, etc. But as the terms vary, the problem comes in translating them into practice. Because it is easy to say what equality is and then identify that we are giving equal opportunities to all, the system is devised in such a way that it treats all individuals as equals. But as we said, in some circumstances, particularly in a country like India, where there are principles of affirmative action and inclusion, equality needs to go beyond non-discrimination. It has to account for the representation of those who are otherwise likely to be excluded. So, the challenge in translating NITI's RAI principles, as well as the earlier ones we saw from the committee appointed by NITI, lies in precisely defining what we mean by equality, non-discrimination, and then non-bias.



The slide features a yellow background with a decorative orange and white wave at the top. On the left is the NPTEL logo, and on the right is the NALSAR logo. The title "NITI's work on FRT and RAI" is centered in red. Below the title is a bulleted list of points. To the right of the list is an illustration of three people (two men and one woman) looking at a computer monitor. At the bottom right is a photograph of a man in a blue shirt with his hands clasped in a prayer-like gesture.

NITI's work on FRT and RAI

- By applying the principles of RAI to FRT and to Digi Yatra, Niti published a series of recommendations for policy makers and others.
- It emphasised on inter alia, principle of accountability, principle of transparency, privacy by design, Principle of protection and reinforcement of positive human values.
- It has pointed out need for a comprehensive Data Protection Act.
- It's recommendations are similar to what is discussed in literature e.g. Jauhar, Ameen (2020) "Facing up to the Risks of Automated Facial-Recognition Technologies in Indian Law Enforcement," Indian Journal of Law and Technology: Vol. 16: Iss. 1, Article 1.
- Available at: <https://repository.nls.ac.in/ijlt/vol16/iss1/1>

Second, the problem also arises of how we prioritize among these 7, 8, or 10 values. In the sense that when we develop an AI system, which ones are the core ones that should be there at all times, or can we say that all the 10 or all the 7 are equally important? It could be said that all seven are equally important, but then the problem is that mapping these things into principles and then to guidelines, and then seeing how they need to be

incorporated into the rules for developers is the real challenge. Of course, companies have come up with some rule books, some guidance, and then a lot of companies, including Google, Microsoft, and Apple, are working on various principles, so there is a lot to pick up from that. Even the book "The Developers Playbook" by NASSCOM itself has good examples, so certain things can be picked up from there. But the real problem, particularly in the context of law and justice, is that given the wide diversity of AI systems that will be developed or that are being developed across sectors, how do we make them sector-specific? For example, what equality means in one sector has to be clearly defined or clearly mentioned. But then equality in another sector may be completely different. For example, equality in employment law is very different from equality in some other sectors because the purpose of the AI system must also be aligned with the sectoral or legal values of that specific sector. So, translating these into practice is a real challenge.

Having said that, we will now look into what NITI's work is on facial recognition technology and responsible AI. So NITI did apply these principles to facial recognition technology as it is being practiced elsewhere and also looked at what was being done in India, particularly in Digi Yatra. It came up with a series of recommendations for policymakers and others. I am not repeating the same thing here for two reasons. One: This was done in 2022. Today, if the same work has to be done on FRT, we need to revisit many of them, update them, and see what the major issues are with applying facial recognition technology, particularly in India and in the context of apps like Digi Yatra. Then the second point is that facial recognition technology is being used in India now in different places and in different contexts, but so far, we do not have a specific law to regulate it or govern it. So, the lack of a specific regulation or guidance is also a major factor. So, if we have to re-evaluate NITI's work on FRT and RAI, we need to understand it in the specific context. So, it gave emphasis to the principle of accountability, principles of transparency, privacy by design, the principle of protection, and the reinforcement of positive human values.

Now this reinforcement of positive human values, even if you take that out as something that is very vague or very difficult to translate into practice, the other four are core ones that should be applicable to any facial recognition technology app or application. But NITI also said that we need a comprehensive data protection act, and that act has now been brought in, but the full implementation has not taken place. Rules have been notified, rules have also been implemented, but the whole gamut of rules is not in place, and the authority is not in place. And then, NITI's work can be compared to what is discussed in the literature. For example, I am citing an article in the Indian Journal of Law and Technology where this is mentioned. Similar ideas were raised regarding the risks of automated facial recognition technologies in Indian law enforcement. So, by now

we have seen a couple of examples where the ideas of RAI are being translated into practice either through guidelines, principles, or specific guidelines for developers.



RAI in Law and Justice



- Although RAI in Law and Justice in India is discussed in the literature there are gaps that may hinder fuller development and realization
 1. Absence of a data protection regime: The DPDP Act is yet to be implemented in full scale.
 2. Absence of studies on specific aspects in RAI in Law and Justice: While Niti's report on FRT is important we need similar studies on other technologies including algorithms
 3. Governance framework on AI including algorithms and deployment of ADM.



But the real challenge in the context of law and justice is much more complex for the simple reason we have discussed: that RAI in law and justice is mentioned in the literature, but then there is not much available, and there are gaps in the literature itself. Following are the ones that could hinder the fuller development of the realization of responsible AI in law and justice. We did discuss in brief the criticism by Vidhi on Responsible AI in the judicial sector in India. We cited that. That is one thing we can keep in mind. But let us look at the broader picture. As of now, the DPDP Act is not yet implemented on a full-scale. So, translating RAI in law and justice would definitely need something that works only when there is a data protection authority, the rules are in place, the mechanisms are in place, and then it is fully enforced. Then we do have only very limited studies on responsible AI in India, and particularly in the context of law, that are much more limited. However, the availability or non-availability of these studies is also a concerning factor because, as we have discussed in so many previous instances, we need to contextualize RAI in law and justice, particularly in light of the earlier points we raised in the context of automated decision-making, AI, and the rule of law. And, of course, AI application in different laws is something that we can set aside for a moment. And we also need to look into the way algorithms are regulated in India. So, we need more studies, including on algorithms; we need to see how they are getting implemented, what the guidelines are, if any are available, what the pros and cons of the current guidelines are, and then the current implementation. Then we do not have a governance framework for AI, including algorithms and the deployment of ADM in India.

This, again, is a larger problem of non-availability of a governance framework for AI. The Government of India has not come up with a governance framework for AI because the idea seems to be that the government wants to promote innovation and does not want to bring upfront a governance regime that will more or less regulate the industry. The government's role or intention as of now is to find a comprehensive binding regulation for AI governance, but to allow things to evolve over a period of time, considering first the risk and safety angle, allowing for some voluntary guidelines from the industry. The NITI committee then came up with some recommendations on AI governance principles—see how that gets translated into practice and specific sectors. But if we have full-fledged AI governance, it will be easy for that to be implemented into law and justice, more specifically. For example, the EU GDPR Act is something that can be translated into data protection privacy and a whole lot of other related concerns when we apply AI. Similarly, a broader Act like the EU AI Act will cover many aspects relating to law and justice, particularly algorithmic discrimination, bias, non-discrimination, and privacy protection by the Act itself. Translating those specific principles in law and justice can be done easily by borrowing certain relevant articles and provisions and linking them with other relevant articles and provisions, which will really guide or enable developers as well as other stakeholders to understand what exactly they should do. In the absence of a governance mechanism, there will be a lot of questions that remain unanswered.

NPTEL

RAI in Law and Justice

NALSAR

- Guidelines for advocates, courts and others on using AI
- Guidelines on assessment of AI systems in Law and Justice
- Absence of specific policy on RAI in Law and Justice
- Lack of information on use of predictive policing in India
- Delhi Police's Crime Mapping Analytics and Predictive System

(CMAPS ,predictive policing software which determines 'hotspots' in the city and used to decide how much force is required and where to be deployed. The Status of Policing in India Report (2019) (https://www.commoncause.in/uploadimage/page/Status_of_Policing_in_India_Report_2019_by_Common_Cause_and_CSDS.pdf) But there is no comprehensive study of such use elsewhere

Then, coming to the specific sector, we don't have any guidelines for advocates, courts, and others on using AI. Of course, the Supreme Court has come up with certain things in the sense that the Supreme Court is doing what needs to be done at that level. And then there is a lot of clarity. I mean, there is a lot of clarity available among the judges and among advocates on what to do and what not to do when it comes to AI. And it has been

made clear, time and again, that AI should not be used for writing judgments or for decision-making. But what exactly should be done? Guidelines that will help the advocates and others understand tools, what they should do in specific circumstances, and what they should avoid at all costs are not available. More importantly, we don't have any guidelines for the assessment of AI systems in law and justice, but this again is a problem of the larger issue where we don't have any guidelines to assess AI systems. Of course, there are a lot of guidelines available from private players; the private industry has its own guidelines, which are either given by those who develop the systems, by consultants, or by various agencies. And there is no specific policy on responsible AI in law and justice in India, which is again something we need to mention. Although there are a lot of things happening in AI in the context of law and justice in India, we don't have a clear-cut policy as to which apps should be prioritized and who will be coordinating what, or what exactly the role of the state is in that. Whether the state and private players will come together and develop voluntary guidelines for regulation or for assessment is not clear.

And one important gap in this is the lack of information on the use of predictive policing in India. The predictive policing in India is being used, but there is no literature that really maps what is happening in India. There are reports here and there. For example, I have given one: CMAPS predictive policing software that determines hotspots in the city and is used to decide how much force is required for law and order and where to be deployed. A 2019 report on the status of policing in India cites this. But there is no comprehensive study on such use elsewhere in India, and as far as I know, there is only one study based on Delhi by Shivangi Narayan; it's more like her thesis, which looked into how the Delhi police use predictive policing for crime control, management, and identifying potential spots where crime can take place. But we really do not know who the police are who are using predictive policing and under what guidelines and rules it is being done. So, as I said, we don't have specific case studies; we don't have a lot of literature available on algorithmic auditing, and we don't have literature available on how to deal with algorithmic decision-making, along with a whole lot of issues. So, in the absence of all this, translating RAI into law and justice is really a big challenge in India.



RAI in Law and Justice



- On the positive side we have a well aware civil society Groups like Vidhi and Daksh and public interest AI organizations like agami <https://agami.in/>
- Responsible AI in the Law and Justice in India needs a nuanced balance between innovation and safeguarding citizens' fundamental rights. For this we need inter alia, robust regulatory frameworks, accountability mechanisms, Guidelines on ADM and a functioning data protection regime.



But let us look at the positive side, as well. On the positive side, you would find that there is a lot of awareness in RAI in India, which we can also see on many law and legal websites, as well as in the writings of many people. Groups like Vidhi and Daksh are working on it. Daksh is again a civil society group that is specifically working on the technology law interface. They have also done a lot of work on it. And then there are organizations like Agami. Agami, again, is a public interest organization working very closely with the stakeholders to bring AI into law and justice, particularly for apps that could benefit the larger public, and also to make AI available through law and justice access. So, there are institutions like Agami, and there are other initiatives that are working on various tools for enhancing access to justice in India. So, based upon all this, we can say that responsible AI in law and justice in India needs a nuanced balance between innovation and safeguarding citizens' fundamental rights. For this, we need inter alia a robust regulatory framework, accountability mechanisms, guidance on ADM, and a functioning data protection regime. So, if these things were to materialize, translating the RAI principles in law and justice could be much easier or could be done in a much better and faster way. But since there are so many constraints already in place, talking about responsible AI in law and justice in India is very relevant, so that at least there is some nudging towards translating this into practice. But irrespective of that, within the next four or five years, as things converge, in the sense that a fully functional data protection regime, availability of guidelines, and other materials for guiding various stakeholders, along with various pressures from different stakeholders, are likely to result in translating RAI into law and justice in India.



Next

- Explainable AI



So, with this, we have covered responsible AI in detail as a concept and practice, responsible AI in the broader context of law and justice, where we discussed what is happening in different countries and the specific context of India, particularly responsible AI in the context of law and justice in India. So, having done this, let us move to an equally important topic, which is explainable AI. So, in the next lecture, we will deal with explainable AI, and that will be followed by another lecture where we will discuss the way we have done now: what we mean by explainable AI in law and justice. Thank you.