

**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 – 07**  
**Lecture – 32**



# Artificial Intelligence, Law and Justice

Session 32

AI and Judges

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



Artificial Intelligence Law and Justice Course: This is Session 32, AI and Judges. We have discussed in this course the various applications in AI, particularly how they have been used in different courts in different judicial systems, but when it comes to the judiciary, particularly judges, AI has some specific implications and impacts. It is not that AI is going to totally replace the judges, nor is it that the AI will one day be the one that will overwrite the judges. No way. But we need to understand how judges can understand AI and what sort of precautions they need to take to use AI effectively and efficiently. And more importantly, what are the pros and cons of reliance on AI in the justice or judicial system for judges?



## Recap



- The last session on overview of current scenario and emerging picture of use of AI in different countries was given.
- We pointed out while some applications are futuristic the hardcore issues cannot be wished away.



So before getting into details, let us look at a recap of the previous session. The previous session was on the current scenario and the emerging picture of the use of AI in different countries in different contexts. We did point out that there were some applications that are futuristic in some countries, like AI judges, AI advocates, and extensive use of AI as part of the multiverse environment. But there are some hardcore issues that cannot be simply wished away. Or in other words, while the future looks quite bright and shining, particularly when it comes to the application of AI in law and justice in many countries, there are some hardcore issues that are almost common in every jurisdiction, and they need to be addressed; they need to be, first of all, overcome if you really want to harness AI fully.



## Overview of AI in Judicial Systems



- **Advantages of AI in Justice**
  - Reduces administrative burden and backlog
  - Increases court efficiency and reduces costs
  - Makes justice more accessible
- **Fairness and Accuracy**
  - Strictly follows precedents
  - Prevents personal biases and preferences
  - Handles large amounts of information
- **AI in Courtroom Functions**
  - Auxiliary administrative functions
- **AI Judicial Tools**
- **Fully Automated Judicial Decision-Making**
- **This session addresses selected issues in using AI Particularly generative AI as a tool for Judges**



So let us get into the topic now. The overview of AI and the judicial system is too well known. I am not going to fully repeat it, such as what the advantages are, what we mean by fairness and accuracy, why we need AI in the courtrooms, judicial tools, fully automated judicial decision-making, this we have discussed in detail earlier.



## Human Judgement in Judicial Decision-Making



- **Human Judgment in Judicial Decisions**
  - Example: King Solomon's judgement to reveal the true mother
  - Based on emotional intelligence and credibility assessment
- **Legal Judgements**
  - Based on evidence and rules
  - Involves practical and reflective judgement
- **AI in Judicial Decision-Making**
  - Discriminative AI for determinative judgement
  - Generative AI for practical judgement
- **Discriminative AI**
  - Trained to distinguish categories (e.g., spam detection)
- **Generative AI**



So, what is human judgment in judicial decision-making? This is important. Because human judgment is not simply driven by reason, dry logic, or dry reasoning. Human judgment is also driven, if not fully, by emotion, empathy, sympathy, and understanding, and by putting things in context. For example, King Solomon's judgment to identify the real mother of the baby was that the true mother of the baby is someone who has

emotions. So, it is based on emotional intelligence and credibility at times. But legal judgments are not giving much importance to emotional intelligence, using them to assess cases. We are looking more in terms of evidence and rules. And then this also involves practical and reflective judgment, in the sense that evidence and rules are there to write a judgment, but then the evidence and rules themselves cannot be translated into a judgment. So, we need practical thinking, practical application, and a reflective judgment that goes back and forth or that addresses reality and then comes with a judgment. The problem with AI we have seen is that discriminative AI could be an issue here, and generative AI can be used for practical judgment. Now, the discriminative AI is a problem; we have also seen that it is trained to distinguish categories like spam detection. Whereas there are a lot of tools available for generative AI, more tools are likely to come in the future.

The slide features a header with the NPTEL logo on the left and the NALSAR logo on the right. The title "Predictive analytics models" is centered in red. Below the title, there are two main bullet points: "Predictive Analytics Models" and "Prescriptive AI Models", each with a sub-list of characteristics and examples. To the right of the text is an illustration of a person in a white lab coat pointing at a large screen displaying data charts, with another person in a purple dress looking at the screen. Below the illustration is the text "Advance Analytics". In the bottom right corner, there is a small video inset of a man in a dark blue shirt speaking.

## Predictive analytics models

- **Predictive Analytics Models**
  - Utilize statistical algorithms and machine learning on historical data
  - Identify patterns and make projections
  - Examples: credit score generation, fraud detection
  - Assign scores to predict future events
- **Prescriptive AI Models**
  - Suggest the best course of action for desired outcomes
  - Integrate predictive analytics outputs
  - Employ AI processes, optimization algorithms, and expert systems
  - Examples: personalized healthcare plans, dynamic airline ticket pricing

Advance Analytics

One of the most significant applications we have seen for judges in law and justice is that, using predictive analytics with statistical algorithms and machine learning, it is possible to identify patterns and then make projections. Like a credit score, fraud detection, assigning a score to predict future events will help judges find it useful in cases where they need to grant parole, set bail, review bail conditions, review parole conditions, or address new demands presented by individuals; thus, one is to predict and the other is to prescribe.

The prescriptive AI models, in fact, give us lots of choices; they will be able to tell us, "This is the best course available to you, Mr. Judge or Madam Judge," on the basis of the desired outcomes, and then this can be integrated into predictive analytical output. The best mode here is to find that there are four options open to you in the case of this bail. The first option is to give unconditional bail; the second option is to give conditional bail

subject to these conditions imposed on it; the third option is to neither grant conditional bail nor deny bail, but rather refer it to a different bench due to some peculiarities of this case or other factors; the fourth option could be that the petition should be dismissed because the negative aspects or factors outweigh the ones that favour the grant of bail. So, simplistically speaking, among the four options, the judge can decide, but the AI model can give a lot of prescriptions like this; this can be integrated into a predictive model so that you can describe if this is the predictive model, what the outcomes are, and then what the best outcome possible is.

So, the prescriptive AI models are available particularly in diagnostics, personal health care, etc. Particularly in airline ticket pricing, where it fluctuates from time to time, from minute to minute, based upon your choices and the least time-consuming route for you, or the ones that have the most stopovers, or the way you want to structure it in the sense that you want to go to a city, spend some time outside, and then come back. So, you would like to prepare at least a gap of four or five hours between the flight that lands and the flight you want to take off, or the flight you want to take further for the next location. So, it is possible to use a prescriptive AI model to do a lot of things in the sense that this model will tell you, among all the flight options, which is the best option available to you. Therefore, predictive analytics and prescriptive analytics are two important analytics models.

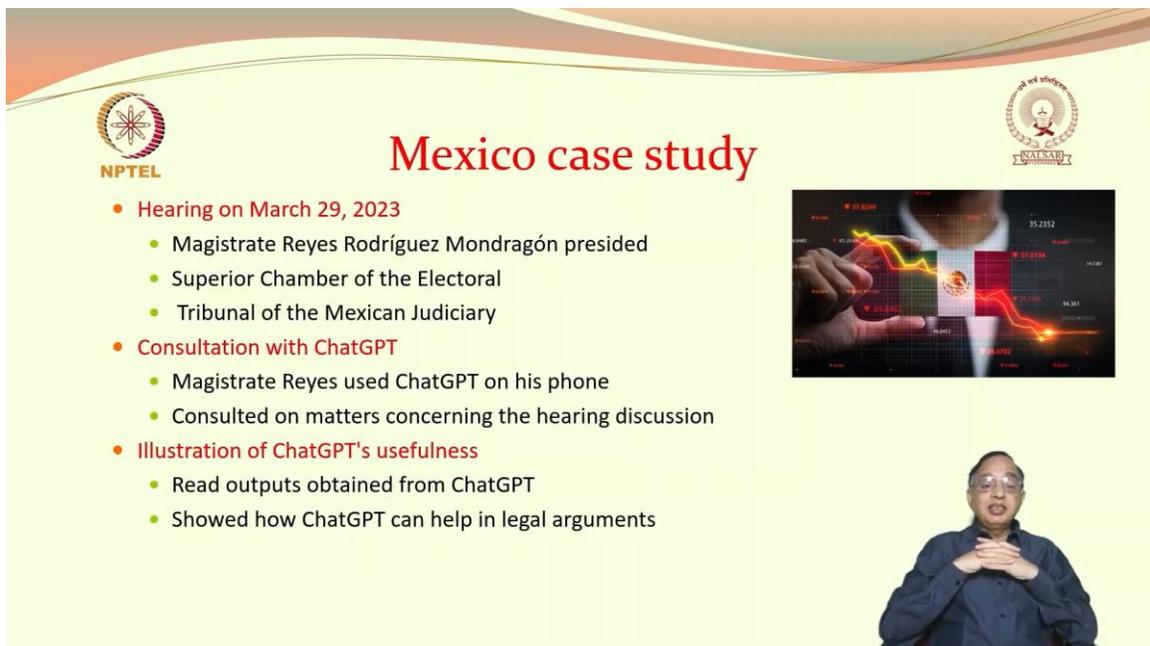
The slide features a yellow background with a decorative orange and white wave at the top. On the left is the NPTEL logo, and on the right is the NALSAR logo. The title 'Generative AI models' is centered in red. A list of six bullet points is on the left, and a small image of a person's head with a digital brain is on the right. A video inset of a speaker is in the bottom right corner.

## Generative AI models

- **Predictive AI in Judicial Decision-Making**
  - Suggests likely outcomes based on past cases
- **Prescriptive AI for Systemic Optimization**
  - Optimizes resource management for case resolution
- **GenAI's Role in Legal Research**
  - Drafts arguments
  - Identifies relevant statutes
  - Summarizes complex documents
- **Influence on Judges**
  - Aids in research and understanding of law
  - Impacts application of legal principles
- **Opportunities and Challenges**

Predictive analysis in judicial matters will involve outcomes based on past cases; this is actually something very similar to what people normally do. But it can also optimize resource management for case resolution. And then GenAI, as we have seen, can be part of legal research; it can draft arguments, identify limiting statutes, and summarize complex documents for judges. It will help judges understand research and identify flaws

by assisting in the research. Furthermore, it can also help them understand the application of various legal principles. So, it can directly or indirectly influence the judges. But the problem here would be that if that influence prolongs or continues for a long time, judges may think that it is better to rely on AI-based systems than to scrutinize it or to think that whatever AI gives us as an output need not be cross-checked once again. So, there are opportunities and challenges in both areas. The opportunity is that it will be helpful for judges who are pressed for time to go through too many documents and then come to a conclusion, or who do not have sufficient support to conduct good legal research and then arrive at some preliminary findings. The challenges we have mentioned are the problems with the use of AI in law and justice and how to make the systems more responsible, ethically aligned, and aligned with the values of the institution.



The slide features a light green background with a decorative orange and white wave at the top. On the left is the NPTEL logo, and on the right is the logo of the Superior Chamber of the Electoral Tribunal of the Mexican Judiciary. The title "Mexico case study" is centered in red. A bulleted list on the left details the case. To the right of the list is a small image of a hand holding a smartphone displaying a financial chart. At the bottom right is a video inset of a man in a dark blue shirt speaking.

## Mexico case study

- **Hearing on March 29, 2023**
  - Magistrate Reyes Rodríguez Mondragón presided
  - Superior Chamber of the Electoral
  - Tribunal of the Mexican Judiciary
- **Consultation with ChatGPT**
  - Magistrate Reyes used ChatGPT on his phone
  - Consulted on matters concerning the hearing discussion
- **Illustration of ChatGPT's usefulness**
  - Read outputs obtained from ChatGPT
  - Showed how ChatGPT can help in legal arguments

In Mexico, in March 2023, Magistrate Reyes Rodriguez presided over the Superior Chamber of the Electoral Tribunal of the Mexican Judiciary. He used ChatGPT on his phone. He showed that. So, he consulted ChatGPT regarding matters related to the discussion. So, what he did was read the outputs from the mobile phone and then said this is what ChatGPT said, and then he also showed how ChatGPT can help in legal arguments. So, this is one way for judges to directly use ChatGPT without any mandatory requirements; they can use it as they wish.



## India case study



- **Judge Anoop Chitkara's Bail Decision**
  - Refused bail to a man accused of serious crimes
  - Consulted ChatGPT for guidance
- **Purpose of Using ChatGPT**
  - Ensure impartiality in decision-making
  - Balance personal bias with AI input
- **Legal Basis and Consistency**
  - Judge's consistent view from past cases
  - High level of cruelty allegations considered



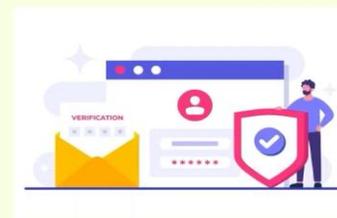
In India, Justice Anoop Chitkara refused to grant bail to a man accused of serious crimes, and then he consulted ChatGPT for guidance. The purpose here is to ensure impartiality in decision-making and to balance personal bias with AI input. Additionally, this AI's consistent view from past cases was also taken into account. There was also the problem of a high level of cruelty allegations against the person who had applied for bail. So, taking these factors into account, Anoop Chitkara decided not to grant bail, and he used ChatGPT to arrive at his decision. But the problem here is that ChatGPT can be consulted; however, ChatGPT cannot be the main reason to deny or grant bail.



## Ex-ante verification process



- **Ex-Ante Verification Process**
  - Guarantees GenAI meets basic standards
  - Applies to third-party and in-house developed GenAI
- **Relevance to Judicial Use**
  - Particularly important for GenAI due to extra-legal data sources
- **Licensing and Verification Regime**
  - Assesses GenAI systems for functionality and legal adherence
  - Minimizes potential risks in judicial processes
- **Integration Standards**
  - GenAI must meet high standards for judicial decision-making



There can be some sort of mechanism to identify what happens when we use general AI

tools to predict an outcome or to give judgments, and then to what extent they have been truly truthful or whether they have some impact on the person after the judgment. So, the ex-ante verification process can be done for AI-generated tools or gen AI tools in law and justice. You can apply third-party in-house developed generative AI tools. It guarantees basic standards in the sense that ex ante verification is that this is the bail given, this is the judgment rendered, or this is the parole granted. What happens after that?

Ex ante verification can be given in the sense that if a judge using the AI tools granted a hundred bails, how many of the persons who went out on bail broke the bail conditions, did not adhere to the bail conditions, committed a crime, or tried to commit a crime? Similarly, behaviours can be monitored by general AI tools in the sense that we can build a database using this ex-ante verification, which will further contribute to legal training as well as to deciding on future courses. It is important because, due to extra-legal data sources, we can build general AI systems in a much better way, but the general AI systems need to be really assessed for functionality and legal adherence. The legal adherence means that they should be legally coherent; they should not give inconsistent answers, or they should rely on the right precedent, and then they should use the right code or law for giving judgments or at least helping the judges. We should also identify the minimum risk that could happen and then try to minimize it. So, when it comes to GenAI, the judicial decision-making bar should be set so high that it is as good as that of a human being. So, this ex-ante verification process and other steps will help us ensure that these tools become more effective.

The slide features a title 'Algorithmic fairness and bias mitigation' in red text. On the left, there is an NPTEL logo. On the right, there is a logo for the National Judicial Academy (NJA) of India. Below the title, there is a bulleted list of three main categories, each with sub-bullets. To the right of the list is an image of a hand holding a tablet displaying a network diagram. At the bottom right, there is a video inset of a man in a blue shirt speaking.

## Algorithmic fairness and bias mitigation

- **Verification Process for Algorithms**
  - Scrutinize algorithms for potential biases
  - Ensure implementation of fairness testing and de-biasing methods
- **Data Utilization in Risk Assessment**
  - Data should be representative of diverse populations
  - Avoid perpetuating historical biases
- **Specific Techniques to Minimize Bias**
  - Avoid algorithmically integrating biometrics in judicial decisions
  - Enhance unbiased data
  - Preserve data privacy

So, we have talked about algorithmic fairness and bias mitigation; we have talked about the need for understanding and then looking at the algorithms, how they work, and what the problems are with them. We need to identify the ways to minimize bias, how

unbiased data can be used, and then the diverse population data. When we integrate algorithmic fairness and bias mitigation in a Gen AI tool, it is equally important to look at the precautions we have discussed in detail in the previous classes.



## Oversight by independent body

- **Oversight by Qualified Independent Body**
  - Expertise in AI, law, and ethics required
- **Demonstration of Standards Compliance**
  - Developers must show Gen AI systems meet standards
- **Pre-Usage Requirement**
  - Standards must be met before use in courtrooms



But when we develop such Gen-AI-based systems for judges, they should be overseen by an independent body with expertise in AI law and justice ethics, as we need to first establish the standards. As of now, there is no ISO standard; there is no standard for any judicial system using AI for writing or delivering judgments. But then the time has come for the system to meet minimum standards, which could be technical and can evolve over a period of time. More importantly, the standards must be met before they are put to use. Or it cannot be a criterion that the standards will be met two months after the system is put to use. And even if the standards are met two months after the system is put to use, the adherence to standard norms should come first before the system is tested, even under trial conditions.



## Internal verification for in-house Gen AI



- **Internal Verification for In-House Gen AI Models**
  - Courts developing their own Gen AI Models should implement rigorous internal verification processes
  - These processes should mirror the external licensing regime proposed in the article
- **Adherence to High Standards**
  - Internal verification ensures in-house Gen AI systems meet high foundational standards
  - Standards should be equivalent to those of third-party developed systems



One way to look at it is that Gen-AI models and then Gen-AI systems can be developed in-house by the judicial authorities or by the judicial system itself. So that the internal rigorous verification process is not compromised, the judicial authorities or the judicial system itself knows where the shoe will pinch or where the problems could occur. So, they may be the ones who could easily identify the problems and potential issues first, and then identify them. These processes should mirror the external licensing regime that have been proposed in the article cited; the literature is given there. And then the internal verification should ensure that it meets the high foundational standards for AI use in justice. Standards should be equivalent to those of third-party developed systems or put in another way, there should not be discrimination or bias in favour of in-house developed Gen AI models, nor should the same standards apply to in-house models as well as to those developed by third parties.



## Importance of high-quality datasets



- **Importance of High-Quality Datasets**
  - Gen AI system's potential is determined by the quality of its training data
  - Biased or incomplete data can perpetuate societal biases
- **Impact on Fairness and Accuracy**
  - Inaccurate legal reasoning can undermine justice system cornerstones
- **Considerations for Integration**
  - Factors to be considered when integrating Gen AI into courts





Then, of course, we know high-quality datasets, and this we have discussed even earlier as well.



## Data quality monitoring and strategies



- **Importance of Data Quality**
  - Gen AI systems rely on high-quality data
  - Data must be accurate, unbiased, representative, and complete
- **Data Quality Monitoring and Mitigation**
  - Strategies to monitor and mitigate biases
  - Historical tracking of court hierarchy and legal developments
  - Weighted analytics on legal interpretation progressions
- **Strategies for Ensuring Data Quality**
  - Employing diverse data sources
  - Using procedural and material fairness metrics
- **Closed-Network Datasets**
- **Success of Gen AI in Judicial Decision-Making**





Then data quality needs to be monitored on a continuous basis because data has to be unbiased, persuasive, and, more importantly, accurate. So how do we do that? Data quality monitoring and mitigation have to be ongoing regular processes, and if the system is inbuilt, and if the system is also trained on account of this inbuilt data rather than relying on third-party data or third-party systems, there should be a mechanism as part of the whole system development, upkeep, and maintenance that data quality monitoring and maintaining the quality should be done on a regular ongoing basis. So, the success of

GenAI in judicial decision-making depends on many of these factors, and even a small glitch or issue in them will affect the overall output.

 **Data access, Explainability, and origin visibility** 

- **Data Access Protocols**
  - Clear protocols for data access are essential for transparency and fairness
  - Judges and relevant parties should understand the data sources used by Gen AI systems
- **Explainability of Gen AI Outputs**
  - Explainability helps judges and litigants understand the reasoning behind Gen AI outputs
  - Fosters trust and transparency in decision-making
- **Origin Visibility of Data**
  - Gen AI systems should disclose data sources and characteristics
  - Includes precedents, laws, and regulations used in conclusions
- **Data Summaries and Anonymized Access**
  - Provide summaries of data used
- **Audits and Verification**



Then, of course, we have the data access protocols. Clear protocols are essential for transparency and fairness because when data is used for judicial decision-making or for assistance, we need to look into the higher criteria; we need to look into the higher standards and then see whether they are able to meet them. That is more important. So, this should again be part of the regular evaluation process.

 **Developer accountability** 

- **Developer Accountability**
  - Developers should be accountable for algorithmic design and functionality
  - Clear disclosures of AI's modeling and reasoning process are required
- **Liability Allocation**
  - Avoid overly punitive developer liability frameworks
  - Balance responsibility to avoid high barriers to entry
- **Certification Process**
  - Ex-ante certification process to balance incentives and safety
  - Certification or licensing process to assess development practices



Then the developers' accountability should be made part of the system where, in addition to meeting the standards set, developers should be accountable for the algorithmic system

and functionality, which they should demonstrate without any doubt. And then clear disclosure of the modelling and reasoning process is required because this is for two purposes: one, for the explainable AI purposes; two, in case of any dispute in the future and if the judgment is challenged or the use of the AI system is being challenged, the court, institution, or judge will be in a better position to counter those challenges. So, the system should be designed in a way that the overly punitive developer liability framework is avoided; developers should be punished or made to pay the price for the problems, but it should not be too harsh so that it becomes a disincentive. We need to look at the barriers to entry, the level of innovation, and then we need to consider the responsibility. More importantly, ex-ante certifications should also be done so that the system's evaluation can be conducted after it has been put to use for about three to four months or even a longer period, ensuring that the pre-trial verification or the pre-trial evaluation norms have been met by the system. And then the certification or licensing process to assess the development process is important so that the understanding of the development process remains well within the judiciary or the justice system itself.



The slide features a header with the NPTEL logo on the left and the NALSAR logo on the right. The title "Judicial discretion and training" is centered in a large, bold, red font. Below the title, there are four main bullet points, each with a red circular icon and a list of sub-points:

- **Judicial Discretion and Responsibility**
  - Judges retain ultimate decision-making authority
  - Absolute responsibility for Gen AI outcomes could discourage use
- **Training and Resources**
  - Comprehensive training for judges to evaluate Gen AI outputs
  - Understanding limitations of Gen AI
- **Liability Distribution**
  - Tiered approach integrating training and risk-based responsibility
  - High-risk cases require rigorous review
  - Low-risk cases require lesser scrutiny
- **Review Process and Scrutiny Obligations**
  - Higher degree of responsibility for judges in high-risk cases

There are two images on the slide: one showing a judge in a black robe standing at a desk in a courtroom, and another showing a man in a blue shirt speaking, likely the presenter.

Finally, the ultimate authority or the decision-making should be left to the judges. So, they should be the ones who take responsibility and decide whether they should use the outputs or not. More importantly, they need comprehensive training for such outputs, and then they need to understand the limitations of Gen AI. In other words, they need to understand what could go wrong, what might go wrong, and what the potential is for the system not to provide sufficient output of the desired quality. So, the liability here needs to be balanced or distributed. High-risk cases will require a rigorous review for obvious reasons, particularly in criminal justice. Lower cases will require less scrutiny. Lower cases could be the ones for traffic violations, minor penalties, or minor fines, or where you substitute imprisonment for a person by paying a fine, or you substitute a punishment

for a person by paying an extra fine. So, the responsibility of judges in high-stakes cases will always be higher. So, the systems have to be dealt with in such a way that they meet the higher standards set by the judiciary on its own as well as for the systems.



## Shared responsibility model



- **Proposed Shared Responsibility Models**
  - Developers and court system share liability
  - Depends on case circumstances and algorithmic output
- **Factors Influencing Liability**
  - Judges' Gen AI training
  - Complexity and risk profile of cases
  - Transparency of Gen AI system
- **Goals of the Framework**
  - Encourage responsible development and Gen AI integration
  - Empower judges and increase efficiency
  - Protect integrity of judicial decision-making



One way is to develop a shared responsibility model. The shared responsibility model is presented as an idea so that the responsibility does not unduly lie on the shoulders of a single party. When this happens and a system is being developed, co-developed by the developer as well as the justice system, this is possible. So, depending on the cases in the sense of whether they are civil, criminal, labour, or environmental law, and then the need for the algorithm and the algorithmic output, this could be decided. So, the liability will again vary depending on the judges' training and understanding, the complexity of the issues involved, and then be transferred to the system. More importantly, the framework should encourage responsible development and deployment of general AI integration systems. It should empower judges, improve efficiency, and protect and promote the integrity of the judicial system. On no grounds should the integrity of the judicial system be questioned by the system, nor should the system emerge as a challenge to the integrity of the judicial system. So, the shared responsibility model will be one where the developers are initially very reluctant to come and develop models due to the higher liability and their unfamiliarity with the way systems work.



## Human-designed prompts and expertise



- **Role of Human-Designed Prompts**
  - Used by all judges in case studies
  - Assume human expertise in prompt creation
- **Effectiveness of Human Prompts**
  - Depend on human's expertise
  - Effective for tasks with well-defined input and structured output
- **Expertise Required for Designing Prompts**
  - Understanding of specific field and human-machine interaction
  - Competency and experience vary across jurisdictions



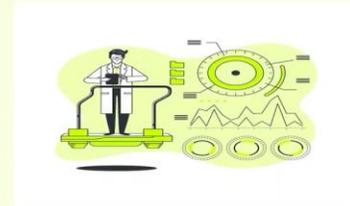
One way to do that is that, as we all know, AI systems need training. We need people to be trained in how they can use prompts and how prompts can influence the decision-making power of AI systems. So, judges can be trained, and judicial authorities can be trained in prompt creation and prompt management on how to create effective prompts that will be the right ones to deal with the system. More importantly, well-defined input and structured output are required, so the set of precise prompts that should go in and the sort of questions that should be brought in should also be part of the training program. In other words, we need to build capacity around human-designed prompts and expertise within the judiciary.



## Challenges and automated prompt design



- **Quality of Human-Designed Prompts**
  - Ensuring quality is challenging
  - Affected parties cannot assess quality to challenge decisions
- **Limitations of Human-Designed Prompts**
  - Researchers exploring automated prompt design
  - Improves efficiency and adaptability
- **Automated Prompt Design**
  - Generated using various algorithms and techniques
  - Already deployed in sectors like the medical field
- **Future Exploration**
  - Choice of prompt structure depends on task and resources
  - Needs further exploration in judicial systems and Gen AI



Of course, there are some challenges which are mentioned here. The quality of the output and the quality of the system itself will be challenging. And then human design prompts have their own limitations because they come with the human's understanding of the issue. Whereas AI's understanding, training modules, and self-learning practices could be totally different. For example, my prompt, which I input to the system, can be trivial, or the system may misunderstand it as my thinking and then may come up with a totally unanticipated output that could be as good as junk. So, the researchers exploring automated prompt design also try to tell us that automated prompt design will give the judges the option to decide whether they should use an automated prompt or provide the prompt themselves. Automated prompts are something like an available option that is pre-tested depending on the algorithms and techniques, and that will also be useful for people to decide. For example, an automated prompt for a radiologist will be one that has already been tested in the medical field and that will also have an understanding of what exactly the radiologist will be asking for. So, this prompt engineering and the prompt structure dependence will need to be looked into in detail.



# Risk-based deployment and accuracy thresholds



- **High-Risk Use Cases**
  - Significant potential consequences for defendant's rights
  - Examples: criminal sentencing, loss of liberty
- **Stricter Accuracy Thresholds**
  - Mandated for high-risk cases
  - Based on metrics relevant to the legal domain
- **Higher Responsibility in Reviewing Process**
  - Enhanced scrutiny within the judicial system
  - Ensures accuracy and fairness
- **Scope of GenAI Application**
  - Exclusion from deployment in high-risk cases
  - Limited application: reviewing case law, weighing evidence



And then risk-based deployment comes with what are the defendant's rights; for example, as of now, the judicial system is accountable, but then I cannot go and sue a judge or the judicial system for a judgment that is given wrong, but I can appeal or I can seek a review. But with the integration of AI-based systems in judicial decision-making, the potential for these systems to present risks or to curb the defendant's rights or to not oppose the defendant's rights or restrict them is also possible. Particularly in criminal sentencing, this can happen in laws of liberty, in laws of fundamental rights, in laws of financial freedom, and so on. So, the potential consequences for the defendant's rights should be looked into. Otherwise, the defendant can say that the AI system that was used was biased, and then the judicial system that used such a system is also equally accountable, and the burden cannot be shifted back to the AI system.

So, the accuracy threshold should be set at a higher level, and then the metrics relevant to the legal domain should be identified and put there. So, the higher responsibility in reviewing assessments is also part of that. So, one way to look at it is that you shouldn't use GenAI-based applications where the risks are high. For example, it is not advisable to use GenAI-based systems in criminal sentencing for cases involving the relevant sections, for life term sentences, for the death penalty, for very serious judgments, or where the offense is very serious. Similarly, in civil cases, it could also be argued that the general AI-based system should not be assigned cases where there are a lot of emotional factors involved, or the factors are too complex, or the judgment involves a better human understanding of the circumstances than the systematic, typically data-based approach. So, depending on the risks and the category of the system cases involved, whether to use the GenAI system or not should be left to the judge's discretion.



# Graduated approach to review and responsibility



- **Review and Scrutiny Levels**
  - Low-risk cases: Basic review by the judicial system
  - High-risk cases: Thorough examination by judges, clerks, or specific organizations
- **Increased Review Requirements**
  - Legal professionals delve deeper into GenAI's reasoning process
  - Examine data used, such as case precedents, laws, and extra-legal information
  - Identify potential biases
- **Heightened Responsibility**
  - Judges: Evaluate GenAI outputs and ensure alignment with legal reasoning
  - Developers: Accountable for algorithmic design and functionality



So, we need to look at things in a graduated approach; we need to move from low-risk cases where AI systems are well trained and well adapted, and then try to go further, progressing from low risk to medium risk, and then try them later in the high-risk category. This does not mean that they should not be put to use at all; rather, they should not be the ones that decide the judgment. And then we need to have review mechanisms that are internal, and we need to examine the data use; we need to assess the output from the system or service, how a human would have understood and then delivered a judgment, so these sorts of evaluations should go on in a graduated approach on a regular basis. Judges cannot evade responsibility even if the AI systems provide the right judgment with wrong reasoning, fake cases, or hallucinations; therefore, the judges' responsibility and liability must be seriously taken into account. The same applies to developers, so the heightened responsibility also comes with the need for heightened accuracy and increased responsibility.



## Disclosure to parties

- **Transparency in Judicial Processes**
  - Mandatory disclosure of Gen AI use to all relevant parties
  - Ensures fairness and accountability
  - Promotes trust in the judicial system
- **Key Purposes of Disclosure**
  - Informs all parties involved about the use of Gen AI
  - Allows for scrutiny and evaluation of Gen AI's role facilitates informed decision-making



Then the disclosure to the parties is equally important because the parties have the right to know to what level the AI was used and to what level AI was used to write or give the judgment, and all the stakeholders—whether the defendant, the plaintiff, the state, or the appellant—should have the right to know exactly what the level of AI usage is and what the AI actually did. It is possible that some defendants or even the government may say that in a high-risk or high-profile case, we will not go for an AI-based system on account of various factors, including either a lack of confidence or the way people think that AI systems may go wrong; so they may say that we want to opt out of the AI-based systems and would use only human-based systems. In such cases, prior disclosure will help them decide on it. More importantly, the system should keep this option open at least in some cases where the stakes are too high, in the sense that it involves electoral disqualification or situations where a person's life and death are at stake, or cases where a person's liberties are being taken away by the court or by the system on account of some serious crime. So, this disclosure to parties should be made mandatory, should be compulsory, and then the disclosure should be adequate enough for the parties to make a decision.

 **Empowering parties and ensuring fairness** 

- **Transparency in Legal Proceedings**
  - Parties are informed about Gen AI involvement
  - Allows for informed decision-making regarding legal strategies
- **Options for Challenging Gen AI Output**
  - Parties can choose to challenge the Gen AI output
  - Request for traditional review process if needed
- **Promoting Fairness**
  - Ensures all parties understand the tools being used
  - Enables parties to shape their participation accordingly



And then we need to really empower the parties; we need to ensure that there is fairness in the system.

 **Promoting trust and procedural awareness** 

- **Transparency in Judicial Process**
  - Builds trust among parties involved
  - Ensures awareness of procedural status
- **Use of Gen AI**
  - Informs parties of their rights
  - Boosts confidence in due process



More importantly, there is a need to promote trust and personal awareness so that everyone has a better understanding of the systems.



## Methods and content of disclosure



- **Tailored Disclosure Methods**
  - Adapted to the risk profile of the case
  - Based on the type of Gen AI used
  - Consideration of court's means and resources
- **Due Inclusions and Potential Approaches**
  - Specific methods can vary
  - Approaches depend on case specifics



So, the tailored-made disclosure can be brought in, and then depending on the type of AI, the disclosure norms can vary; and then, depending on the court's means, objectives, and resources, this can go along; this can be done. So, there should be a tailor-made list of inclusions, what sort of inclusions are mandated at different stages, who should be informed, and what should be included should also be made part of the decision-making process. So, the specific methods may vary. So, when it comes to some specific cases, high-profile cases, or high-risk cases, the judgment and AI system that we run in a mode of how it will predict can come up with three or four different judgments from the same system. So, systems can come up with different judgments based upon different parameters and different understanding. It will be ideal; it will be the better approach to understand those things first and then to really decide whether the system should be used for it. Because if the system comes up with four or five judgments that are mutually contradictory or applies the same logic but reaches different misleading interpretations of case law, then the whole judiciary's credibility will be at stake and will be questioned. So, in such cases, there should be due inclusion, and there should be a way to opt out of using AI.



## Scrutiny tailored to risk



- **Scrutiny Tailored to Risk**
  - High-risk cases require rigorous verification
  - Involves criminal sentencing or significant impacts on rights and freedoms
  - Independent verification by court personnel or qualified third parties
  - Cross-referencing data sources and verifying legal citations
  - Ensuring factual accuracy
- **Low-risk cases**
  - Reduced review process
  - Tailored to the needs and rights affected by each specific process



And of course, risk evaluation has to be done; this we have discussed earlier.



## Gen AI as an enhancer, not a bottleneck



- **Gen AI as an Efficiency Enhancer**
  - Verification should not slow down the judicial system
  - Goal is to enhance efficiency while mitigating risks
- **Streamlined Verification Procedures**
  - Utilize technology-assisted verification tools
  - Achieve balance between efficiency and risk mitigation



But we should understand GenAI as an enhancer, not as a bottleneck. And an enhancer means we should try to see that how it can enhance efficiency while mitigating risk. And then, technology-assisted verification tools should be made part and parcel of the system. So, we need to achieve a balance between efficiency and risk mitigation. So, the use of Gen-AI as a tool for judges comes with a lot of benefits and then a lot of risks. At the same time, the benefits outweigh the risks.



## Literature (Selected)



- David Uriel Socol de la Osa and Nydia Remolina (2024) Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision making through generative AI *Data & Policy* (2024), 6: e59 doi:10.1017/dap.2024.53
- Vasiliy A. Laptev · Daria R. Feyzrakhmanova (2024) Application of Artificial Intelligence in Justice: Current Trends and Future Prospects *Human-Centric Intelligent Systems* (2024) 4:394–405
- <https://doi.org/10.1007/s44230-024-00074-2>
- Australian Institute of Judicial Administration (2023) *AI Decision-Making and the Courts : A guide for Judges, Tribunal Members and Court Administrators*
- Felicity Bell and Michael Legg (2025) *Judicial Impartiality: AI in Courts Cambridge Handbook of AI in Courts* (forthcoming)



So, I have given some explanations, I have provided some ideas in it, but the field is evolving and rapidly developing. So, the literature before you, which I have cited, takes into consideration some of the points and also dwells on some of the things. More importantly, some of the future prospects discussed in the literature may not be easily feasible or may take years to materialize, and equally important is that different countries have different judicial norms for different expectations of judges; furthermore, different countries deal with judicial ethics, procedural fairness, process fairness, and due process in law in different ways, so one cannot come up with a general or very conditional answer that is applicable to all circumstances. So, we need to go on a case-by-case basis regarding how the system understands the judiciary in that country or in a specific context. So, the very idea of judicial impartiality may vary from country to country. In some countries, the judicial impartiality may be of such a high standard that judges should be beyond any doubt. In some countries, judicial impartiality may not be of such a high standard that judges themselves could be debated or impartial; people think that impartiality is not as important as efficiency. So, the literature I have cited does address some of the issues, but then, as we have said, we are only scratching the surface of the issues. We need to go beyond; we need to understand more. So, with that, in this session, I stop. Thank you.



## Next



- AI and Human Rights (Session 33)



The next session is going to be on AI and human rights; that is session 33. Session 33 will look into the implications for AI and human rights, and then we will also look into some of the emerging issues in AI and human rights.