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 – 31

Artificial Intelligence, Law, and Justice Course, Session 31: AI in Law and Justice in the USA, Some Examples and Points.

Let us recap the earlier session. In the last session, we discussed the implementation of AI in law and justice in selected countries and gave some examples. We highlighted that while the experience is uneven, some of the issues are common. Moreover, we said that

in many countries, different institutions, including Bar Associations and Departments of Justice, are developing guidelines or coming out with some guidelines and reference materials. Specifically, we mentioned that in this session we will be looking into the USA as a case study and then identifying some interesting examples and particular points from that.

So, let us move into that now. The USA, as you all know, is a federal system where there are states, and then within states, there are counties. So, the counties have their own legal system, judicial system. One of Palm Beach County's digital innovations is considered to be a very pioneering one. They have deployed AI-driven tools in courts and clerk's offices over the past five years which reduced the inefficiency and errors in document processing. They came up with a light-out document processing program. It analyses tags and then indexes document filings. So, this enables improvement of the court and other mechanisms easily. So, they started with a 2018 pilot program first, then they trained hundreds of documents on that, and finally, they audited all the documents that had already been trained and organized by the software.

Efficiency and Accuracy Improvements

- High Accuracy Rate
  - Machines achieved 98% to 99% accuracy
  - Significantly higher than human counterparts
- Workload Capacity
  - Five robotic systems equaled the workload of 19 human employees
- Employee Benefits
  - Freed up workers for more thoughtful jobs
  - Enabled skillset growth and increased earning potential

So, this resulted in machine achieving 98% to 99% accuracy much better than human counterparts. So, what they did was use 5 robotic systems or AI systems that equated the workload of 19 human employees, and this resulted in the human employees spending more time on thoughtful work rather than the routine mundane tasks that they had to do repeatedly for hours. And this enabled the growth of skill sets and increased earning potential for those employees.

Public Defender Offices: Enhancing Legal Advocacy

- Miami-Dade County Public Defender's Office
  - Advocates for AI tools in legal document drafting and research
  - First in the US to use AI for attorneys and teams
- Los Angeles County Public Defender's Office
  - Integrated AI solutions into their toolkit
  - Migrated 24 legacy systems to cloud-based platforms

And then in the USA, there are public offices called public defender's offices. Who is the public defender? A public defender is an advocate or a legal professional appointed by the court or by the government to defend a person who otherwise lacks the means to defend himself or herself or lacks access to the legal system. So public interest is taken care of by appointing a public defender who otherwise would be serving the cause of the

people, those who would have been left without anyone to defend their case or to seek justice for them. So, the public defender offices are also using AI tools in the USA. For example, the Miami-Dade County Public Defender's Office advocates for AI tools in legal document drafting and research, and it is the first to use AI for attorneys and teams. Often the public defender is a single person; as such, the designation may be misleading, but they work in tandem; they work as teams. So, the office of the public defender will have various public defenders working as individuals or as teams to ensure that those who are in prison or those who are seeking legal justice but are otherwise unable to pay for attorney's fees or unable to reach out to the legal system can use their services.

In the Los Angeles County Public Defender's Office, they integrated all the AI tools into a single toolkit and then migrated 24 legal systems to cloud-based platforms. What are the 24 legal systems? See, we have computers of different generations and different capacities available. Today's computers are much more sophisticated and much more powerful in terms of processing speed and other parameters than those that were available 10 years or even 5 years earlier. But what happens in many institutions, as computers are acquired over a period of time and used for different purposes, is that some of the systems may not be up to date or may be on the verge of becoming obsolete. Lots and lots of data get stored in different categories of such systems. Some of them may be considered legacy systems or systems that are not of the current generation or that are on the verge of becoming obsolete. So, this public defender's office migrated 24 of such systems into cloud-based platforms. Now the cloud-based platforms make things much more easily accessible and much easier to work with as well. So, shifting from the legacy systems, which are otherwise obsolete, to cloud-based platforms will enable better utilization of AI in such services.

So, the idea here is that technology's role should be to improve the case outcome, which means that people should not unnecessarily get punished or land in jail for no fault of their own. And then the person who is to be defended, or who has to be defended, has to be treated with care, so we need to humanize them. And then, to do that effectively, in addition to legal expertise, we need lots and lots of data, so a public defender by choice

will have to really use troves of data for that purpose. Again, in this case, what they did was organize the data by person's name, not by case, and then they also looked at the data in human narratives, and then they suggested alternative treatments or alternative punishments rather than simply sending them to jail or sending them back to jail. So, the idea of the public defender's office in such cases is to reduce the prison sentence or at least reduce the term of the prison sentence if it has already been awarded.

So, what can AI do in reducing incarceration and predicting recidivism? Recidivism; we have seen enough examples in the past. First of all, AI tools in judicial decision-making can play a major role there. The Tulane University study of AI in over 50,000 convictions in Virginia pointed out that AI scores offenders' risk of re-offending in the sense of whether they will recommit the crime. And then AI advises judges on sentencing options. And then AI helps correct gender bias and racial bias in judges' decisions. Judges often decline alternative punishments to defendants of colour. So, this again is a major issue because the alternative punishments could be either the person is let out on bail with some conditions or conditional bail, or a person is released with electronic surveillance placed on them in the sense that their movements are monitored electronically. So, although technically the person is free, he is free from jail, but his or her movements will be monitored. So, then there are alternative punishments. But what this study found is that for people or the descendants of colour, the AI-based or AI-enabled system offered certain choices, but then the judges declined to make use of those alternative punishments.

The New York Parole Board uses COMPAS for risk and needs assessment; we have discussed about COMPAS earlier. It factors in educational level, age, and re-entry plans, and the studies of AI effectiveness in these matters show mixed results in various studies in the USA. One is that some of them are critical in the sense that some point out that while these systems seem to be efficient and are able to deal with large numbers, their efficiency is debatable because their judgment is no better than that of humans. Or put it in other words, human judgment in such cases is as good as, and preferable to, using AI systems for it. Second, some of them are critical of such studies for the simple reason
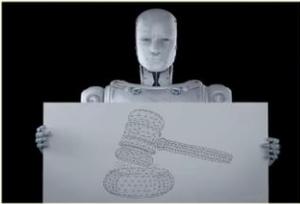
that, like the Parole Board's study on COMPAS, some other studies pointed out that the systems are inherently biased or inherently discriminatory. Similarly, it has also been found that while some systems may seem to be better or may seem to be less biased, what happens is that the decision given by them is something a judge may not accept because the judge may be more compassionate or may take into account some other factors.

Moreover, going by the findings of the system and then deciding on that basis by the system itself is not a better choice because some things are better left to humans than to the systems themselves. So, some studies are positive, like this study which said that it corrects gender bias and then racial bias in judges' decisions. So, the results are mixed, and they vary from application to application. But the question here is what ethical considerations need to be brought in? And how do we assess these systems in terms of ethical parameters? In the USA itself, there is a lot of literature on it; there is a lot of debate ongoing as to how we bring in ethical considerations and how we ensure that some of the problems associated with PSA, which we have seen in the earlier class, or with examples like COMPAS can be combated. Of course, there is no final word on these things; there is no final decision on many of them, but experience shows that over a decade after many studies, the justice system is convinced that there are really some issues with that, and they need to be addressed, or the blind faith in these systems has really come down.

And more importantly, frequent use of these systems also results in many studies that are showing the potential for harm or the probable issues with these systems. So, AI rules in reducing jail terms as well as predicting recidivism, resulting in a direction where, although ultimately the role of AI may not diminish, the decision support systems may improve on one hand, and on the other hand, the whole idea of predicting recidivism may not be left fully to AI in the future.

The study that we have now pointed out by Tulane University stated that it scored offenders by their risk of re-offending and advised judges on sentencing options. So, this is possible with tools, and then the final decision is made by human judges. But it also said that there was a disproportionate decline in offering alternative punishments. So, this

means that the use of this system is very mixed. AI is often ignored by some people, but they also ignored systems that provided an alternative punishment. So, in such cases, the use of AI should be done with care, or AI should not be the one that is the primary decision maker.

And then the parole-based tool COMPAS for risk and needs assessment, which we have seen earlier that the state parole board uses are again, a controversial one. It included education level, age at conviction, and re-entry plans. The University of California-Davis analysed 4,000 individuals later released on parole over a period of three years (2012-2015). It then looked at the documents based on the COMPAS analysis and the parole decisions. It found that parole was often denied to low-risk individuals due to the severity of the initial offenses. Now, what are the initial offenses and why does this make a difference? Because the COMPAS gives undue weight to the severity of the initial offenses. For example, if a person is initially punished for an offense with a specific prison term for some crimes that are classified as severe, then when the same person applies for bail or parole, the parole is denied to them, even though the overall risk is lower because of the earlier punishment or conviction. This means that the COMPAS software is problematic because it is unduly harsh or unduly discriminatory against people who were previously convicted of a severe crime. Now let us look at a different perspective here. A person who has already committed a severe crime may or may not commit another severe crime in the future. And it is also possible that if a person is released on parole, he may come back, or he may not even commit a crime. Because we need to look into the human tendency to reform, to repent, to make good, and to come to terms with the past.

But when undue weight is given to the past as well as the punishment for past crimes, the potential for refinement and the potential for a person redeeming himself or herself are reduced. So, when parole is denied on these grounds, it only reinforces the faith of the person that the system is inherently discriminatory and biased against him or her. So ultimately the risk may be low, but then the system twists things in such a way that even low-risk individuals are denied parole. Now the problems with COMPAS have been

studied in detail, along with the ProPublica investigation. There again, they said the algorithm is no better than human judgment, and the potential bias in the COMPAS algorithm is too well known by now.

Ethical Considerations and Human Oversight

- Ethical Gen AI Principles
  - Formulation of guidelines for ethical AI use
  - Addressing issues of bias, fairness, and accuracy
- Governance Frameworks
  - Establishment of structures to oversee AI implementation
  - Ensuring reliability and data privacy
- Interdisciplinary Expert Engagement
  - Continuous collaboration with experts to tackle ethical issues
- Human Oversight
  - AI as a tool, not a replacement for human Judgment
  - Essential for evaluating and identifying unconscious bias

So, what are the ethical considerations and human assessments we need to look into? First of all, in such cases, we need to have specific guidelines for ethical AI use. And then we need to address the specific cases of bias, fairness, and accuracy. Then, with the implementation of AI, we need to look at structures to oversee that, ensuring that they are reliable and that data privacy is ensured. And then these are the recommendations given by various studies. One way to address this is to work collaboratively and continuously with AI experts and others to take care of ethical issues. Then human judgment should prevail, not AI judgment, and more importantly, we need to look into identifying unconscious bias. So human oversight should play a major role when it comes to decisions like parole or the granting of bail.

Preliminary Findings

- Access to Justice
  - Improving accessibility through AI
- AI & Future Technologies
  - Exploring future applications of AI in the legal system
- Best Practices in Courts & Administration
  - Implementing AI for efficient court administration
- Establishing Gen AI Literacy in Courts
  - Educating court staff on generative AI
- Generative AI
  - Utilizing AI to generate legal documents
- Government
- Justice Tech- Need for Justice Tech that can use AI tools
- Effectively for access to Justice

So, the overall findings are that improving accessibility to justice through AI is a positive one. And then future applications of AI in the legal system need to be looked at. And then we can look into the best practices for courts and administration and try to implement them elsewhere as well. And then, if we educate the court staff on generative AI, their literacy improves, and they are able to use it better. And then we should use AI to generate legal documents more effectively. So, what is often called for is justice technology. And then the idea of justice tech is also very similar to what is called public technology, technology for the public, technology for good, or inclusive innovation by different terms. But the need for justice tech means that we can use AI tools to further the objectives of technology, using AI as an effective tool. So, the call for justice tech again goes hand in hand with access to justice. Effective access to justice can be achieved if we develop sufficient tools that can be considered and used as justice tech.

Civil Rights Division Guidance & Other Documents

- **Guidance for Employers Using Automated Software**
  - Issued on December 1, 2023
  - Discusses considerations for using software to handle Form I-9
- **Guidance on AI and Disability Discrimination in Employment**
  - Released on May 12, 2022
  - Describes how AI can lead to disability discrimination in hiring
- **Article on Civil Rights in the Digital Age**
  - Published in January 2022
  - Overview of issues with AI in employment decisions
  - Discusses work by the Department of Justice and other federal agencies

On the other hand, we looked up something called a public defender. But in the USA, there is also a civil rights division, which is part of the government. This Civil Rights Division does a lot of things. I mean, it looks into cases of discrimination, potential discrimination, and cases where discrimination has already occurred. Because civil rights protection also includes protection against discrimination on account of colour, age, or any undue discrimination. So, the Civil Rights Division of the Federal Government has issued many guides and other documents. For example, the guidance document of 2023 examines employers' use of automated software. It gives specific guidelines and considerations to handle this software and then it looks into guidelines for AI and disability discrimination in employment why because in case of employment disability or physical disablement could be a criterion that may go against many people who otherwise would be qualified. Under the US Act or under the US legal system, using disability as a factor to discriminate is an offense. And then the civil rights division can step in and try to settle this matter by suing an employer or a potential employer who uses this idea of disability as a discriminating factor. So, this guidance looks at how AI can lead to disability discrimination in hiring, or it identifies the ways AI can be used in hiring on account of discrimination. Then the Civil Rights and Digital Age published an article and said, "Look, these are the major issues that have to be taken into account by federal employees in employment decisions." Then it also looked into what the Department of Justice and other agencies are working on. In the subsequent slides, we will look at some of the instances where the Civil Rights Division has effectively intervened and then tried to get justice to combat discrimination.

AI and Disability Discrimination in Employment

- Department of Justice's Technical Assistance Document
  - Released on May 12, 2022
  - Title: "Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring"
- Content of the Document
  - Describes how algorithms and AI can lead to disability discrimination in hiring

The Department of Justice Technical Assistance Document dated 2022, titled "Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring," describes how algorithms in AI can lead to disability discrimination in hiring. So, it is alerting the employees or its own staff and others that there is a significant potential or possibility of AI being used to discriminate on account of disability.

Civil Rights in the Digital Age

- Overview of AI in Employment Decisions
  - Examines the impact of AI on employment practices
  - Highlights predominant issues arising from these practices
- Department of Justice's Role
  - Discusses the work being done to address AI-related issues
  - Collaboration with other federal agencies

So, the idea of civil rights in the digital age looks at the impact of AI on employment practices, predominant issues surrounding that, and then what work is being done by different federal agencies.

Cases and Matters

- Robocalls and Voting Rights Act
  - Filed statement supporting private plaintiffs
  - Challenged AI-generated robocalls as coercive
- Fair Housing Act and Tenant Screening
  - Filed SOI in Louis et al. v. SafeRent et al.
  - Alleged discrimination against Black and Hispanic applicants
- Accessibility at University of California, Berkeley
  - Filed consent decree for inaccessible online content
- Meta Platforms Housing Advertisements
- Discrimination in Job Advertisements
- Microsoft Citizenship Status Discrimination
- Ascension Health Alliance Investigation

And then there are some cases that we will look into. The Robocalls and Voting Rights Act: Robocalls are nothing but automated calls in the sense that you receive a call on your phone number from an automated calling system that talks to you immediately and then collects some information or asks for your response. In such a case, the Civil Rights Division filed a statement supporting private plaintiffs who said that they were discriminated against by the robocalls because it stated the robocalls were coercive; we will see later the details on that. In the Fair Housing Act and tenant screening, it filed a motion (SOI) in Saint Louis against a tenancy leasing firm called SafeRent et al., stating that their pattern or practice discriminated against Black and Hispanic applicants in that they used AI to screen applications. Then it went to a specific instance at California Berkeley where it stated that California Berkeley's online content was not available to some selected stakeholders and that this constituted a case of discrimination. And then at Meta Platforms, there were issues with housing advertisements for hiring. So, it went into job description advertisements, then Microsoft citizenship status discrimination, and then a health alliance investigation. So, we will look into these things in a little detail now.

Robocalls and Voting Rights Act

- Department of Justice's Involvement
  - Filed a statement of interest
  - U.S. District Court for the District of New Hampshire
- Nature of the Lawsuit
  - Challenging AI-generated robocalls
  - Known as 'deepfake' robocalls
- Violation of Voting Rights Act
  - Section 11(b) violation
  - Intimidating, threatening, or coercive robocalls
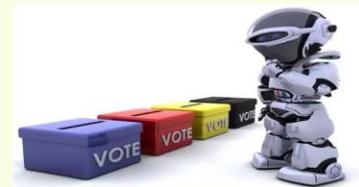
In the robocalls, the Department of Justice (DoJ) filed a statement of interest, which then went to the US District Court for the District of New Hampshire. It said that the AI-generated robocalls were, in fact, known as deepfake robocalls because they threatened, intimidated, and coerced the people who were on the other end. So, it stated that it violated section 11b of the Voting Rights Act. So, the idea here is that people should not be intimidated, coerced, or non-voluntarily forced to disclose information that they would not otherwise disclose.

Algorithm-based Tenant Screening Systems

- Statement of Interest Filed
  - Filed by Department of Justice and Department of Housing and Urban Development
  - Filed in January 2023
- Case Details
  - Louis et al. v. SafeRent et al.
  - Complaint against SafeRent, formerly CoreLogic Rental Property Solutions, LLC
- Allegations
  - Discrimination against Black and Hispanic rental applicants
  - Applicants using federally-funded housing choice vouchers
  - Violation of Fair Housing Act and Massachusetts state laws

In the case of the Department of Justice and the Department of Housing and Urban Development, they went to a specific instance where The SafeRent et al. is a company that was formerly called CoreLogic Rental Property Solutions, LLC. It used algorithm-based tenant screening. Now the tenants are screened by these systems based on some

parameters, including whether they are people of colour, as well as other factors, including their citizenship criteria and more. So, it found that the AI-based algorithmic system discriminated against Black and Hispanic rental applications and particularly discriminated against them in instances where they were using federally funded housing choice vouchers. See the federally funded housing choice vouchers are made available to poor and some other tenants who otherwise would not be able to afford certain houses that are unaffordable to them. So, using those vouchers, they will be able to pay the rent or at least bid for those houses where they can go as tenants. Then it said that this action by the algorithmic-based tenant screening system violated the Fair Housing Act as well as Massachusetts state laws. So, it found that the use of an algorithm-based tenant screening system was prima facie algorithm-based discrimination against specific people of black and Hispanic origin, particularly when they applied for rental applications.

University of California, Berkeley Consent Decree

- **University of California, Berkeley Settlement**
  - Filed in November 2022 by the Civil Rights Division
  - Allegations of inaccessible online content for individuals with disabilities
  - Inaccurate automated captioning technology for hearing impairments
  - Decree approved on December 2, 2022
  - University to provide accurate captions and not rely solely on YouTube's AI-based technology
- **Meta Platforms, Inc. Settlement**
  - Filed in June 2022 by the Civil Rights Division and U.S. Attorney's Office for the Southern District of New York
  - Complaint and proposed settlement agreement in United States v. Meta Platforms, Inc.
  - Settlement agreement signed on June 26, 2022, and final judgment entered on June 27, 2022

So, such use of discrimination—identifying them and then suing against that by way of identifying the specific provisions—indicates that while discrimination is not novel, what is novel is the use of AI-based algorithmic decision-making. So, the law has to keep up with that, and then the agencies have to keep up with that to ensure that these discriminations, which are technologically sophisticated, being used do not survive or do not continue even in this age. So what is happening is that some of the earlier ideas or earlier provisions, like fair housing or the voting rights of the people, when they are getting negated in part or in full on account of various AI-based strategies, including deepfake calls and then AI-based systems based on faulty algorithms, agencies have to step in to ensure that civil rights are protected and that there is no discrimination. So, in the case of the University of California, Berkeley, the Civil Rights Division stepped in and said that the online content made available by the university is inaccessible for people with disabilities; people with disabilities here obviously mean those who are not able to see or hear.

Often, you would find these days that many of the websites enable you to hear or, by voice, give you the option to hear them rather than see, and many websites have options for visually disabled people to increase the length or access the site otherwise. So, it said

that inaccurate automated captioning technology for hearing impairments was not valid, so it got a decree approved on that, and then it said that the university has to provide accurate captions and not rely solely on YouTube's AI-based technology. Mark the point here; it said, "It is the duty of the university to provide accurate captions there instead of relying on the technology developed with AI by YouTube for the use on YouTube." So, it is prima facie the responsibility of the university in this context to ensure that the captions are accurate and accessible to people with disabilities. In Meta Platforms' case, the case was settled in its favour. In June 2022, the Civil Rights Division and the U.S. Attorney for the Southern District of New York jointly filed a case. They said that there were some issues with Meta Platforms, and then they entered into a settlement agreement in June 2022 that was finalized by June 27, 2022.

What is the case here? So, in June 2022, the initial round of settlement stated that when meta-platform advertisements are posted on college and online recruitment platforms, including Georgia Tech's platforms, the advertisements discriminated against non-U.S. citizens. Here, the point is, even if the advertisements discriminate against non-U.S. citizens, the Civil Rights Division is ready to step in and state that this is a blatant case of discrimination and a violation of the Equal Rights Act.

Settlements with Employers Using Recruitment Platforms

- **Settlement with Microsoft Corporation**
  - Resolved claims of discrimination based on citizenship status
  - Microsoft engaged in unfair documentary practices
  - Used employment eligibility verification software improperly
- **Settlement with Ascension Health Alliance**
  - Investigation settled in August 2021
  - Unfair documentary practices in employment verification
  - Improper programming of verification software
  - Sent unnecessary reverification e-mails to non-U.S. citizens
- **Immigration and Nationality Act's anti-discrimination provision**
  - Prohibits requesting more or different documents than necessary
  - Protects against discrimination based on citizenship, immigration status, or national origin

So, what happened in this case was that Microsoft Corporation received claims of discrimination based on citizenship status. The case alleged that Microsoft's documentary practices were discriminatory; it used employment eligibility criteria improperly to verify citizenship and then was biased against non-U.S. citizens. In the case with Ascension Health Alliance, the case alleged that the practices of employment verification were unfair, and then the documentation was improper; improper programming sent email for re-verifications that were unnecessary to non-US citizens, or non-US citizens were harassed by the emails that they should not have received in the first place. So, the Immigration and Nationality Act's anti-discrimination provisions prohibit requiring more than one different document, seeking unnecessary documents, and evidence itself amounts to an act of discrimination. Then discrimination based on citizenship, immigration status, or natural origin itself is discrimination. So, when these factors were used to discriminate against individuals by different entities through different platforms, the civil rights division stepped in and found that these violations cannot be accepted and that they violated federal laws, particularly the Immigration and Nationality Act.

Purpose of the Guidelines

- Collaboration of Experts
  - Five judges and a lawyer/computer science professor
  - Members of the Working Group on AI and the Courts
- Development of Guidelines
  - Part of the ABA's Task Force on Law and Artificial Intelligence
  - Consensus view of Working Group members
- Purpose of Guidelines
  - Provide a framework for responsible AI use
  - Targeted at U.S. judicial officers

So, the various guidelines that the Civil Rights Division provided were that you need to collaborate with experts; five judges, a lawyer, and a computer science professor should be there to work on the groups of AI and the courts. Then you need to look into the purpose of the American Bar Association Task Force on Artificial Intelligence and then the consensus of the working group members. So, the idea here is that these provide a framework for responsible AI use and then use certain judicial officers. So, the purpose of different guidelines is to move towards responsible AI deployment in law and justice, particularly because various field's expertise is necessary rather than simply relying on computer scientists or data scientists.

Judicial Authority and AI

- Indispensable Judiciary
  - Independent, competent, impartial, and ethical
  - Essential for justice in society
- Judicial Authority
  - Vested solely in judicial officers
  - Not in AI systems
- Technological Advances
  - Offer new tools to assist the judiciary
- Core Obligations of Judicial Officers
  - Maintain professional competence
  - Uphold the rule of law
  - Promote justice

The judicial authority and AI: the judicial authority should be independent, competent, impartial, and ethical, and then it should be solely vested with the judicial authority not

with the AI systems. This is what the guidelines say. And then technological advances, the Civil Rights Division, and other federal agencies are willing to work with various entities to provide new tools to assist the judiciary. But the core obligation is to provide, maintain professional competence, uphold the rule of law, and promote justice.

So, what exactly can AI do for judicial integrity? First and foremost, AI should strengthen judicial integrity but not compromise it. Judicial officers must remain impartial to ensure public confidence in the sense that they should not unduly rely on AI, and their behaviour and use of AI systems should ensure that they appear impartial and are also impartial so that public confidence is maintained. And then they are solely responsible for their decisions, so they need to come to grips with and understand how the AI tools work; more importantly, they should know that if they are going to use any extrajudicial information from AI, then there is a risk of relying on that. They should decipher what information AI is providing them and whether it is really necessary. If you are relying on extrajudicial information from AI, you may end up taking a risk or giving a judgment that may not be correct. So how do we balance AI's principles and promise with the core principles of the judiciary? So, AI can produce increased advanced justice, but don't over-rely on AI. And then judicial officers ensure that AI enhances their responsibility, but AI cannot be held responsible; they cannot substitute their responsibility by using AI tools, nor can they assign their responsibility to AI tools.

Limitations of Gen AI

- **Understanding Gen AI Tools**
  - Gen AI tools generate content based on prompts and training data
  - Responses may not be the most correct or accurate
  - Gen AI does not engage in traditional reasoning or exercise judgment
- **Vigilance Against Bias**
  - Avoid becoming anchored to AI responses (automation bias)
  - Account for confirmation bias
- **Disclosure Obligations**
  - May need to disclose AI use under local rules
  - Obligation to avoid ex parte communication
- **Verification of Work Product**
  - Judicial officers are responsible for materials produced in their name

Having said that, we need to look into the limitations of Gen AI. Gen AI content is based on prompts and training data. We have seen that how you give the prompt often decides what the potential outcome could be. So, this is again an issue when it comes to Gen AI tools. If the prompt gives a decision and a hint, and if the prompt itself is taken as a hint, then the answer is tailored to that. So, the training data and the prompt also make a difference. So, the judges should be aware of this and should know exactly what they should ask when they use Gen-AI tools, or they should know the limitations of using the wrong prompts, as the prompts can be misleading. Judges should know that Gen-AI does not engage in traditional reading, listening, or exercising judgment, so they should understand the biases involved and avoid becoming anchored to automatic bias in AI responses. They should also account for confirmation bias, and they may need to disclose AI use under local rules. More importantly, they should be under an obligation to prevent ex parte communication, which refers to communication with one party without the other party being aware, and using AI tools incorrectly to send ex parte communication should be avoided. More importantly, the work product materials produced in their name should be verified by them, or when AI gives an output in their name, the work product obtained using AI should be verified and checked by them.

Confidentiality and Privacy Concerns

- **Gen AI Tools and Information Usage**
  - Prompts and information may be used to train models further
  - Developers may sell or disclose information to third parties
- **Handling Confidential Information**
  - Do not use health data, or privileged information in prompts
  - Ensure the Gen AI tool treats information in a privileged manner
- **Settings and Prompt History**
  - Pay attention to the tools' settings
  - Consider retaining, disabling, or deleting prompt history after sessions

Then Gen-AI tool prompts may be used to train models further, so what happens is: If I am a judge and if I am using Gen-AI tools effectively again and again, the data or the prompts that I am using with the Gen-AI tools may be used by the developer to train subsequent models, and this information may get sold or shared with a third-party developer or another party. Therefore, it is important for the judge or the judicial officer not to specifically identify himself or herself when using prompts. For example, the prompt should be something where the privacy or the identity of the person who is using the prompt is not disclosed, and then the information should also not be disclosed. Suppose if I'm a judicial officer, by mistake I give the prompt where, unknowingly or knowingly, I display my role, my designation, and my name, or I say I am a county judge asking you this, then the prompt will address me as the county judge of that county and then will give me the output or the decision or whatever I want from it.

So, this sort of information should be kept confidential. And then the judicial officers or the judicial team should not use health data or privileged information in prompts. This is important because health data is very critical and something fundamental. I should give a prompt and then say an accused before me, so and so, is suffering from this and then wants bail on that ground or is wanting some concession from me on that health ground. So, the moment I give these details, ChatGPT or any other AI tool will give me the answers, but then I am unknowingly or knowingly destroying health data related privacy, since the tools may treat data in a very casual way. The data may be used for training purposes, or it may be used again with some other data set and then go somewhere else. So, this sort of confidential information sharing should be avoided. So, look at the tool setting, look at what exactly the prompts are, and then consider whether the prompt history should be made available or should remain in memory after the session. As a good practice, it is better to disable and then delete the prompt history after sessions, irrespective of the AI tool you are using, because knowingly or unknowingly you might have given some information that you otherwise would not have shared, and the prompting sessions may be using the inputs you provided to train the system further.

Quality and Reliability of Gen AI Responses

- Importance of Training and Testing AI Tools
  - Critical for pretrial release decisions and criminal convictions
  - Ensures validity, reliability, and minimizes bias
- Quality of Gen AI Responses
  - Depends on the quality of the prompt
  - Responses can vary even with the same prompt
- Training Data Sources
  - May include general Internet information or proprietary databases
  - Not always trained on non-copyrighted or authoritative legal sources
- Review Terms of Service
  - Check for confidentiality, privacy, and security considerations
- Accuracy Verification

And then comes the question of the quality and reliability of Gen-AI. So, this is again an area we discussed much earlier. So, we won't go into greater detail here.

Operational Data Analysis

- Time and Workload Studies
  - AI and Gen AI tools assist in analyzing time and workload
- Real-Time Transcriptions
  - Gen AI tools create unofficial/preliminary transcriptions in real-time
- Translation of Documents
  - Gen AI tools provide unofficial/preliminary translations of foreign-language documents
- Operational Data Analysis
  - AI tools analyze court operational data and routine administrative workflows
  - Identify efficiency improvements
- Document Organization and Management
  - AI tools assist in organizing and managing documents

So, AI and GenAI tools can be used in code systems to assist with time and workload studies, and then they can use unofficial preliminary transcription in real time, which, again, we all know that most of the conference tools like WebEx and Teams also have inbuilt tools that provide you with the summary or updated recording or synopsis of the sessions. So, real-time transcription is also possible through this. Then, of course, for foreign language documents, they also provide translations, and they can analyse data for routine administrative workflows. They can be used for efficiency improvements and to organize and manage documents. In the case of operational data analysis in the USA, Gen-AI tools are being used very extensively.

Editing and Proofreading

- Editing and Proofreading
  - AI and Gen AI tools for checking spelling and grammar in draft opinions
- Legal Filings Review
  - Gen AI tools to check for misstated law or omitted legal authority in filings
- Generating Court Communications
  - Gen AI tools for standard court notices and communications
- Court Scheduling
  - AI and Gen AI tools for scheduling and calendar management
- Enhancing Accessibility
  - AI and Gen AI tools to assist self-represented litigants and improve accessibility services

Then, of course, editing and proofreading, which many of us may not be using AI extensively for, but that, again, is a good potential use for judges and judicial officers.

Implementation

- Regular Review and Updates
  - Reflect technological advances
  - Incorporate emerging best practices in AI and Gen AI usage
  - Improve AI and Gen AI validity and reliability
- As of February 2025
  - No Gen AI tools have fully resolved the hallucination problem
  - Human verification of AI and Gen AI outputs remains essential
  - Some tools perform better than others

And then, how to implement this, reflect the technological advances incorporating emerging best practices in AI and Gen-AI usage; usage alone is not sufficient; you need to look at the validity and reliability of the tools you are using. So, the situation as of February 2025 is that the studies conducted on the EU application of AI in law and justice mentioned that no Gen AI tools have fully resolved the hallucination problems. So, it remains there. So, human verification of Gen AI and AI outputs is essential. Some tools perform better than others, but there are still a lot of issues with many tools that are used in law and justice.

AI and How to Get Started

- Decide Whether to Use Open or Closed AI Models
  - Evaluate the benefits and limitations of each model type
- Ensure Permission and Understand the Terms of Use
  - Review legal and ethical considerations
- Select a Few Simple "Low Risk" Tasks
  - Start with tasks that have minimal impact if errors occur
- Use a "Human-in-the-Loop" Approach
  - Incorporate human oversight to ensure accuracy
- Train Staff and Judges on AI Systems
  - Provide comprehensive training on AI functionalities
- Prepare for Advanced Tasks
- Engage in Knowledge Sharing

So how to get started? This is what the manuals state: evaluate the benefits, review the legal and ethical considerations, look at the low-risk tasks, and then keep a human-in-the-loop approach with human oversight to ensure accuracy, train staff and other systems, and then prepare for advanced tasks. When you graduate from a low-risk task, you go to an advanced task and then engage in knowledge sharing. See, these are available in the literature that I have cited, although I have not specified in which literature this is available. So, for further details, you can go to the literature and find out.

Accusatory vs. Inquisitorial Models

- Accusatory Model (Common Law)
  - Used in the US
  - Two equal and autonomous parties: suspect and prosecuting party
  - Each party makes its own case before a neutral judge
- Inquisitorial Model (Civil Law)
  - Used in most European continental systems, like France
  - Official authority collects evidence independently
  - Evidence is used to uncover the truth without consulting any party

Now there are two ways of judicial proceedings: one is the accusatory, and the other is the inquisitorial. Accusatory is a common law principle. It is used in the USA where two autonomous equal parties, party A files a case against party B; they are the suspect and the prosecuting party. Or it is state versus I or I versus the state. So, there are only two

parties: either private versus private, state versus private, or private versus state. And then each party brings its own case and then argues before a federal or a neutral judge, so the adverse accusatory model is that I acquiesce or I accuse or you are accused. This is where a neutral judge has to act as a balanced arbitrator, listen to both parties, and then come to the judgment. Whereas in the inquisitorial mode, in civil law particularly, this is the culture in Europe: cases' official authority collects evidence independently, and then the evidence is used to uncover the truth without consulting any party. So, in the inquisitorial mode, the authority acts as an inquisitor. An inquisitor is one who poses questions or is there to furnish or find the potential for punishment, so the authority collects evidence independent of the party. Now, why are we discussing it? We will come to it in a bit.

Exclusion and relevance of AI in evidence: this is again very important, we need to create something that is deemed relevant, federal relevance for evidence has to be assessed first, enhancement and augmentation of low quality footage has to be avoided, significant modification should be done, creation of new content may be slow but what really happened had to be looked into, relevance may be constrained if AI fails to show actual events, so this comes when you use AI evidence, particularly the visual and audio visual evidence as evidence.

Implementation of AI Policies at State Level

- State-Level Regulations
  - Rules vary depending on the stage of the criminal procedure
  - Concrete obligations imposed on law enforcement and criminal justice authorities
- California's 2024 Rules of Court
  - Set standards for risk assessment technologies
  - Used specifically for sentencing purposes
- Privacy Concerns
  - Use of risk assessments impacts the right to privacy

So, in the USA, different states have their own policies, and how the systems are being used also varies among different courts. California's 2024 Rules of Court state that risk assessment technology has met some standards and is used specifically for sentencing purposes. And then the risk assessment impacts on the right to privacy have to be taken into account.

Human Experts to Understand AI

- Adversary System for AI Use in Court
  - Challenges AI evidence
  - Ensures fair trial
- Federal Rules on Evidence (2023 Amendments)
  - Updated to address AI
  - Improves judicial gate keeping
- Regulatory Interventions
  - 2023 amendments to Federal Rule of Evidence 702
  - 2024 AI Policy

Then, in the adversarial system, we can use AI in court where, if it is used there, the AI evidence can be challenged, ensuring a fair trial. So, the federal rules on evidence have been updated to address AI-based or AI-generated evidence, which improves judicial gatekeeping. The 2023 amendments to the federal rules of evidence (702) and the 2024 AI policy have to be taken into account. Now the humans exposed to understand AI have received updated guidance and the guidance principle notes are available to them.

Legal Framework for AI Use

- Authorities' Access to Information
  - Necessary for accurate investigation and adjudication
  - Includes personal data
- AI in Criminal Procedure
  - Processes personal data in sophisticated ways
  - Raises privacy and data protection concerns
- Need for Regulations
  - Data integrity and quality
  - Reliability and security
  - Storage, retention, and sharing
- General Discussion on Privacy
  - Reference: Daniel Marshall and Terry Thomas, PRIVACY AND CRIMINAL JUSTICE

So, the legal frameworks are already in place, necessary for accuracy in criminal procedure; privacy concerns need to be addressed, and data integrity quality must be considered; this, again, we have discussed so many times earlier.

Broadened Ex-ante Regulation Approach

- Avoid AI in Critical Decisions
  - AI usage should be comprehensible and scrutinizable by human experts
  - Required by evidence-related law and due process
  - Critical decisions need adequate and concrete reasons
- Permissive AI Use in Less Critical Contexts
  - AI can be used where statistical precision is required
  - No human rights should be at stake
  - AI should corroborate human-made decisions
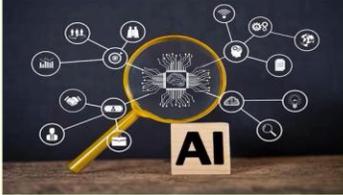 - AI performance should be subjected to enhanced checks and balances

So, the broadened ex-ante regulation approach is also emerging now. In this, AI usage should be comprehensible and scrutinisable by human judges. So, here AI should not be used in critical decisions. And then critical decisions need adequate and concrete reasons. First, the judge should decide whether he or she is going to use AI. And if he has to decide it, then it has to be justified with adequate and concrete reasons. Then AI can be used where statistical precision is required in greater detail, where there is no human rights law at stake, where this is a typical commercial transaction, or where the matter is more of data crunching than of violating any privacy or human rights. Then AI should

corroborate human-made decisions here, and the performance should be subjected to enhanced checks and balances.

In 2025, the Trump administration came up with new AI guidelines that were applicable to all federal agencies, which stated to keep minimum risk management practices, reduce vendor lock-in, improve transparency, and protect IP and public data. These are applicable to AI systems used in the judicial systems in the USA.

So, with this, we have seen a good example of, or many examples of, how AI is being used in the US at different levels, and then why there is so much variance available and why the views on COMPAS itself are not concurring with each other. More importantly, we also looked into and found that some of the tools in the USA, although controversial,

are similar tools being used extensively, particularly in criminal justice for decisions on parole or bail.

Therefore, we can safely conclude a few things here, although these are very preliminary hypotheses. One: Civil Rights Division and other institutions in the USA are now better equipped to handle cases based on AI discrimination, and they are updating themselves to manage such cases. Point number two is that, irrespective of the wide variance, most of the courts and institutions are moving towards greater use of AI, and to facilitate that, many guidelines, including bar association guidelines, guidelines by federal agencies, and guidelines by court counties, have been provided. Some of the evidence or some of the literature that I have cited specifically give good examples of this, for example, the AI rapid response team at the National Center for State Courts, or the one given as "Navigating AI in the Judiciary: New Guidelines for Judges and Their Chambers," which is forthcoming in 2025, deals with these topics in greater detail. So, the USA prima facie seems to be progressively addressing some of the concerns associated particularly with AI or discrimination based on bias to ensure that some of the civil rights are defended, some of the human rights are protected, and more importantly, the institutions are up to date with dealing with AI in law and justice.

Or in other words, since we are dealing with responsible AI in law and justice, such a topic needs a large amount of good literature survey and then good practices that need to be picked up from different jurisdictions and different countries. We also need to look at what people have done right or what people have not done right in the specific context. When we contextualize them, we will be able to identify what the major issues are in putting responsible AI to use in law and justice, and more importantly, in the Indian context. So, with this, we are completing the discussion on responsible AI in law and justice.

We have seen examples. We have seen the specific fundamental principles. We have gone into the general issues in law and justice when it comes to responsible AI. And then we have conducted two sessions on the use of AI, particularly in law and justice. So, our discussions by now, including responsible AI, ethical AI, evidence of explainable AI, and the use of AI in law and justice, would have given you a good overview and idea of some of the key issues regarding the use of AI in law and justice, of course related to responsible AI and explainable AI. But the overall picture is that a lot of things are happening; a lot of important developments are occurring, and we need to keep our eyes, ears, and minds open to learn from all these things and to pick up what is most relevant to us.

- AI and Judges

The next thing will be a very logical sequence to that, which is AI and judges. There, we will look into some of the things that we have discussed earlier regarding how different institutions are coming to grips with the use of AI in law and justice. Thank you.