Virtual Reality and the Law: Immersive Learning for Judges and Prosecutors – PRESENTATION

AI, Human Rights, and Immersive Judicial Training

Artificial intelligence (AI) is rapidly transforming many sectors, including the justice system. While AI can improve efficiency, it also raises new human-rights challenges. Experts warn that AI tools often rely on large datasets and complex algorithms that can pose risks to privacy, data protection, non-discrimination and access to justice[1]. At the same time, virtual reality (VR) and 360° simulations are emerging as modern training tools for judges and prosecutors. Immersive VR courtroom scenarios (with decision-tree structures) engage trainees in realistic cases, allowing them to make choices and see consequences in a controlled environment.

This brief handout explores the intersection of AI, human rights, and VR training, examining how biases in AI affect vulnerable groups, the legal safeguards courts must apply, and how VR can prepare jurists for these challenges. It builds upon and deepens the insights presented during the session *'Virtual Reality and the Law: Immersive Learning for Judges and Prosecutors'* held on 17th June 2025 in Krakow, by further elaborating on its key findings, extending its practical applications, and placing greater emphasis on the core legal and ethical issues surrounding AI.

1. Algorithmic Bias and Discrimination

As we are told AI systems learn from historical data and human-designed rules. When that input data reflect social prejudices or inequalities, the AI can "learn" and reinforce those biases[2]. In plain terms, "if the data fed into [an AI] system is biased, the result will be biased too"[3]. Two main causes drive algorithmic bias: (1) **Historical human biases** embedded in data (for example, past policing or hiring practices skewed against certain groups) and (2) **Unrepresentative training samples** (if a group is under-represented, the model makes more errors on it)[4]. The net effect can be to perpetuate or even deepen existing inequalities[5].

- **Disparate Impact on Minorities:** Biased AI often hits minorities and vulnerable groups hardest. For example, studies found that leading facial-recognition programs correctly identified light-skinned men but misidentified darker-skinned individuals and women at much higher rates[6]. In one case, a law-enforcement face ID system in Detroit (US) misidentified suspects 96% of the time and led to wrongful arrests of Black residents. Another analysis showed Amazon's Rekognition mislabeled 19% of women as men and 31% of dark-skinned women as men, while making no errors on lighter-skinned males. In practice, this means an innocent Black person might be falsely flagged by a biased face scanner, or a woman's face might be misread by AI in public-space surveillance.

- **Predictive Policing:** AI tools that predict crime hotspots can reinforce over-policing of minority neighborhoods. Predictive models trained on historical crime data will concentrate enforcement on the same areas that were already heavily policed. If judges or prosecutors rely on such tools (e.g. risk maps or predictive scores) without scrutiny, they may inadvertently validate these biased patterns.

- **Hiring Algorithms:** AI recruiters can encode gender or racial bias if trained on biased resumes. A famous example is Amazon's 2014–2017 experimental hiring tool, which downgraded resumes containing the word "women's" or graduates of all-women's colleges[7]. Because the tech industry's historical applicant pool was

mostly male, the AI "taught itself that male candidates were preferable," penalizing women's resumes[8]. Amazon ultimately scrapped the program after it became clear the bias was too deeply ingrained. This case shows how even without intent to discriminate, an AI can output unfair recommendations if its training data mirror a male-dominated workforce.

- **Criminal Justice Algorithms:** Risk-assessment tools used in courts may misjudge defendants. For instance, the COMPAS algorithm (used to predict recidivism) was found by ProPublica to misclassify Black defendants as high-risk far more often than white defendants with similar profiles[9]. Black defendants who did *not* reoffend were mislabeled high-risk 45% of the time (vs. 23% for comparable white defendants)[10]. Such errors could lead to harsher bail or sentencing for minorities. In effect, an AI recidivism score reflected "human prejudices such as arrest records" in the data[11].

These examples illustrate that algorithmic bias is not a hypothetical worry but a real human-rights concern. AI can **reinforce systemic inequalities**: when historically marginalized groups appear more often in "bad" data (e.g. crime records, denied loan applications, etc.), automated decision-making risks perpetuating those patterns[12]. As one analyst notes, "AI systems have the potential to deepen existing systemic inequalities" in healthcare, employment and criminal justice[13]. The message for judges and prosecutors is clear: they must be alert to how AI-derived evidence or decisions might carry these biases into the courtroom.

2. How Bias Arises

AI bias can be explained in simple terms: **machines learn from what people show to them.** If the training data contains prejudice, the AI will mimic it[14]. For example, if a dataset of past loan approvals contains fewer minority borrowers (even for reasons unrelated to creditworthiness), a credit-scoring AI trained on that data will be less likely to approve loans for those groups. In technical terms, a machine-learning model typically

optimizes overall accuracy or predictive power, which may inadvertently **trade off fairness between groups**. With imbalanced data, the model minimizes aggregate error by favoring the majority group. In practice, this means the AI's objective function has no built-in concept of "equal treatment," so it reinforces existing disparities. A vivid demonstration of this phenomenon can be found in the interactive game *Survival of the Best Fit* (LINK to the game: https://www.survivalofthebestfit.com/), which shows how automated decision-making systems can inherit and amplify societal biases. The game allows players to simulate hiring decisions and observe how seemingly neutral algorithms can lead to unfair outcomes, especially when trained on historically biased data.

Key factors include:

a) *Training Data Bias***:** Historical data may reflect discrimination. For instance, if policing data shows more arrests in certain areas due to biased enforcement, an AI using that data will see those areas as high-risk*[15]*. Rich data on one group and sparse data on another also causes skewed performance.

b) *Label Bias:* The outcomes used to train (e.g. "loan repaid", "crime occurred") may themselves be influenced by prejudice or social factors. If minority groups have worse outcomes for socioeconomic reasons, the AI will learn a false correlation.

c) *Feature/Proxy Bias:* Sometimes AI uses proxies (like zip code or credit score) that correlate with race, income or other protected traits. Without careful design, these proxies embed bias into the model.

d) *Model Design and Oversight:* Developers' own assumptions or a lack of diverse input can introduce bias. If a model is not audited for fairness metrics (e.g. equalized false-positive rates), disparities can go unnoticed.

The plain-language takeaway is that **algorithmic bias is often a reflection of real-world bias** in data and design*[16]*. Combating it requires awareness and deliberate testing, not just trust in "objective" software.

3. Legal and Ethical Implications for the Judiciary and Public Prosecutor Office

Courts are already grappling with how to handle AI in evidence and decision-making. Several legal and human-rights frameworks guide judges when AI is involved:

- **Fair Trial and Non-Discrimination (ECHR & EU Charter):** Under the European Convention on Human Rights (ECHR), Article 6 guarantees a fair trial and Article 14 prohibits discrimination. The EU Charter of Fundamental Rights similarly forbids bias and ensures due process (Articles 21 and 47). Any AI-based evidence or tool used in court must respect these principles. For example, if a prosecutor presents AI-generated analysis, the defense must have the chance to challenge it. Judges must consider whether using "black-box" AI violates the accused's right to understand and contest the evidence. The Council of Europe has emphasized that AI must remain "a tool in the service of the general interest" and that its use must "respect individual rights". In fact, the Council's new *Framework Convention on Artificial Intelligence* (opened for signature in September 2024) will establish binding standards to ensure AI upholds human rights, democracy and rule of law. This signals that future AI use in courts will be measured against stringent human-rights norms.

- **GDPR and Automated Decisions:** European data-protection law (GDPR) directly addresses automated decision-making. Article 22 of the GDPR gives individuals the right *"not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning"* them*[17]*. Even when automated decisions are allowed (for instance, under contract performance or consent), the controller must provide safeguards: notably, **human intervention**, an explanation, and the ability for the individual to contest the decision. In practice, this means a judge should be wary of accepting a purely AI-generated outcome (like an algorithmic risk score) without human review. A defendant should have the right to know how the AI arrived at its conclusion. If an

AI tool influenced sentencing or evidence presentation, GDPR grants the accused a line of defense against opaque profiling.

- **EU AI Act:** The European Union's AI Act (currently under adoption) will impose rules on AI systems according to risk levels. High-risk AI (those affecting fundamental rights, such as criminal risk assessment, facial recognition in law enforcement, or hiring algorithms) will face strict requirements: bias testing, mandatory documentation, transparency and human oversight. For instance, the Act generally **prohibits real-time biometric identification in public spaces** (like live face scanning), except narrowly for serious crimes[18]. It also requires all AI handling personal data to comply with the GDPR's principles (data minimization, purpose limitation)[19]. In theory, these rules aim to prevent AI from producing discriminatory outcomes[20]. However, as critics note, deep-seated bias in data may persist despite "neutral" design, and explanations provided by AI may be too technical for laypersons[21]. The Act emphasizes that AI should not replace human judicial decision-making[22], reinforcing the idea that judges remain ultimately responsible.

- **Fairness, Accountability, Transparency:** Legal experts stress **FAT principles** in algorithms. Courts may demand *explainability* of AI evidence ("show me the code or logic behind this report/outcome"), *auditability* (can we verify its performance across groups?), and clear *accountability* (who is liable for a flawed AI decision?). The EU's ethics guidelines and the Council of Europe's studies have repeatedly called for transparency and oversight in algorithmic justice, so that "public trust and the integrity of legal systems" are maintained. Judges should apply traditional evidentiary scrutiny to AI: assess reliability, consider bias, and ensure equal treatment. In short, courts must evaluate AI evidence under the same human-rights standards as any other evidence – upholding fairness and the right to a reasoned decision.

4. VR in Judicial Training for AI and Ethical Issues

Immersive VR simulations are an innovative way to train judges on these complex issues. Rather than passive lectures, VR places the trainee *inside* a lifelike scenario. For example, the Polish National School of Judiciary's VR/AI lab creates **decision-tree courtroom scenarios**: users choose actions at each step and witness either correct or incorrect outcomes, with legal explanations[23]. This "interactive film" approach engages judges' senses and judgment in real time. Judges report that VR training feels **realistic and safe**: one commented the VR tool "offers a more realistic courtroom environment than traditional mock trials"[24], and another called it "ideal… with no drawbacks"[25].

- *Immersive Decision Practice:* VR simulators can be programmed to include AI-related dilemmas. For instance, a VR case could involve evaluating evidence from facial recognition software or a computer-generated forensic report. By making decisions in a 360° scene, trainees experience the nuance of bias, privacy and procedural fairness firsthand.

- *Engagement and Reflection:* Immersiveness (the quality of fully engaging one's senses) helps judges internalize lessons. Participants face the consequences of "wrong" choices in a low-stakes setting. The VR scenarios in Poland's program include both correct and incorrect endings, with immediate feedback on the legal basis of decisions[26]. This decision-based learning helps clarify complex points (e.g. when an AI-generated piece of evidence might be unreliable).

- *Feedback from Trainees:* Judges who have tried VR training praise it. They say it reduces performance anxiety (no audience of peers) and focuses learning on the judge's role[27]. New judges find it "excellent" for building courtroom skills, while even experienced judges use it to rehearse unusual or tech-driven cases[28]. Many see VR as a powerful complement to traditional methods, especially for preparing decision-makers to recognize and navigate the ethical-legal dilemmas posed by AI[29].

In summary, VR-based training has a potential to translate abstract concerns (like algorithmic fairness or digital privacy) into concrete scenarios that judges can "live

through." By blending VR simulations with discussion of rights and rules, judicial education can help ensure courts are ready to uphold human rights in the age of AI.

Dariusz Szawurski-Radetz

Head of the VR/AI Section

National School of Judiciary and Public Prosecution

Sources: Authoritative studies and reports from the Council of Europe, academic research and legal analyses have informed this handout, as have evaluation materials from the National School of Judiciary and Public Prosecution VR/AI Lab training program.

[1] European Union Agency for Fundamental Rights. *Bias in Algorithms – Artificial Intelligence and Discrimination*. Publications Office of the European Union, 2022,

https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

[2] Lifshitz, B. (2021, May 6). *Racism is systemic in artificial intelligence systems, too*. *Georgetown Security Studies Review*,

https://georgetownsecuritystudiesreview.org/2021/05/06/racism-is-systemic-in-artificial-intelligence-systems-too/

[3] (Lifshitz, 2021)

[4] (Lifshitz, 2021)

[5] (Lifshitz, 2021)

[6] (Lifshitz, 2021)

[7] Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.

https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

[8] (Dastin, 2018)

[9] Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). *How we analyzed the COMPAS recidivism algorithm*. ProPublica.

https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[10] (Larson et al., 2016)

[11] (Lifshitz, 2021)

[12] (Lifshitz, 2021)

[13] (Lifshitz, 2021)

[14] (Lifshitz, 2021)

[15] (Lifshitz, 2021)

[16] (Lifshitz, 2021)

[17] **General Data Protection Regulation (GDPR), Article 22. (2016, May 4). Retrieved from GDPR-Info website:**

General Data Protection Regulation (Regulation (EU) 2016/679), Article 22. Automated individual decision-making, including profiling. In GDPR-Info.eu. Available at: https://gdpr-info.eu/art-22-gdpr/

[18] **Das Gupta, S.** (2025, February). *The EU AI Act and its adherence to the European Convention on Human Rights*. EMILDAI. Retrieved July 17, 2025, from

https://emildai.eu/the-eu-ai-act-and-its-adherence-to-the-european-convention-on-human-rights/

[19] (Das Gupta, 2025)

[20] (Das Gupta, 2025)

[21] (Das Gupta, 2025)

[22] (Das Gupta, 2025)

[23] **National School of Judiciasry and Public Prosecution VR/AI Lab presentation delieverd on 17th of June 2025:** *Virtual Reality and the Law: Immersive Learning for Judges and Prosecutors*: https://shorturl.at/1fAwN

[24] (VR/AI Lab presentation, 2025)

[25] (VR/AI Lab presentation, 2025)

[26] (VR/AI Lab presentation, 2025)

[27] (VR/AI Lab presentation, 2025)

[28] (VR/AI Lab presentation, 2025)

[29] (VR/AI Lab presentation, 2025)