# Fair Trial and the AI Act in Criminal Investigations and Digital Forensics

Adi Stoykova, Groningen University

Image: © Creation of A.I. After Michelangelo's Sistine Chapel by Je Moretti

# Agenda

**1. Fair Trial & Challenges with AI**

**2. AI Act: Case Studies**

- **Prohibited AI for law enforcement**
- **High-risk AI: computer vision + predictive policing**
- **Minimal-risk AI: speech-to-text**
- **GPAI: Chat GPT, Llama etc.**

**3. Criminal AI & Research Agenda**

# 1.Fair Trial and AI Evidence?

# The right to a fair trial

Article 6 ECHR, CoE
Article 10 UDHR, UN
Article 14 ICCPR, UN
Article 47 CFR, EU

- Universally recognized principle
- A standard for criminal procedure in accordance with the rule of law
- Art.6 ECHR by far is the most granularly developed

Source: simpson33 | Credit: Getty Images/iStockphoto

# ARTICLE 6 ECHR


1. In the determination … of any criminal charge against him, everyone is entitled to a *fair and public hearing* within a *reasonable time* by an *independent and impartial tribunal* established by law…

2. Everyone charged with a criminal offence shall be *presumed innocent* until proved guilty according to law.

3. Everyone charged with a criminal offence has the following minimum rights:
(a) to be informed promptly … of the nature and cause of the accusation against him;
(b) to have adequate time and facilities for the preparation of his defence;
(c) to defend himself in person or through legal assistance…
(d) to examine or have examined witnesses against him and to obtain the attendance and examination of witnesses on his behalf under the same conditions as witnesses against him;…

# Fair trial: Equality of Arms (Art. 6 (1) + (3) ECHR)

1. Fair procedure to evaluate the lawfulness and the lawful use of evidence

2. Possibility to challenge the evidence: fair disclosure of and to information about the evidence

3. Maintaining equality of arms against expert evidence …

**Stoykova R, 'The Right to a Fair Trial as a Conceptual Framework for Digital Evidence Rules in Criminal Investigations' (2023) 49 Computer Law & Security Review 105801.**

Source: simpson33 | Credit: Getty Images/iStockphoto

# Fair trial: Presumption of innocence (Art. 6 (2) ECHR)

**4.** Accurate fact finding

**5.** Protection against prejudicial effects in evidence procedure

**6.** Protection against Reverse burden of proof

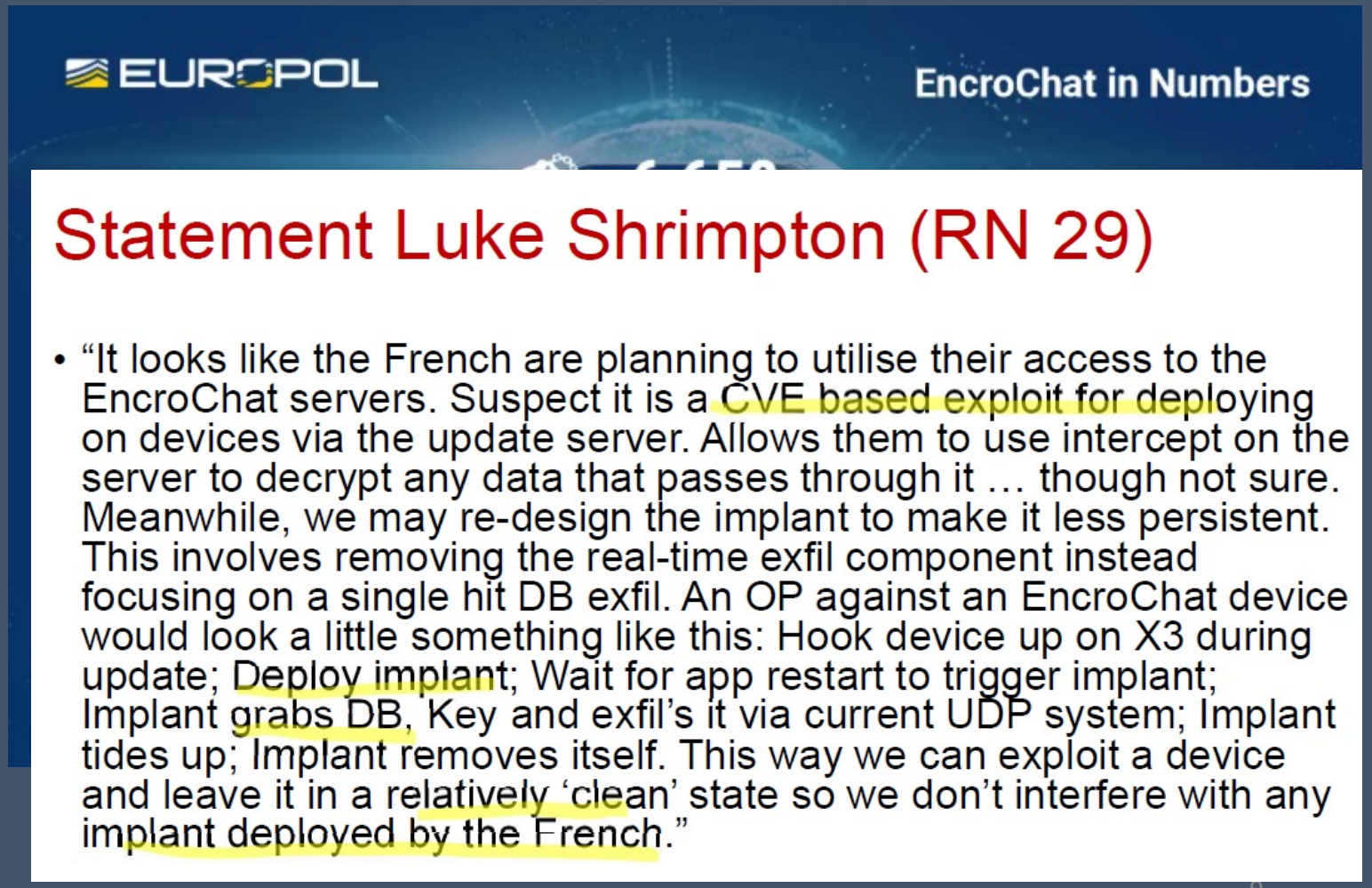
Source: simpson33 | Credit: Getty Images/iStockphoto

# 1. Lawfulness and lawful use of evidence

- **Lawfulness**: Substantive and procedural assessment
  - especially when the technology available for use is continually becoming more sophisticated.
  - intrusive measure to be based on presented facts, time limits, authorization, notification after termination and supervision for notification.
  - communicating record to judge and defence.
- **Lawful use:**
  - **Quality**: whether the circumstances in which it was obtained cast doubt on its reliability or accuracy.
  - **Contestability** opportunity of challenging the authenticity of the evidence and of opposing its use.
  - **Supporting evidence**: questionable evidence must be evaluated in the light of supporting evidence.

# 1. Lawfulness: *Challenges Encrochat*

- Encrochat lawfulness?
- Authorisation? French warrants? UK? NL?



# 2. Fair disclosure: Possibility to challenge the evidence

- obligation for the prosecution to disclose evidence.
- other evidence that might relate to the admissibility, reliability, and completeness of the former.
- a positive obligation to investigate and collect evidence in favour of the accused.

# 2. Fair disclosure: Large Datasets

- Rook v. Germany:
  - 78,970 telecommunication data sets
  - *14 million* electronic files
  - *1,100 files as relevant to the case*
- Requirements:
  - No obligation to disclose the full collection of data
  - the defence to be involved in determining the search criteria when filtering the full collection of data
  - to conduct further searches for exculpatory evidence

See also: Sigurður Einarsson and Others v. Iceland

# Ecrochat Slang: in Denmark and UK?

EXCLUSIVE: BRITAIN'S biggest ever sting against organised crime sting has been dealt a blow after the case against a man charged with conspiracy to supply drugs collapsed because the prosecution was unable to link him to an encrypted phone device central to the operation.

# Art. 86 AIA

*Article 86*

**Right to explanation of individual decision-making**

1. Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system [… ]and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

# Defendants and Rec. 59 AIA

- AI systems are characterized by a significant degree of power imbalance and may lead to surveillance, arrest or deprivation of a natural person's liberty as well as other adverse impacts on fundamental rights
- The impact of the use of AI tools on the defence rights of suspects should not be ignored, in particular the difficulty in obtaining meaningful information on the functioning of those systems and the resulting difficulty in challenging their results in court, in particular by natural persons under investigation.

# COMPAS in US

- Task: High or low risk?
- Input: historical arrest data + criminal history + criminal associates, substance abuse..
- Features: 137 points questionnaire?
- Problems:

- arrest data not representative
- discriminatory bias
- selected features not correlated to recidivism!
- data scientists in private company decide on balancing public interests and individual rights???


Larson J and others, 'How We Analyzed the COMPAS Recidivism Algorithm' (*ProPublica*, 2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=Tu5C70R2pCBv8Yj33AkMh2E-mHz3d6iu> accessed 02 June 2025.

# *State v. Loomis*

- Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing.
- Mr. Loomis challenged the Circuit Court's use of COMPAS at sentencing because it violated his due process rights when it interfered with his right "to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevent[ed] him from assessing its accuracy."

https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html and https://harvardlawreview.org/print/vol-130/state-v-loomis/

# 3. Equality of arms: technology-assisted expert evidence

- effective procedural measures to challenge expert evidence reliability, to contest and comment on the expert's findings.

- to be presented with the expert report and expert findings on exculpatory evidence.

- to be present at expert interviews, but also to access the full documentation on which the expert report was based.

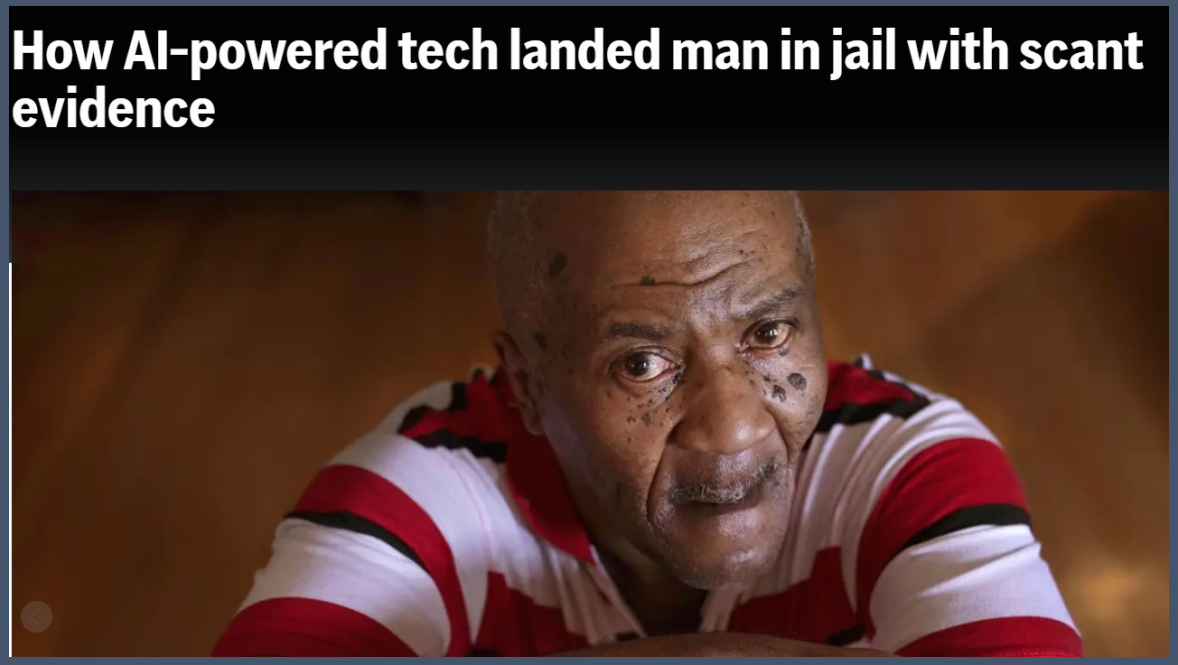
NBC NEWS

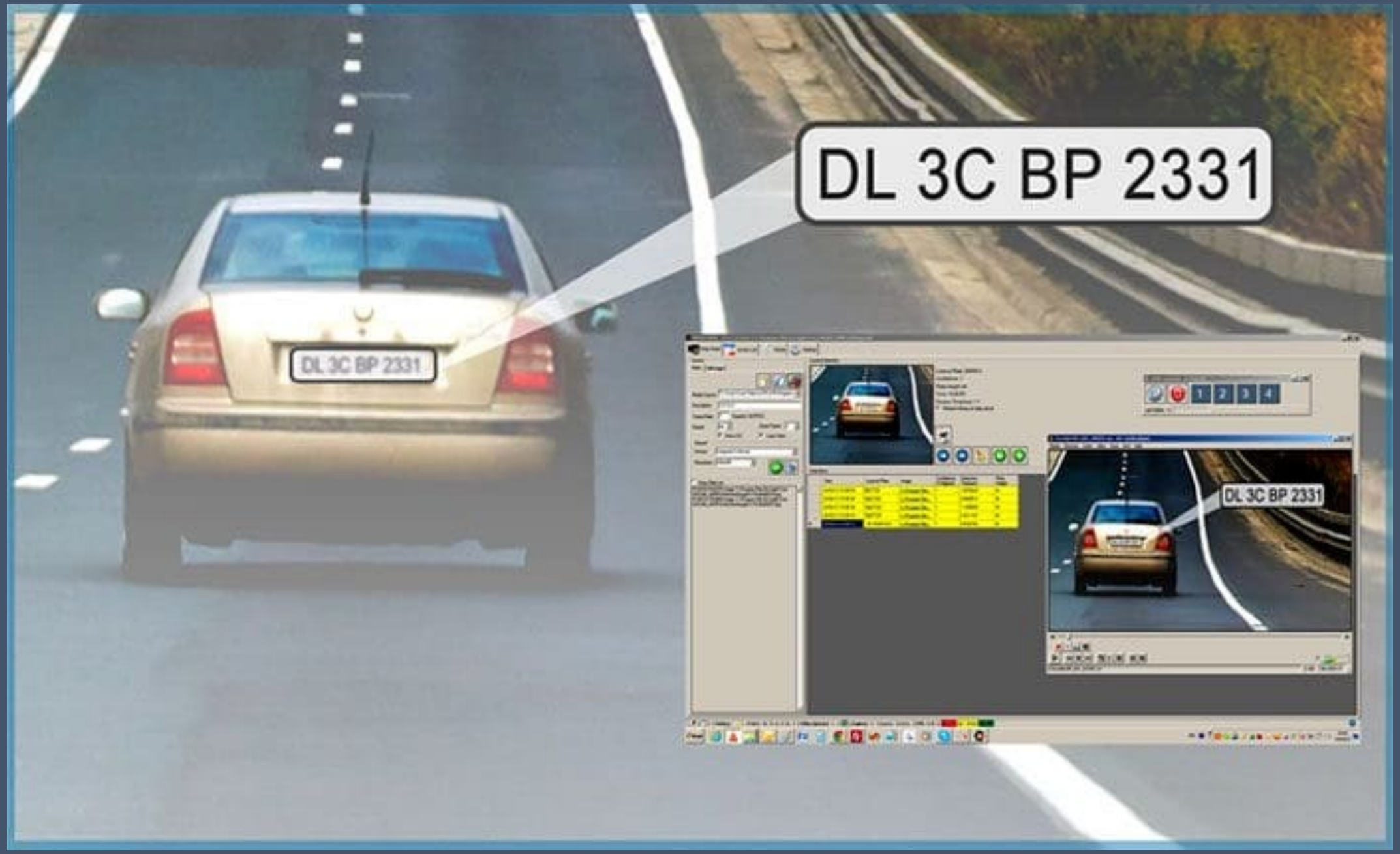
EXCLUSIVE May 3, 2024, 3:00 PM GMT+2

ARTIFICIAL INTELLIGENCE

# An AI tool used in thousands of criminal cases is facing legal challenges

Cybercheck's founder has said the software tops 90% accuracy. Defense lawyers have said he lied under oath about his expertise and made false claims about when and where the technology has been used.

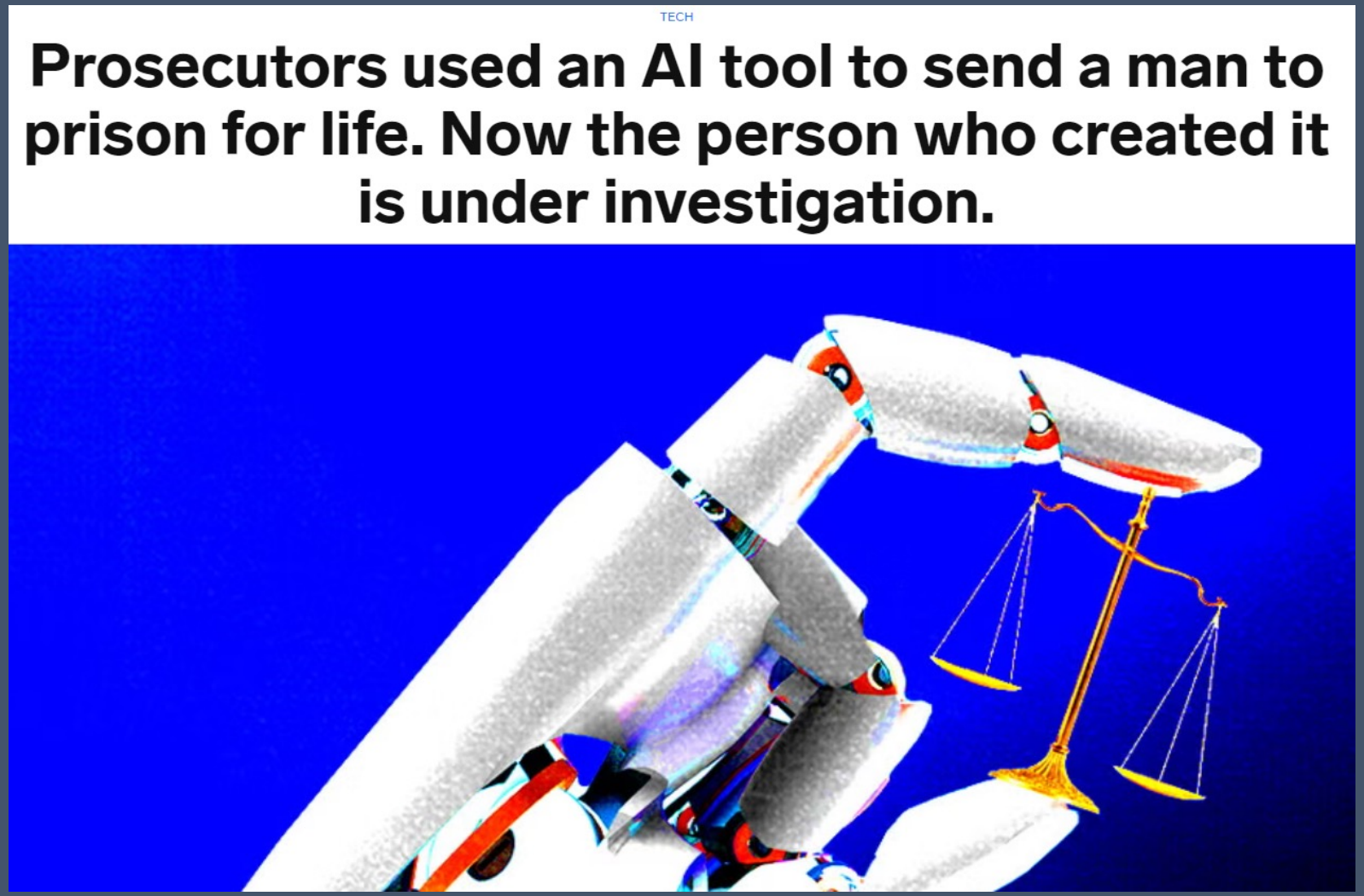





Schirrmacher F and others, ‘Benchmarking Probabilistic Deep Learning Methods for License Plate Recognition’ (2023) 24 IEEE Transactions on Intelligent Transportation Systems 9203

# 3. Equality of arms: AI-assisted expert evidence

- Commercial AI tool providers?
- Overreliance on expert opinion
- Who should comply with fair trial requirements for expert evidence cross-examination?




https://www.businessinsider.com/ai-crime-tool-cybercheck-founder-adam-mosher-investigation-2024-8

# 4. Accurate fact finding: *Challenges with AI*

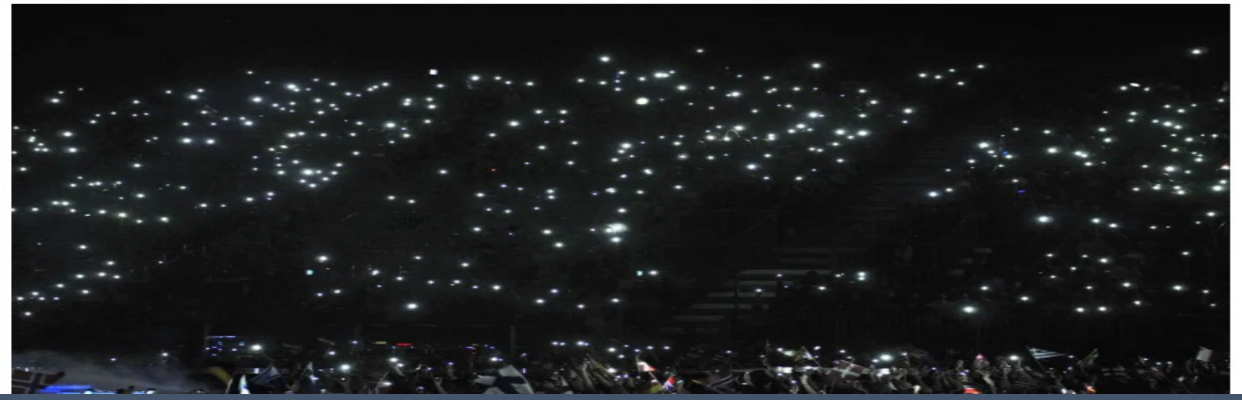
- What level of accuracy or probability should be achieved in order to conclude that the digital artefacts support reasonable suspicion? or
- What are the criteria for suitable hypotheses and methods to test them in order to comply with the presumption of innocence?

A screenshot of the DataWorks Plus Case Management interface, showing two algorithms run concurrently.
(Source: Hudson County, NJ Prosecutor's Office)

# 4. Accurate Fact-Finding?

The New York Times

## *Flaws in Cellphone Evidence Prompt Review of 10,000 Verdicts in Denmark*



TERS® World Business Markets Sustainability Legal Breakingviews Technology Investig

United Kingdom

## What is Britain's Post Office scandal?

By **Sachin Ravikumar**

January 10, 2024 11:11 PM GMT+1 · Updated 16 hours ago


A post office sign hangs above a shop in Belgravia, in London, Britain January 7, 2024. REUTERS/Hollie Adams *Acquire Licensing Rights*

LONDON, Jan 10 (Reuters) - One of Britain's biggest miscarriages of justice, the wrongful conviction of hundreds of Post Office workers due to faulty software, has exploded into the public domain following a TV drama, triggering demands for justice.

# 5. Prejudicial effects in evidence procedure

Protection against prejudicial **statements** about the facts by:

- The court
- State officials at the pre-trial
- The prosecutor

**Prejudice:** harm or injury that results or may result from some action or judgement.

**Evidence** that has a tendency to unduly influence the fact-finder to decide a matter on an **improper basis**:

- lengthy **delay** in bringing charges
- decision not based on facts but **discriminatory**, preconceived idea of guilt
- excessively long periods of pre-trial detention

# 5. Prejudicial effects: *Challenges in digital investigations*

- Prejudicial effects embedded in technology?
- Algorithms trained with discriminatory data
- Excessive long surveillance? Excessive data collection without bringing charges?
- Technology protection fallacy?

## Even algorithms are biased against black men

A study on offenders in Florida refutes the notion that computers are more objective than people

> One in three black men can expect to be incarcerated (compared with one in six Latinos and one in 17 whites)

# 6. Reverse burden of proof

- when the burden of proof is shifted from the prosecution to the suspect or defendant.

Presumption of fact and of law
e.g. Drug-smuggling

# 6. Reverse burden of proof?

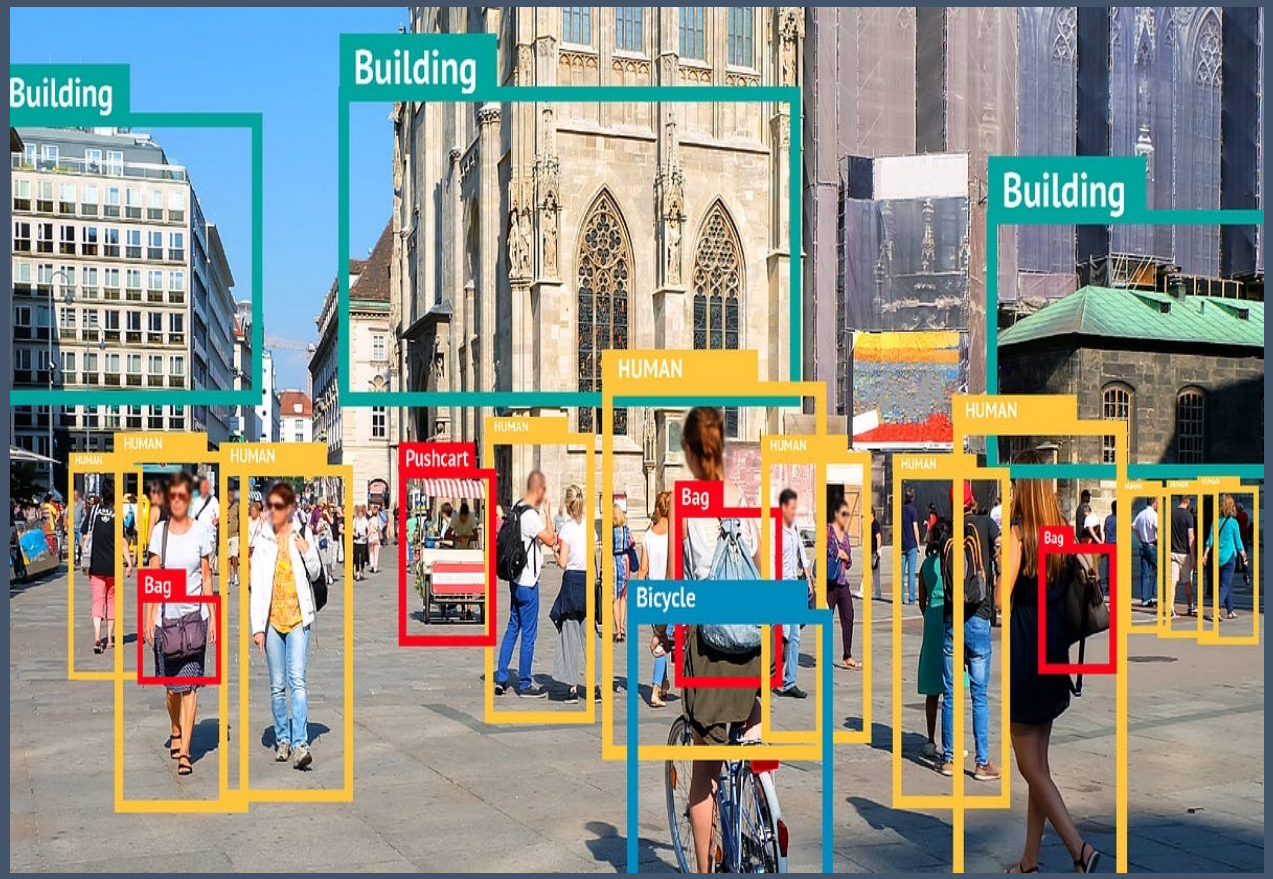

- What software was used?
- What was the reference database?
- Where the reference photo was taken from and what was the quality of it?
- How do they analyze if the sample suspect photo matches the reference one?
- What is a match?

- No criminal conviction merely based on outcome of FRT?
- False positives leading to false identifications and wrongful arrests?

# Instance search and Personal Data Protection

- Visual representations of individuals are personal data
- if they ‘can be linked to a particular person’;

- If the purpose of video surveillance is to identify the persons to be seen in the video images […] the whole application as such has to be considered as processing data about identifiable persons. (EDPB)




https://medium.com/@apandey_24903/automating-object-detection-62f4b432673c

# Training on non-personal data?

**From: Sensitive attributes in anonymized data**

- ZIP code + birth date + sex
- Netflix rating of 3 movies
- face anonymization provides minimal protection

To: **Overlearning**

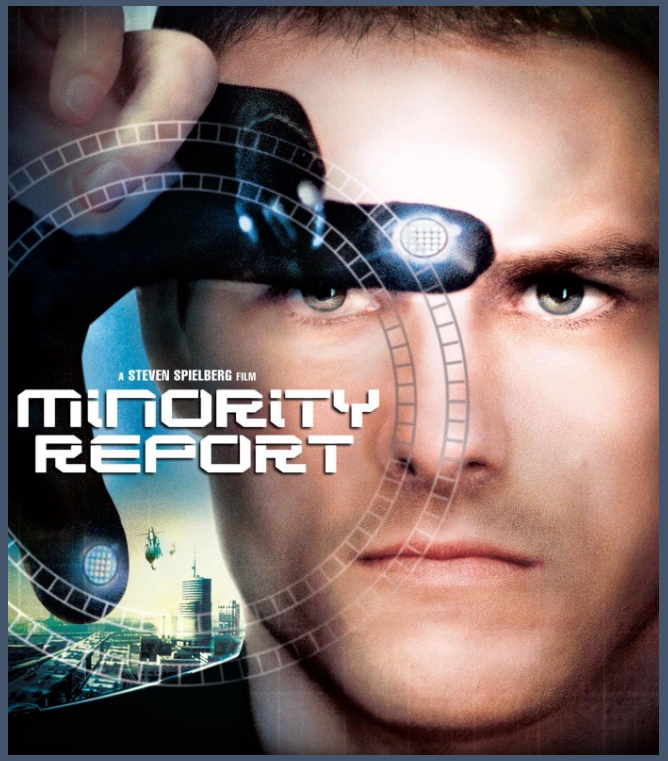
- Emergence of features that are much more general than the learning objective
- instance search models trained only on non-personal data still develop person re-ID capabilities.
- Personal data processing - from the moment the algorithm is deployed to a dataset with visual representation of people.

See Ohm, 2010; Song & Shmatikov, 2020; Dietlmeier, 2021; Nguyen & Stoykova, 2025 – under review.

# Profiling?

Art. 11 LED - decision based solely on automated processing, including profiling, that allows law enforcement to evaluate personal aspects of individuals and produce adverse legal effects or significantly affects them

# Annex III, point 6, letter (d) AI Act: High-risk Profiling?




- Article 5(1)(d) AI Act Prohibition of offenders risk assessment based solely on profiling

- **Exception:**

this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity

- it will be classified as a high-risk AI system (Annex III, point 6(d))?

AI systems intended to be used by law enforcement authorities ... for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling or to assess personality traits and characteristics or past criminal behaviour of natural persons or group

- NO RETROATIVE EFFECT? NO ANCILLIARY EFFECT? OUT OF SCOPE?

# Bibliography


- Nguyen, A. T. 2023. Adapting Video Instance Segmentation for Instance Search. In 20th International Conference on Content-Based Multimedia Indexing, CBMI 2023, 256–260.
- Dietlmeier J and others, 'How Important Are Faces for Person Re-Identification?', *2020 25th International Conference on Pattern Recognition (ICPR)* (2021)
- EDPB, Guidelines 3/2019 on processing of personal data through video devices, adopted on 29 January 2020.
- P. Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," *UCLA Law Review*, vol. 57, p. 1701, 2010.
- C. Song and V. Shmatikov, "Overlearning Reveals Sensitive Attributes," Feb. 08, 2020, *arXiv*: arXiv:1905.11742. doi: 10.48550/arXiv.1905.11742.
- Stoykova R, 'The Right to a Fair Trial as a Conceptual Framework for Digital Evidence Rules in Criminal Investigations' (2023) 49 Computer Law & Security Review 105801.
- Stoykova R, Porter K and Beka T, 'The AI Act in a Law Enforcement Context: The Case of Automatic Speech Recognition for Transcribing Investigative Interviews' (1 August 2024) <https://papers.ssrn.com/abstract=4913090> accessed 13 September 2024.
- Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act) , Brussels, 4.2.2025.
- Stoykova R. 'Encrochat: The hacker with a warrant and fair trials?' (2023) 46 Forensic Science International: Digital Investigation 301602.

# Thank you for your attention!


R. (Adi) Stoykova, PhD
Assistant Professor

Faculty of Law

Telephone:

E-mail: r.stoykova@rug.nl

Security, Technology & e-Privacy Research Group


The 'Security, Technology and e-Privacy (STeP) Research Group', is an interdisciplinary team of researchers – from early stage researchers to advanced researchers – organised within the Department of Transboundary Legal Studies (TLS). As can be derived from the name, the STeP Research Group is involved in research in three main areas in their broadest sense which are very much inter-related: security, technology and privacy.

https://www.rug.nl/staff/r.stoykova/?lang=en