EUI DEPARTMENT OF LAW

Detecting Deep Fakes in Criminal Proceedings

Opportunities and Challenges of Forensic AI

27 April 2022

Webinar ‘Forensic AI in eEvidence’

DG Justice and Consumers - European Commission

www.eui.eu

Structure

Deep Fakes in criminal proceedings

Detection through Forensic AI

The way forward

What is a Deep Fake?

- Manipulation techniques of **image, audio or video** content that ‘would falsely appear to a person to be authentic or truthful’ (Article 52(3) Proposal AI Act)
- ‘Harmless’ uses: social media, conversational agents, movie industry
- Harmful uses: deepfake pornography, fake news, blackmailing and so on

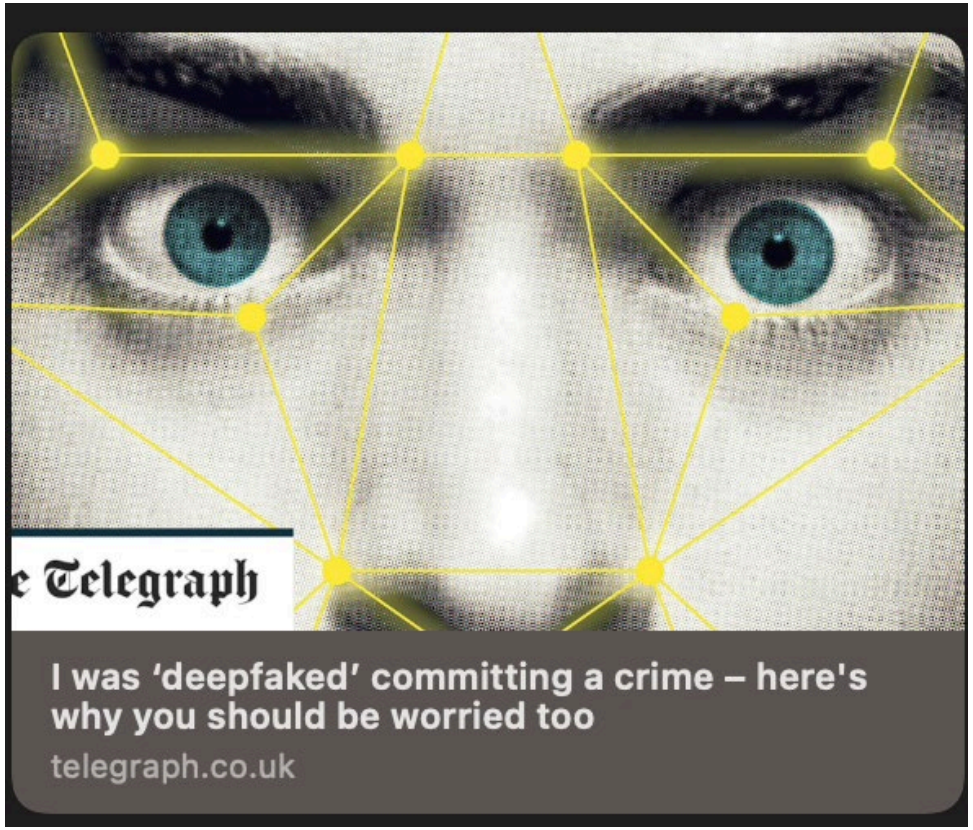
EUI DEPARTMENT OF LAW

Deep Fakes at Trial: Criminal Offence

- One of the ‘most dangerous form of AI-related crime’ (study by UCL, 2020)
- Deep fake as a crime itself?
  - EU: ongoing debate
    - Only labelling requirement in AI Act (Article 52(3) AIA)
    - Policy recommendation (EPRS Study, 2021): extend framework with regard to criminal offences
  - Germany: distribution that violates personal rights (§ 201a German Criminal Code)
  - UK: calls for specific crimes
  - South Africa: Domestic Violence Amendment Act now includes deepfakes (focus on gender-related harm)
- Safeguards through ‘traditional’ crimes
  - revenge porn, child pornography, privacy-related crimes, harassment, defamation, blackmailing

Co-funded by the Erasmus+ Programme of the European Union

Deep Fakes at Trial: Tampered Evidence



- Michael Grothaus, author of the book: "Trust no one: Inside the World of DeepFakes"
- *"There's a video of me attempting an armed robbery. […] But despite this video showing me committing the crime, it never actually happened*"
- First reported case
  - In the UK (2020) deepfake audio in a custody battle to portray the father as violent

Co-funded by the Erasmus+ Programme of the European Union

Tampered Evidence

Risks

1. High accessibility of the technology
   - No high skills required
   - Little data to train the Deep Fake AI (one video may be enough)
2. Very hard to detect
   - In 2019, Professor Hao Li: genuine videos and deepfakes indistinguishable in half a year

Consequences

- DeepFakes risk eroding trust in video, image or audio evidence
- If not detected, could lead to wrongful convictions or acquittals

Detection through Forensic AI

Detecting Deep Fake is challenging for humans

- Case in Malaysia: forensic analysis not able to verify authenticity
- Try the quiz developed by Microsoft “Spot DeepFakes”

Why not AI to detect AI?

- Study from UC Berkeley, machine learning system: accurate 92% (Agarwal S and others, 2019)
- Facebook “Deepfake Detection Challenge” (2019)
- Google Projects
- Microsoft Video Authenticator

Opportunities of Forensic AI

Human detection is labour-intensive, expensive or too hard

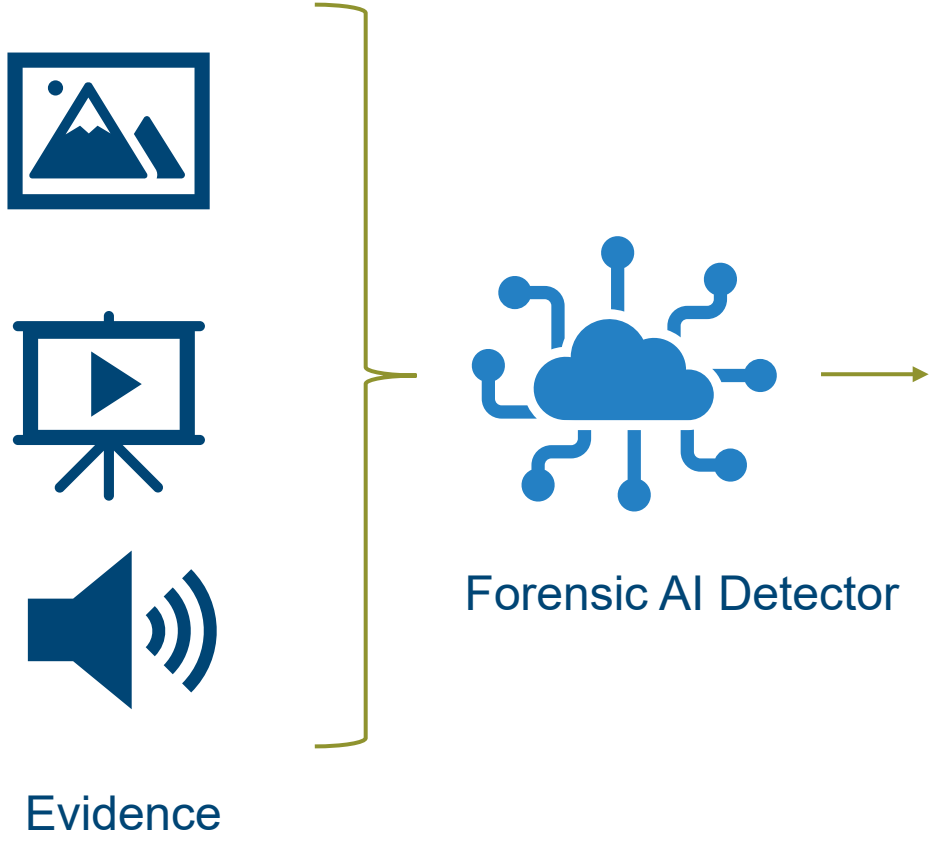
Tool for public prosecutors

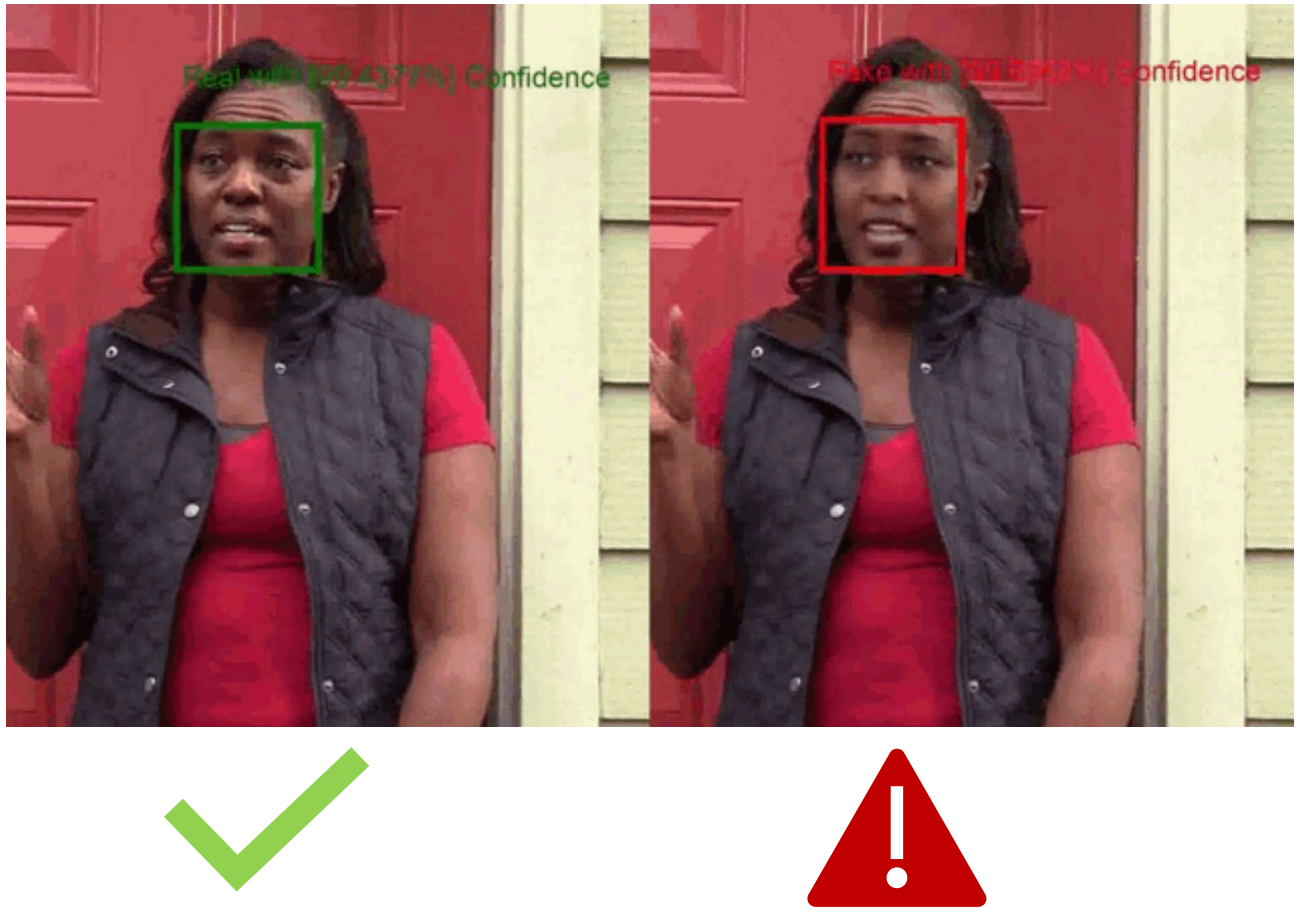
- To prove the ‘manipulation’ if deepfake is a crime
- To test authenticity of evidence (imagine a manipulated video for an alibi)

Tool for defendants

- In Virginia (2020), a law firm reported increasing use of false recordings against defendants
- To assess the authenticity of evidence brought against them

Forensic AI in action

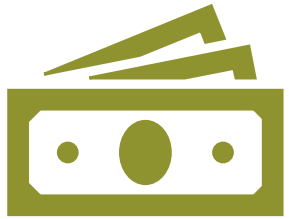




Challenges of Forensic AI

Accuracy

Transparency



Costs

Co-funded by the Erasmus+ Programme of the European Union

Accuracy

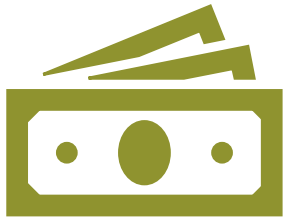
- Forensic AI assesses reliability of evidence potentially manipulated, who assess the reliability of the Forensic AI?
- Study by Agarwal S and others (2019)
  - Compared their tool to FaceForensics++
  - Results: less accurate, especially with lip-sync deep fakes
- False positive → wrongful convictions
- Accuracy relates to admissibility of evidence
  - Avoid ‘presumption of reliability’ of Forensic AI
  - Careful scrutiny before entering criminal trial

Co-funded by the Erasmus+ Programme of the European Union

Transparency

- Gradable transparency (not only source code!)
- Transparency relates to confrontation phase
  - A human expert can be questioned
  - Can you put AI on the stand?
- Access to technical info on the Forensic AI tool
  - Tests, standards applied, validation
  - Accuracy levels
  - Info on training dataset

Co-funded by the Erasmus+ Programme of the European Union



Costs

- Access to experts and trained lawyers
  - UK family law case
  - Lawyer and his experts analysed the recording's metadata
  - "The judge was really shocked. It would have never occurred to him to look into that"
- High costs for experts and tools
  - Legal aid may not cover them!
  - Less likely cover costs for buying Forensic AI tool
- Divide between public prosecutors and defendants, and among defendants, based on their economic status
- Ensure fair access to experts and tools (principle of equality of arms)

Co-funded by the Erasmus+ Programme of the European Union

The way forward: the AI Act

- Proposal COM(2021) 206 final for an «AI Act»
- «Work in progress» (expected amendments by MEPs in May)

- **Forensic AI to detect deep fakes**
  - Pre-designated high-risk system (Annex III, 6(b) AIA)
  - *AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3) AIA*

- Therefore
  - Compliance with Requirement (Ch III) and Obligations (Ch IV)
  - CAP with internal control + EU Declaration of Conformity + CE marking of conformity
  - Registration in the EU Database

Co-funded by the Erasmus+ Programme of the European Union

The way forward: the AI Act

- AI Forensic follows requirements in Chapter III
  - Including risk management system, technical documentation, record-keeping, transparency to users, human oversight, accuracy, robustness and cybersecurity

✓ Accuracy challenge?
- Relieves burden on judges in the admissibility phase

✗ Transparency challenge?
- If used by public prosecutors –> transparency not enforceable for defendants
- Exceptions in law enforcement field: e.g. no instructions in the database (Annex III, points 1, 6 and 7)
  - Important info including level of accuracy, foreseeable risks or misuse, performance as regards the persons or groups, information on training, validation, testing data sets used (art. 13 AIA)

Co-funded by the Erasmus+ Programme of the European Union

EUI DEPARTMENT OF LAW

The way forward: can providers help?

Accuracy challenge?

Best equipped to test software's accuracy

State clearly if AI tool can be used for law enforcement purposes

Transparency challenge?

Open to share technical documentation with defendants

Protected forms of disclosure (NDA, in-camera hearings)

Costs?

Free license for limited time period to defendants

Technical and expert support to defendants

Co-funded by the Erasmus+ Programme of the European Union

Conclusions

Huge potential for Forensic AI to detect Deep Fakes in criminal proceedings

Need strong legal safeguards for challenging Forensic AI

Software providers can help increasing transparency and reducing costs for defendants

Co-funded by the Erasmus+ Programme of the European Union