doughty street chambers

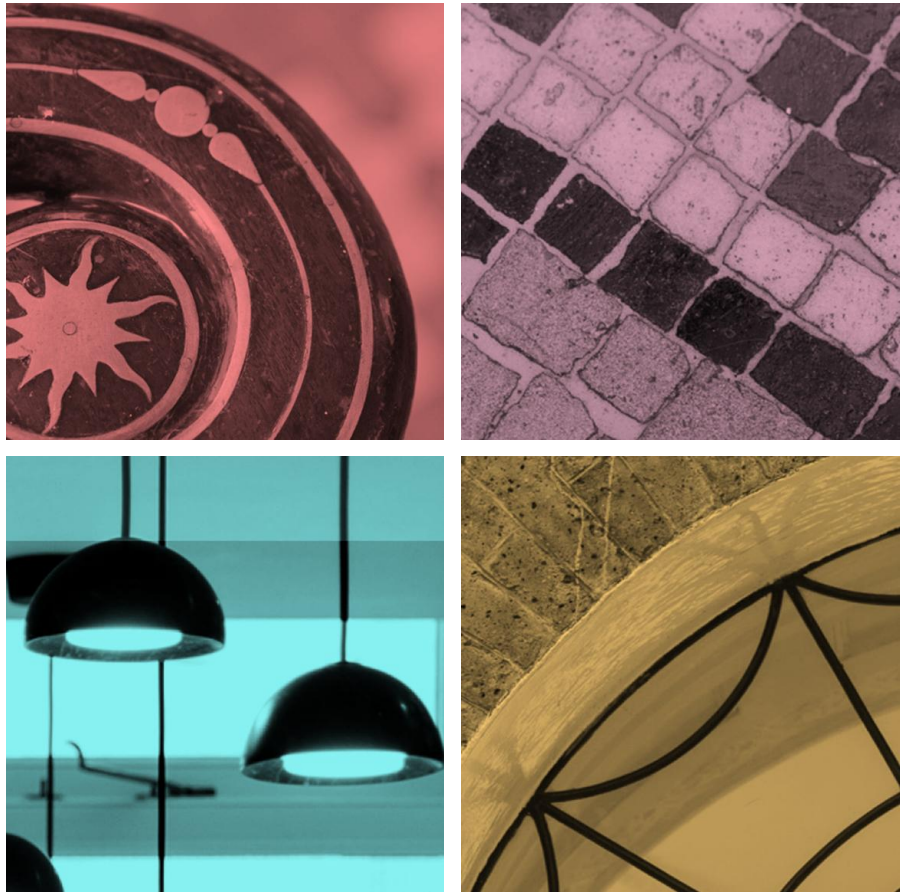
ARTIFICIAL INTELLIGENCE IN THE CRIMINAL JUSTICE SYSTEM

FARRHAT ARSHAD KC
Barrister
Doughty Street Chambers

KARLIA LYKOURGOU
Barrister
Doughty Street Chambers

HARRIET JOHNSON
Barrister
Doughty Street Chambers

KIM LAWRIE
Global Group Head of Technology
Beyond



AGENDA

Part I — AI Deepfakes

Law, remedies & 2025 reforms

Part II — Fake Evidence

Defendants, complainants & the backlog

Tonight's Focus: Two Ways AI Is Reshaping Criminal Justice

A common thread: the law is struggling to keep pace with the technology.

I

AI-Generated Deepfakes

- Who is affected?
- Criminal & civil remedies
- The 2025 legal reforms
- Remaining gaps

II

Fake Evidence in Crown Court

- What does fake evidence look like?
- Impact on defendants
- Impact on complainants
- Impact on the backlog

PART I

AI-Generated Sexually Explicit Deepfakes

Law, Remedies & the 2025 Reforms

What Are Deepfakes?

AI-generated images, audio, or video that fabricate a realistic portrayal of a real person doing or saying something they have never done.

Freely Available Technology

Creating a convincing explicit deepfake — superimposing a person's likeness onto another body — requires only a handful of source images. In many cases, no technical expertise is needed at all. Kim will demonstrate shortly.

Voice Cloning

AI can replicate a voice from short recordings. HMRC and financial services increasingly use voice as a security check — a significant and growing vulnerability.

Not Just Sexual Content

Deepfakes are relevant to fraud, evidence fabrication, and coercive or controlling behaviour — not only explicit imagery. This is a technology challenge for the whole of criminal practice.

The Pre-2023 Legal Landscape

BEFORE 2023 — The Gap

- Liability arose only on sharing or threatening to share — NOT at point of creation.
- Law Commission recommended against criminalising 'making' — harm was seen as arising from publication, not existence.
- A person could create unlimited explicit deepfakes and commit no offence, so long as they were not shared.
- For victims, the mere existence of such images was deeply distressing — and could feed corresponding offending: stalking, harassment, coercive control.
- NB: For children, 'making' an indecent image was already an offence. The asymmetry with adult victims was stark.

ONLINE SAFETY ACT 2023 — s.66B SOA 2003

- Inserted s.66B into the Sexual Offences Act 2003.
- Offences of sharing or threatening to share intimate images — including deepfakes — without consent.
- A meaningful step — but the creation gap remained entirely untouched.
- Creating as many explicit deepfakes as desired, while keeping them private, was still lawful.

The 2025 Reforms: Criminalising Creation

Data (Use and Access) Act 2025, s.138 → inserting ss.66E & 66F into the Sexual Offences Act 2003

s.66E SOA 2003

Creating a Purported Intimate Image

- Intentionally creating an image without consent
- No reasonable belief in consent required for offence
- 'Purported intimate image': appears to show an adult in an intimate state, whether or not it is a real photograph
- Does not require the image to have been shared

s.66F SOA 2003

Requesting Creation of Such an Image

- Companion offence — captures those who commission deepfakes
- Recognises harm does not only arise from the person who operates the tool
- Targets those who solicit creation from online platforms or third parties

Extended limitation period
Up to 3 years from offence, or 6 months from sufficient evidence coming to the prosecutor's knowledge

Deprivation orders
Courts may order that the offender be stripped of the image and any device containing it

Remaining Gaps & Civil Remedies

PRACTICAL CHALLENGES IN CRIMINAL PROSECUTION

Identification

Tracing who created an image — especially via anonymous platforms or overseas-hosted tools — remains technically challenging.

Proof of Intent

Proving the defendant had no reasonable belief in consent is likely to remain an issue, particularly in domestic cases.

Platform Enforcement

Deepfakes shared person-to-person or hosted overseas may be removed slowly, if at all, despite Online Safety Act obligations.

CIVIL CAUSES OF ACTION

Defamation (Defamation Act 2013)

Falsely portrays the claimant in a way that damages reputation

Misuse of Private Information

Tort protecting intimacy and dignity — particularly apt in a sexualised context

UK GDPR / Data Protection Act 2018

Processing personal data (including photographic images) without a lawful basis

Copyright

Where victim is the copyright owner of the source images used to generate the deepfake

PART II

Fake AI-Generated Evidence in Crown Court Trials

And what it means for judges, defendants, complainants & a backlogged system

What Does 'Fake Evidence' Look Like?

Deepfake Video

Fabricated CCTV footage placing a defendant at a scene they never visited. Highly convincing, increasingly accessible without technical expertise.

AI-Cloned Audio

Fabricated voice recordings of incriminating conversations, indistinguishable from authentic intercepts without expert analysis.

Manipulated Images

Realistic fabricated photographs — or, conversely, the denial of genuine images as potentially AI-generated.

Fabricated Documents

AI-generated emails, texts and documents with plausible metadata — and, as in Ayinde [2025], entirely hallucinated legal citations.

The Impact on Defendants

RISK 1 — Conviction on Fabricated Evidence

A defendant may face prosecution on the basis of AI-generated evidence that is entirely false: a deepfake video placing them at the scene, fabricated audio of an incriminating conversation, or a falsified document trail.

Challenging such evidence will usually require prior authority — at best, further delays; at worst, refusal. The risk of wrongful conviction is real at all levels, but particularly in the Magistrates' Court where summary justice rarely allows the level of scrutiny this evidence demands.

RISK 2 — The 'Liar's Dividend'

Because deepfakes exist and are increasingly sophisticated, defendants may challenge the authenticity of entirely genuine prosecution evidence — CCTV footage, recorded intercepts, digital documents — on the basis that it could have been fabricated. Already observed with Jan 6th defendants in the US, despite that being one of the most contemporaneously documented events in criminal history. Each such challenge increases the evidential and forensic burden on an already over-stretched prosecution.

The Impact on Complainants

Evidence Challenged as Fabricated

In sexual offence cases especially, audio or video evidence captured on a phone — a voice message, a recording of an incident — may be challenged as AI-generated, further undermining the complainant's account in an already adversarial process.

Deepfakes Used Against Complainants

A defendant may submit deepfake material to cast doubt on the complainant's credibility — portraying them in a false light, or fabricating communications attributed to them.

Detection Tools Not Yet Reliable

Existing tools for detecting deepfakes are not sufficiently reliable to provide certainty. Authentication disputes will add significantly to the distress of complainants during proceedings that are already extremely difficult.

Impact on the Crown Court Backlog

The Crown Court is already in profound crisis. AI-related satellite litigation risks making it significantly worse — yet the alternative raises serious Article 6 ECHR fair trial concerns.

The Risk of Delay

- Authentication disputes generate satellite hearings pre-trial
- Expert witnesses required — scarce and expensive
- Prior authority needed to challenge AI evidence
- An already log-jammed system faces further pressure

The Risk of Injustice

- Admitting unscrutinised digital evidence risks wrongful conviction
- Art. 6 ECHR fair trial right at stake on both sides
- No dedicated evidential standard in England & Wales
- Contrast: US Federal Rules of Evidence reform proposals actively under consideration

Ayinde [2025]: Five AI-hallucinated case citations submitted to the High Court → wasted costs orders and referrals to BSB & SRA. Hallucination, not deliberate forgery — but illustrates how quickly AI outputs are treated as reliable without verification.

The Government's Response & What Comes Next

What Has Been Done

- Judiciary AI Guidance (Oct 2025): Judges warned of deepfakes and white text; early authentication challenges encouraged at case management.
- Leveson: AI placed at the centre of justice reform; structured frameworks for responsible deployment recommended.
- £12m Justice AI Unit funded to support courts in managing AI-related challenges (per Lammy, Feb 2026).

What Remains Outstanding

- No dedicated evidential standard in England & Wales for AI-generated material.
- US under active consideration (FRE reform proposals) — England & Wales has recommendation but no statutory framework.
- Platform enforcement and cross-border jurisdiction remain unresolved in practice.
- Practical burden of authentication falls on already over-stretched practitioners and courts.

Conclusion

The 2025 reforms addressing deepfake creation are welcome — but gaps in identification, intent, and enforcement remain formidable.

Fake evidence raises questions that go to the heart of the system's ability to deliver accurate verdicts — for defendants and complainants alike.

These challenges fall simultaneously on defendants, complainants, and institutions that were already under enormous strain before AI entered the picture.

AI literacy is now a core professional competency. The obligation to verify AI outputs, understand deepfake capabilities, and advise clients accordingly is already upon us — whether or not formal rules have caught up.