36 STONE

5 Days, 5 Disputes

13 August 2025

# Day 3

by Paul Schwartfeger

## When AI Spills Secrets: Breach of Confidence

Generative AI (GenAI) systems learn from vast datasets, many drawn from public sources, though some from material that was never intended for public use.

In 2023, Samsung engineers reportedly pasted proprietary semiconductor source code and meeting notes into ChatGPT to debug errors and produce summaries, placing confidential microchip test sequences in the hands of OpenAI. The incident shows how, once sensitive data is submitted to a public model, control can be lost, along with any realistic prospect of keeping it secret. If such material enters a model's training data, whether through lax controls, error or deliberate inclusion, it may reappear, potentially breaching contractual duties or equitable obligations of confidence.

### Applicable legal principles

In English law, the foundation for such claims is the equitable doctrine of breach of confidence, as in *Coco v A N Clark (Engineers) Ltd* [1969] RPC 41. The claimant must show the information has the necessary quality of confidence. It must also have been imparted in circumstances importing an obligation of confidence and there must be unauthorised use or disclosure causing detriment. Each element raises AI-specific challenges.

Public domain information will lack the quality of confidence, though presence in a secure dataset may suffice. A compilation of partly public information may still qualify.

Courts will consider if those imparting the information preserved its secrecy, and whether the AI developer knew or should have known its nature. Arguments may revolve around privacy settings and terms, or whether a model was trained on materials obtained under an NDA or scraped from a secure but poorly protected repository.

The most difficult step is likely to be proving "disclosure". Outputs may not be identical reproductions, though could embed factual details or unique phrasing that betrays the source. Where outputs are shaped by user prompts, disclosure may result from both the system's design and the human interaction. This dual causation can complicate claims, with parties seeking to shift responsibility.

### Potential defendants

Liability might attach to the AI provider for training the system on confidential data; the deploying organisation for using the tool in a way that causes disclosure; or an end user for republishing the material.

### Evidence and interim remedies

The key task is proving that confidential material entered the training set or reappeared in an output in circumstances importing an obligation of confidence. That means not only evidencing its presence, through prompt and output logs, dataset snapshots and source files, but also showing it retained the necessary quality of confidence and was misused without consent. This may require scrutiny of your own system architecture to understand how the information was obtained.

The presence of the material in the model may be established through disclosure or, if the respondent is not a party, a Norwich Pharmacal order.

Since one disclosure can make secrecy impossible, injunctions are essential, with courts prioritising prevention over damages.

### Potential defences

Defendants may argue material lacks the quality of confidence, was already public or was disclosed with consent. They may also argue any similarities are coincidental, the result of statistical generation rather than a reproduction from stored data.

A defendant might also argue that the material entered the training data without their knowledge or fault, though this may be undermined where large-scale scraping makes access to the source material foreseeable, if not inevitable. Weak data-governance controls may be said to amount to constructive knowledge.

### Remedies

Remedies include injunctions to prevent further use or disclosure, delivery up or destruction of materials, as well as damages or an account of profits.

In urgent cases, the court may grant "springboard" relief (*Terrapin Ltd v Builders' Supply Co (Hayes)* [1967] RPC 375) to prevent a party gaining an unfair competitive advantage from misuse.

### Conclusions

In the AI era, "disclosure" may now be triggered by nothing more than a prompt surfacing a latent fragment in the model. Winning these cases demands the ability to trace the fragment through logs, datasets and models, and to translate those findings into a compelling claim against the right defendant before the information escapes.

## 5 Days, 5 Disputes

Inspired by the release of OpenAI's GPT-5 and the rapid evolution of tools like it, *5 Days, 5 Disputes* highlights five types of legal dispute where artificial intelligence is testing established legal principles, offering insights for those handling AI claims.

## Paul Schwartfeger

Paul is a barrister with 36 Stone, specialising in commercial litigation and international arbitration with a particular focus on legal disputes involving data and technology.

psc@36stone.co.uk
+44 (0) 20 7440 6900

The law stated is that of England and Wales as at the article date. Always seek specific legal advice from a qualified solicitor or barrister on your individual case and issues.