Day 2
by Paul Schwartfeger

When AI Copies Too Closely: Copyright Disputes

Generative AI (GenAI) systems can produce striking images, convincing prose and complex code in seconds. But behind their polished outputs lies a process heavily dependant on the work of others. Large language models (LLMs) and image-generation systems are trained on vast datasets to learn statistical patterns in existing material. This allows them to generate works that can closely resemble or even replicate their training data.

Disputes are already surfacing. Visual artists have sued Stability AI alleging its AI outputs reproduce their work. In the software field, GitHub's CoPilot has faced litigation claiming it produces near-verbatim code from public repositories. These cases illustrate that copyright risk in AI is not theoretical, and it can attach to both the *training* process and the *outputs*.

Applicable legal principles
Under the Copyright, Designs and Patents Act 1988 (CDPA), copyright subsists automatically in original literary, dramatic, musical or artistic works that involve the author's own skill, labour and judgement. Infringement occurs when a substantial part of a protected work is reproduced without permission. "Substantial" is to be assessed by quality as well as quantity (*Designers Guild Ltd v Russell Williams (Textiles) Ltd* [2001] 1 WLR 2416).

Direct copying may occur if a model reproduces near-identical portions of its training data. More subtly, "substantial" elements may emerge if the distinctive expressive features of a work appear (*Infopaq International v Danske Dagblades Forening* (C-5/08 [2010] FSR 20)).

In an AI context, two particular evidential hurdles arise. First, without disclosure from model providers, it can be difficult to prove that an original work was ingested during training. Second, because many models use stochastic sampling, identical prompts can produce very different expressions, making causal links harder to establish.

Where a claimant can prove both access to a work and a causal connection, unconscious copying principles may apply (*Francis Day & Hunter Ltd v Bron* [1963] Ch 587), meaning intent is not required.

Potential defendants
Defendants may include an AI provider, if the claim focuses on training processes; a deploying entity, if it uses the model to produce infringing content; or an end user who republishes infringing outputs. Joint liaility or authorisation infringement under s.16(2) CDPA may be relevant here.

Evidence and interim remedies
A Norwich Pharmacal order may be sought to compel non-parties to disclose records showing whether the claimant's work was in the training set, along with prompt and output logs linking a user's request to the infringing material. If the AI developer is a defendant, similar material can be sought through disclosure.

Potential defences
Defences include independent creation, where a defendant argues the output is sufficiently original and does not copy the claimant's work; fair dealing for criticism, review, quotation or reporting current events, though these exceptions are narrowly construed in commercial contexts; incidental inclusion, where a protected work's appearance is merely incidental to the output's main purpose; and the use of functional or unprotectable elements, which is particularly relevant for software design. Text and data mining exceptions under the CDPA may also be pleaded, although these are narrower than many AI training scenarios will be. A defendant may also contend that outputs are generated probabilistically without storing protected works in a fixed form. Claimants will counter that statistical representations still infringe.

Remedies
Remedies include damages or an account of profits, injunctions to restrain further infringement, and delivery up or destruction of infringing copies. Additional damages under s.97(2) CDPA are available for flagrant infringements.

Conclusions
The opacity of AI training and variability of AI outputs make copyright claims complex but far from impossible. Substantiality and unconscious copying doctrines remain powerful tools. Judgment is awaited in *Getty Images (US) Inc and others v Stability AI Ltd* [2023] EWHC 3090 (Ch), the first UK case to address large-scale ingestion of copyrighted works for AI training. Its outcome could reshape how infringement claims against AI developers, deployers and users are framed. Those with potential exposure should be reviewing their risk profile now.

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