• @DoughtyStreet

www.doughtystreet.co.uk

# RESPONSIBILITY FOR AI-GENERATED DEFAMATION




doughty street chambers

By Sophie Mitchell

# AI-GENERATED HARM:

## How do you control - and overcome - difficulties with AI-generated reputational harm, including fake biographies, satirical memes and hallucinated allegations. Who do you sue?

**By Sophie Mitchell**

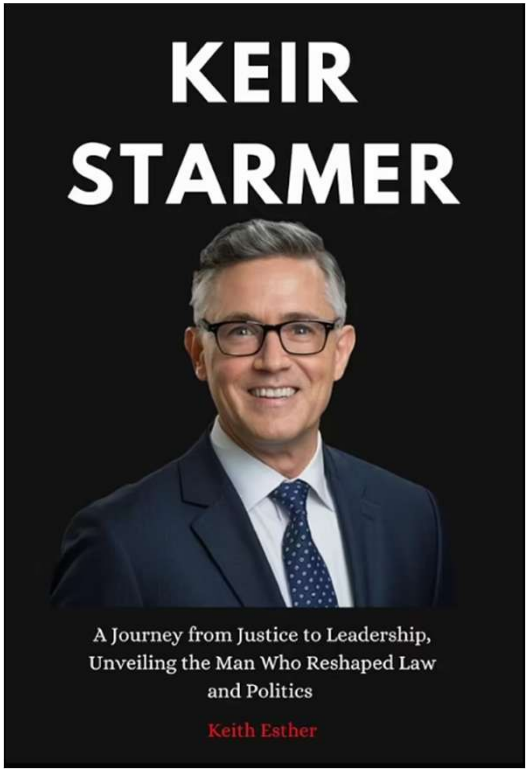
**Doughty Street Chambers**

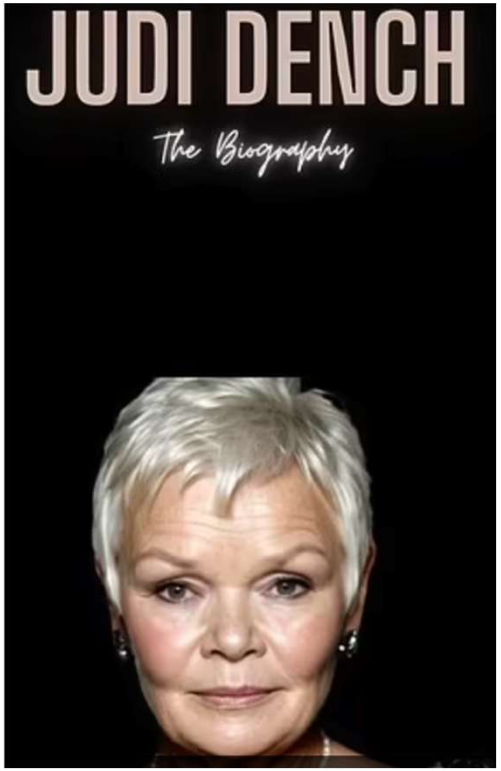
# THE MAIN ISSUE: HALLUCINATIONS

Hallucinations are responses produced by generative AI that contain false or misleading information presented as fact.

Fake biographies and factually incorrect satirical memes are just two examples of the many, many ways that hallucinations may manifest in reputationally damaging publications.

# FAKE BIOGRAPHIES






# RECENT HALLUCINATION INCIDENTS:

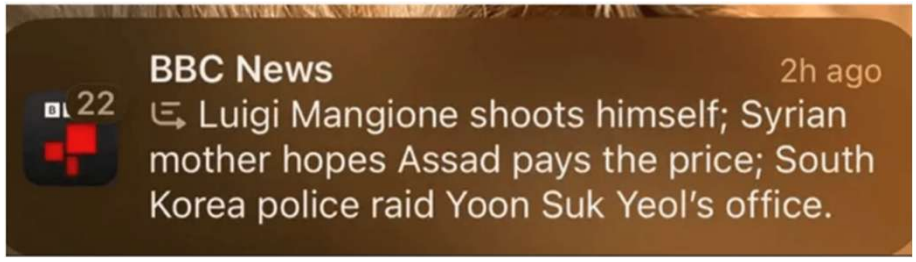
2023: **OpenAI's tool, ChatGPT**, falsely stated that US Professor **Jonathan Turley** had sexually harassed students on a class trip to Alaska. Chat GPT cited a 2018 Washington Post article in support, which never existed. There had been no trip to Alaska and no such allegations made.

2024: **Microsoft's AI tool, Copilot**, falsely described German journalist, **Martin Bernklau**, as being involved in several serious forms of criminality. In fact, these alleged criminal actions related to court cases which Mr Bernklau had reported on in the course of his employment as a journalist.

2024: **Apple's** rollout of its AI news tool which mis-summarised news reports and pushed the inaccurate news alerts out to iPhone users. An AI-generated summary of the headlines falsely told some readers that Luigi Mangione, the man accused of killing United Healthcare CEO Brian Thompson, had shot himself:

# RECENT HALLUCINATION INCIDENTS:



# LITIGATION

US defamation cases involving content produced by generative AI tools:
**- Walters v OpenAI;**
**- Starbuck v Meta Platforms.**

There are currently no defamation cases from this jurisdiction. There is a High Court MPI & data claim underway (not yet determined):
**- Jessica Asato MP v X.AI LLC** - widely published deepfake sexualised images of C on 'X', created by 'Grok', a generative AI chatbot developed, owned and operated by 'X'.

# EXAMPLE SCENARIOS

Scenario A

**Mr C**, is a **newspaper columnist in London**, who writes immigration stories. Mr C becomes involved in an **argument on X** (**formerly Twitter**) **with a television presenter**, **Ms E,** about immigration policy. Ms E **entered the prompt 'Tell me about Mr C' into OpenAI's ChatGPT,** which responded with: "in September 2024, Mr C was **convicted of affray for participating in the anti-immigration Southport riots** and sentenced to **six months imprisonment**".

(Mr C has **no such conviction**. He **did not attend the Southport riots**).

**Ms E has 500,000 followers on X**. Ms E **posts on X a screenshot taken from a conversation she had with ChatGPT**.

**As a result**, Mr C receives **hundreds of abusive messages on X**, had **his newspaper column is suspended**. He is diagnosed with an **anxiety disorder**.

**Mr C wants to sue OpenAI for the harm caused to his reputation**.

# EXAMPLE SCENARIOS

Scenario B

A **technology company's AI news tool** summarises breaking news items from third-party news outlets and **sends them in the form of push notifications to smartphone users**.

**Five million people** in the UK **receive a notification** to their phones stating that a **London-based television presenter, Mr J**, has been **arrested and charged with child sex offences**.

(Mr J has **never been arrested, charged or even investigated**).

The notification prompts **protests outside Mr J's house**, he receives **death threats** and suffers a **mental breakdown**.

**Mr J wants to sue the technology company for the harm caused to his reputation**.

# WHO TO SUE

Potential Defendants – Scenario A:
The developer, owner and operator of ChatGBT (to discuss);
Ms E;
'X' (upon notification & failure to act);
Anyone else who re-posted on X or elsewhere.

Potential Defendants – Scenario B:
The developer, owner and operator of the AI news tool (to discuss);
Anyone else who re-published the comment.

# DEFAMATION ACTIONS INVOLVING THE AI TECHNOLOGY COMPANY: ISSUES LIKELY TO ARISE

It remains to be seen whether, and if so, on what basis, liability could attach to the developer, owner and operator of the generative AI product. The position will depend upon the AI application in question.

*Practical Tips:*

*Some AI companies may not develop, own and operate the AI tool in question. Establish how any multi-entity structure operates;*
*Assess the Terms & Conditions for the product in question (disclaimers, etc);*
*Ascertain model type, version, contemporaneous literature explaining the reliability of it and how 'reliable' as a product it has been marketed to the public;*
*Claimants will find themselves facing Defendants with greater technological expertise at their fingertips, so may need to instruct or consult an expert early.*

# PERSONAL JURISDICTION

Whether OpenAI (in Scenario A) or the technology company (in Scenario B) could be an 'author', 'editor' or 'publisher' of the statement

Section 10 of the Defamation Act 2013:
(1)A **court does not have jurisdiction** to hear and determine an action for defamation brought against a person who was **not the author, editor or publisher** of the statement complained of **unless** the court is satisfied that it is **not reasonably practicable** for an **action to be brought against the author, editor or publisher**.

(2)In this section **"author", "editor" and "publisher"** have the **same meaning** as in **section 1 of the Defamation Act 1996.**

Section 1(2) of the Defamation Act 1996

"**author**" means the **originator** of the statement, *but does not include a person who did not intend that his statement be published at all;*

"**editor**" means a person having **editorial or equivalent responsibility** for the content of the statement or the decision to publish it; and

"**publisher**" means a **commercial publisher**, that is, a person whose business is issuing material to the public, or a section of the public, who issues material containing the statement in the course of that business.

Section 1(3) of the Defamation Act 1996

(3)A person shall **not be** considered the **author, editor or publisher** of a statement if he is only involved—

(a)in printing, producing, distributing or selling printed material containing the statement;

(b)in processing, making copies of, distributing, exhibiting or selling a film or sound recording (as defined in Part I of the Copyright, Designs and Patents Act 1988) containing the statement;

(c)in processing, making copies of, distributing or selling any electronic medium in or on which the statement is recorded, or in operating or providing any equipment, system or service by means of which the statement is retrieved, copied, distributed or made available in electronic form;

(d)as the broadcaster of a live programme containing the statement in circumstances in which he has no effective control over the maker of the statement;

(e)as the **operator** of or **provider** of **access to a communications system by means of which the statement is transmitted, or made available, by a person over whom he has no effective control**.

In a case not within paragraphs (a) to (e) the **court may have regard to those provisions by way of analogy** in deciding whether a person is to be considered the author, editor or publisher of a statement.

In *Tamiz v Google [2013] 1 WLR* at §39, the Court of Appeal held that Google Inc was not a publisher in the context of it hosting its blogger platform, because it **did not 'issue' material to the public**, but instead **provided access to a system** by means of which statements were transmitted.

# RESPONSIBILITY FOR PUBLICATION

Whether OpenAI (in Scenario A) or the technology company (in Scenario B) could be held liable for their respective publications

The general rule is that everyone "who knowingly takes part in the publication of a libel, or authorises or ratifies it" is jointly and severally liable (*Monir v Wood* [2018] EWHC 3525 (QB), §135(i)). However, the common law distinguishes between primary and secondary publishers. To be liable as a primary publisher, a person must be shown to have "knowing involvement in the publication of particular words".

A secondary publisher can rely on the common law defence of innocent dissemination, if they, amongst other things, can show that they neither knew, or had any reason to know, the content of the publication.

# RESPONSIBILITY FOR PUBLICATION

Whether OpenAI could be a publisher at common law - Analogous situations

In *Bunt v Tilley [2006] EWHC 407 (QB),* Eady J summarised the **common law position** on responsibility for **primary publishers**:

"[22] ...to impose legal responsibility upon anyone under the common law for the publication of words it is **essential to demonstrate a degree of awareness or at least an assumption of general responsibility,** such as has long been recognised in the context of editorial responsibility. [...]

[23] Of course, to be liable for a defamatory publication it is **not always necessary to be aware of the defamatory content**, still less of its legal significance. Editors and publishers are often fixed with responsibility notwithstanding such lack of knowledge. On the other hand, for a person to be held responsible there must be **knowing involvement in the process of publication of the relevant words**. It is **not enough that a person merely plays a passive instrumental role** in the process".

# RESPONSIBILITY FOR PUBLICATION

In *Monir v Wood*, Nicklin J similarly summarised the **distinction between primary and secondary liability** as a matter of **knowledge,** as well as **control** [at 178]:

The fundamental distinction between primary and secondary liability is **knowledge** that the **publication contains some defamatory matter** and the **ability to control** its publication or continued publication.

# RESPONSIBILITY FOR PUBLICATION

*Metropolitan International Schools Ltd v Designtechnica Corp & Ors [2011] 1 WLR 1743*, Eady J:

"When a search is carried out by a web user via the Google search engine it is clear, from what I have said already about its function, that there is **no human input from [Google]. None of its officers or employees takes any part in the search**. It is performed automatically in accordance with computer programmes… … … …**There being no input from [Google] therefore…… it cannot be characterised as a publisher at common law**. It has **not authorised or caused the snippet to appear** on the user's screen in any **meaningful sense**. It has **merely**, by the provision of its search service, **played the role of a facilitator**."

At [50, 51] (see also [11], [12], [53] and [74]).

# RESPONSIBILITY FOR PUBLICATION

*Tamiz v Google [2013] 1 WLR 2151*

At [20–25], the Court of Appeal found no "fundamental error of analysis" or wrong conclusion in Eady J's approach to the internet service in [*Metropolitan*], and tacitly approved his view that common-law responsibility for publishing defamatory words **required "a mental element**".
"…it does not create the blogs or have any prior knowledge of, or effective control over, their content" and was not, therefore "in a position comparable to that of the author or editor".

Notification
In *Tamiz*, the Court held that Go**ogle would become a publisher at common law after it had been notified** of the defamatory publications and been given a **reasonable amount of time to remove it**, **but failed** to do so [at 34].

# RESPONSIBILITY FOR PUBLICATION

*Oriental Press Group Ltd v Fevaworks [2018] HKCA 69* and the Singaporean case of *B2C2 Ltd v Quoine Pte Ltd [2019] SGHC(I) 03* at [210]; on appeal at *[2020] SGCA(I) 02* at [97]-[103].

"Where a **developer is defining the inputs, the output parameters and the circumstances of publication**, it is no **longer a mere facilitator** and **arguably should be regarded as the primary publisher,** with the **intent traced back** to the **development process** and the **decision to set the AI into operation**.

**Alternatively**, the operator may be regarded as an **editor**, a role that is defined by statute as "a person having editorial or equivalent responsibility for the content of the statement or the decision to publish it."

Thank you.

Sophie Mitchell
Doughty Street Chambers