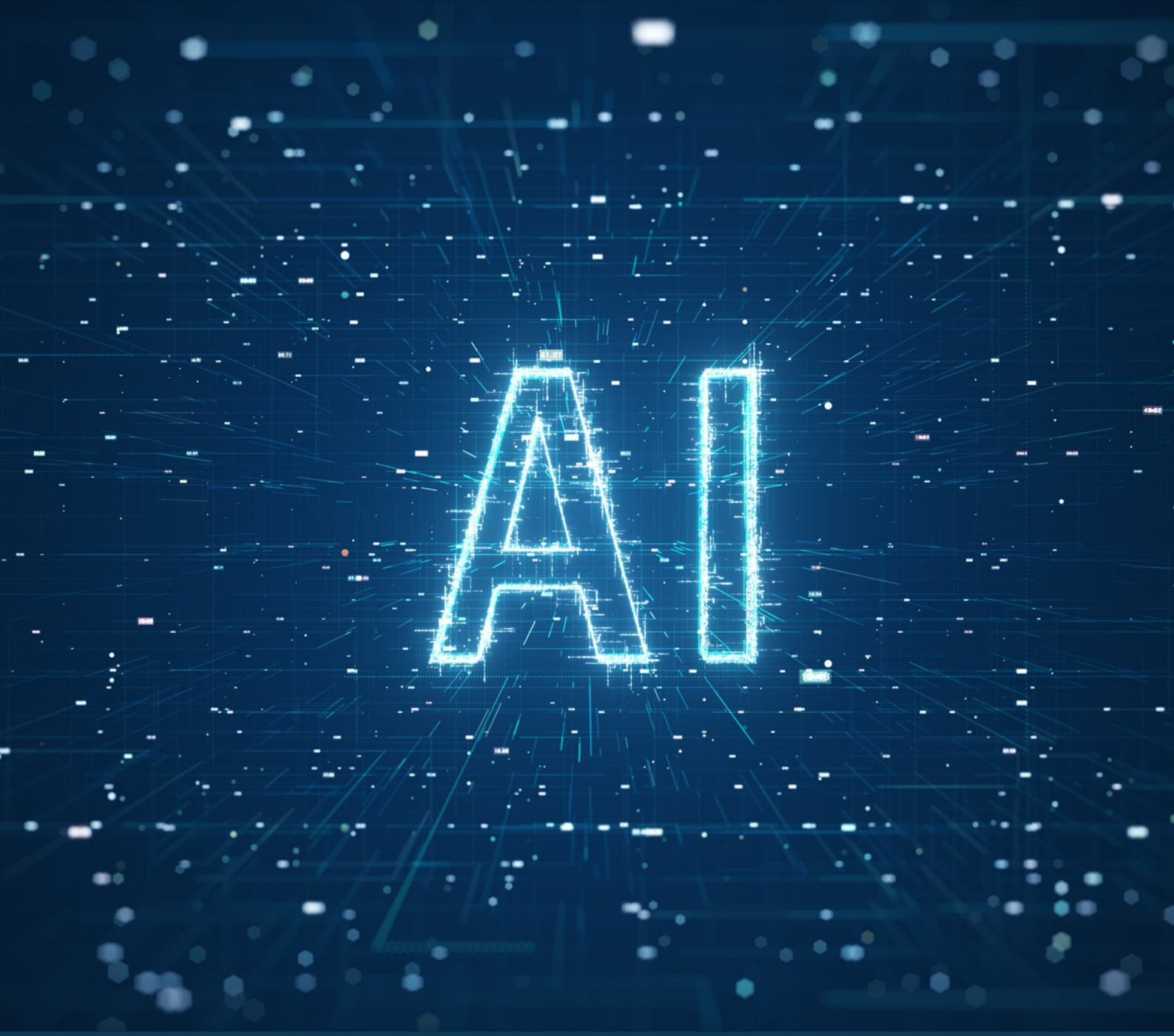

ARTIFICIAL INTELLIGENCE & REPUTATION MANAGEMENT

What clients need to know
What clients can do

Carter-Ruck

DRD PARTNERSHIP

In 2017, the late Stephen Hawking expressed his fear that "AI may replace humans altogether. If people design computer viruses, someone will design AI that replicates itself. This will be a new form of life that will outperform humans".

Hawking's prediction of generative AI has quickly come to pass. Whether the machines outperform and replace humans remains to be seen. However, one immediate impact of the technology is its potential to have an impact on reputation.

AI is capable of generating its own version of the truth. In doing so, it is mostly relying on an Internet already littered with factoids of scant veracity.

For those to whom the truth matters, this is a problem, but not one without a solution.

Even before the legislative and regulatory framework surrounding AI manages to catch up with the pace of technological change, effective legal remedies and communications strategies exist, and there are clear steps that clients can take now to protect their reputations.

Generative AI - generating problems?

Artificial intelligence (or "AI") is the creation of machines to perform cognitive functions more commonly associated with the human mind. AI has in fact been around for decades, and features in online tools deployed by the technology industry (such as Google searches and banking software), which most individuals in the developed world use every day.

The turning point, which is the cornerstone of the current heightened debate, has been the development of Generative AI. This term relates to the use of AI to assimilate pre-existing information or input, used to create new or derivative content in the form of words, images and music. Chat GPT, the language model-based chatbot, DALL-E-2, the image generation system and Jukebox, a music creation device are all Generative AI tools developed by OpenAI, and all are now busy creating content entirely through AI. This is leading not only to an increased volume of content, but also to a lack of demarcation between human-created content and AI-created content. There have even been reports of AI "hallucinating", that is to say, providing inaccurate answers based on impertinent information available to it.

This and other developments have prompted a slew of legal questions, and even claims, about who is responsible for these new machine content creators and the material that they create.

Input, system, output - whose AI is it anyway?

A big question is who will claimants tackle when it comes to claims over AI?

Is the problem at the **input** stage, where information and data are fed to an AI system, from which the system then learns and creates? The input may be an unidentified individual, upon whose words and directions the machine acted. Or is the real challenge the **system** itself, the algorithm, the code, and the programming? The system itself belongs to the AI creator, who may be liable for its workings. Or is the issue the generated **output**, released to the world and published, republished, relied upon and consumed?

The good news is that there is no right answer to that question. Each case will turn on its facts, and there are a number of individuals or companies who may be liable for the output of generative AI, which attracts a degree of flexibility that may even be beneficial for claimants.

The future is already with us: search, social media and online set the tone

A key question for clients is has AI actually moved the goalposts in terms of the challenges that already existed.

The truth is that publication in the 21st century is already a complex and multi-layered arena. Print copies of newspapers – which historically one could rely on derive articles from largely accurate sources, and where publication was relatively easily contained and remedies relatively straightforward to obtain – are no longer the primary source of news consumption.

Most news is consumed via social media. Aggregators and reposts on social media drive dissemination, while search engines mine and represent the choicest picks from what is out there online regarding an event, a person, or a piece of knowledge. The secondary sources not only inform journalists and the interested browser; they shape influential, but not necessarily reliable, compendium reference sites like Wikipedia. You can add to that a layer of clickbait, algorithmic 'pushing', other paid content and Search Engine Optimisation (SEO), designed to distract, divert and ensnare the questing searcher for the truth, or something half-resembling the truth.

The resulting search engine results page can be little more than a mish-mash of reportage, fact, factoid and the plain untrue. Reputations are already at stake, and the task of fixing the damage online can be significant, especially if a statement likely to cause harm is allowed to linger unchallenged.

More recently, problematic sources of "information" online have been amplified through the variety of channels and platforms. This is often driven by conspiracy theorists, smear campaigns or attempts to undermine or weaken an opponent or target party. In 2021, a report by the Center for Countering Digital Hate found that the vast majority of Covid-19 anti-vaccine misinformation and conspiracy theories originated from just 12 people. Religious extremism, cod science regarding the 5G roll-out and opposition to vaccines were common drivers behind campaigns that reached a direct following of 59 million. This is an audience reach which multiplies exponentially as soon as misleading content is reshared.

The lesson here is clear: even before the sudden widening of interest in and deployment of AI this year, online news and information was capable of bearing much content which is untrue, and spreading at a pace that has caught many unawares.

How can clients get "AI-ready"?

For clients – whether individuals or organisations – eager to police and safeguard their reputations, the addition of AI into the information space should be a signal to pause for thought, and to prepare to mitigate threats.

A critical consideration for those clients who have begun to use AI in their day-to-day operations is how AI is being used and how that use is being managed. It is vital that clients understand the legal framework involved in using AI, and certainly if you are responsible for the AI system **input** or **output**.

Crucially, you should never input your confidential or privileged information. This applies not just to your own data, but to the data of others you work with or for. It is of paramount importance if you work with sensitive data (for example if you are in the health, security, or education industries). Organisations should ensure that their staff members are adequately supported to use AI tools effectively and safely, and advised never to input someone else's data without consent. Clear policies should be in place; by way of example, companies such as Amazon, JP Morgan and Samsung have either limited or altogether banned the use of Chat GPT by their employees.

Organisations should also ensure that any contracted third parties or stakeholders follow the same approach - otherwise your confidential information may end up inputted into an AI system without your knowledge or consent. Ensure that the terms and conditions of third-party technology are properly scrutinised and kept under review. Remember that a company's terms and conditions of service often change; for example, earlier this year Zoom updated its terms which entitled the company to use customer data (including audio and video content) in the context of training its latest generative AI features without the need for consent. The change was discovered several months later, and after public outcry Zoom confirmed that it would not use customer content to train Zoom's or third-party AI models.

> If you do use AI within your organisation, be sceptical of its output and always question the veracity of any source information or material

If you do use AI within your organisation, be sceptical of its output and always question the veracity of any source information or material. Do not inadvertently become a publisher or disseminator of false and/or defamatory AI-generated information. Ensure that all AI-generated information is labelled as such and carries appropriate risk warnings, so that the end-recipient is kept fully informed.

What does an AI-driven reputation crisis look like?

SCENARIO ONE

THE STATESMAN

A newly elected MP has become aware of stories circulating on social media which include a false AI generated photograph of him smoking cannabis and a false AI generated police report by Marseilles police which alleges that he was detained in a French prison for alleged supply of cannabis. The stories are being sent anonymously by email to every sitting MP in Parliament. The MP believes his ex-staff member who was fired for disciplinary reasons is responsible for sending the emails, but he does not have any evidence to confirm his belief.

SCENARIO TWO

THE BIG SHORT

An unscrupulous investor wants to force down the share price of a household-name public company, to profit from a quick purchase and resale of equities. AI is used to generate a convincing recording of the chief executive leaving a voice message with the finance director, warning of the need to add a previously unannounced provision for losses in the forthcoming report and accounts. The recording is uploaded to a retail investor chatroom, and goes viral on TikTok.

SCENARIO THREE

A SCHOOL FOR SCANDAL

An embittered former pupil uploads an AI-generated grainy image purporting to show inappropriate conduct by a teacher already the subject of an internal investigation. Parents besiege the school with questions and some withdraw their children. Newspapers carry a redacted version of the image, which show the school's uniform. The police get in touch.

So what should you do if one of you or your organisation's reputation may have been negatively impacted by AI?

Legal strategies

First, you need to identify the target defendant(s). As mentioned above, you could, in theory, direct a complaint to the AI creator or user (whether at the input or output stage), as well as any subsequent publishers of the AI content. Which defendant is chosen will likely depend on a number of factors including the remedy desired, the speed at which resolution needs to be achieved and any jurisdictional hurdles either for commencing proceedings or enforcement.

If you cannot immediately identify the wrongdoer(s) (as in Scenario One), under English law, a claimant can seek what is called a Norwich Pharmacal order against a third party they believe holds information allowing them to identify a wrongdoer: this could in theory be used to compel an AI creator to reveal the AI user, or to compel a social media platform to 'unmask' an individual who has posted the AI content online.

Once you have identified your defendant(s), various laws and remedies may be applicable.

If the information published is false and has caused, or is likely to cause, damage to your reputation, then you may have a claim for defamation. Legal advice should be sought at an early stage to try to control the spread of the information. Remedies for defamation include damages, publication of a summary of any judgment and an injunction to prevent the continued or further publication of the falsity.

> "Legal advice should be sought at an early stage to control the spread of the information

Injunctions can also restrict or prevent publication of private information, regardless of its truth or falsity. The leading case of *McKennitt v Ash [2006]* makes clear that the truth of private information is "an irrelevant inquiry" about which judges should be wary of becoming side-tracked, and a misuse of private information can occur whether the relevant information is true or false. An AI-generated "deepfake" image such as the one described in Scenario Three therefore may fall within the category of private information.

Civil injunctions are also available to prevent harassment; to prohibit the pursuit of a course of conduct that causes a person alarm, fear or distress. This remedy could be relevant to each of the Scenarios above, and can specifically be used to prevent harassment by publication. Financial remedies may also be available where the causes of action of misuse of private information or harassment are made out.

Digital offences under the Computer Misuse Act 1990, Malicious Communications Act 1998, Protection from Harassment Act 1997 and even the Fraud Act 2006, have long provided the police powers to arrest individuals for illegal online communications. Interest in bettering our statutory framework to deal with such offences is only increasing, and the Online Safety Bill is nearing its final stages.

The so-called "right to be forgotten", now the "right to erasure" (enshrined in the UKGDPR and Data Protection Act 2018) enables individuals to seek to have inaccurate or outdated personal data de-listed by search engines, or deleted by the original publishers. Where AI-generated inaccurate and unreliable material is the subject of the delisting request (for example as in any of the scenarios identified above), provided it can be proved as such, the search engines and/or original publishers are unlikely to have a basis for refusing that request. Equally, it may be possible under the data protection legislation to require the creator of the AI system to delete your personal data from the system itself.

Strategic communications

Behind every good legal strategy is a good communications strategy, and those strategies dealing with AI-related content are no different. When it is suspected that AI-derived content may be driving reputational harm, this fact will, in and of itself, be of relevance to the communications strategy.

If the client has suffered reputational harm and AI has been a factor, then the immediate messaging for the response campaign should draw attention to this aspect, as it will drive positive carriage of the client's response.

> If AI has been a factor then the immediate messaging should draw attention to this

Longer term, the response campaign may want to strive to influence policy change which will simultaneously promote reputation rehabilitation. Singling out the technology as occupying a potentially harmful and unresolved position in the law will sharpen the attention of judicial figures and policy-makers as new caselaw and statute take shape.

The choice of channels for carrying the client's message will depend upon the legal remedies available and whether a swift rebuttal of harmful content can be secured. As is already the case, trusted and authoritative news sources such as wire services, public service broadcasters, and major editorial brands should be prioritised for external media.

Client controlled channels, including social media and websites, should be used to host rebuttals, fact-checkers, FAQs and verified data, which can be boosted through paid support, SEO and reposting.

Publicise PR and legal achievements, to the extent that they can be discussed openly, to illustrate the progress of the action, but only where this is on-strategy with the legal approach being taken.

Hope for the future

So where does one look in the face of an apparent existential crisis of technological development? The key message is: don't panic.

Importantly, English law has endured significant technological shifts in recent decades, and demonstrated its flexibility and resilience in dealing with new issues. While statutory changes can take some time to come to fruition, the courts are adept at using and developing long-standing principles to tackle new legal challenges. As witnessed in judgments on social media, data protection, intellectual property and digital offences, the courts have demonstrated that they have a toolkit – based upon strong foundational legal principles – through which they can remedy previously unheard-of legal dilemmas. Cases involving service of documents by social media, Google Adwords, targeting intermediaries for the sale of counterfeit goods online, or the 'right to be forgotten' litigation, all demonstrate the Courts' ability to fashion old law to suit new scenarios.

Further, in the face of mounting concerns among nation states, corporations and publics about the true dimensions of the challenge, the UK and EU are also introducing new statutory law to support the tools already in the hands of our judges. The EU has promised an AI Act with extra-territorial effect (much like the GDPR), while the UK's AI regulation will follow the Online Safety Bill, alongside proposed reform of data protection and privacy laws.

> "Clients should expect a growing train of new statute and regulation which will need to be factored into reputational management strategies

In contrast to its EU counterparts, the UK Government is currently taking an industry-first approach to AI regulation. Its March 2023 White Paper, *"A pro-innovation approach to AI regulation"*, posits a decentralised model of regulatory safeguards with sector regulators each applying common principles to the oversight of their respective industries. This model will be discussed at an autumn 2023 London summit on AI, which aims to make the UK a global leader in AI regulation.

Clients should expect a growing train of new statute and regulation in this field, which will need to be factored into reputation management strategies, in the same way that privacy and data protection laws have taken their place in the armoury.

From a technological standpoint, companies responsible for generative AI are heralding this as a new era of AI output. A real and exciting proposition is that, if AI generated material can be identified as such, then a lot of the legal uncertainty disappears.

If AI generators can be made to take responsibility for their content, then fears that surround the proliferation of deepfakes, false statements and online fraud could be assuaged.

Similarly, if we can establish without doubt that the material is AI-generated, this would facilitate the removal of false damaging material and ensure that end recipients understand what they are consuming. For the UK to become a leader in global AI governance, such measures may be what our legal system would most benefit from.

The law surrounding reputation management and AI liability has many questions yet to be resolved, but they are not beyond the wit of man. Hawking's vision of human substitution may have to wait.

AI Crisis Checklist

- **IDENTIFY THE PROBLEM** - false information? Private information? Personal data?
- **IDENTIFY THE TARGET** - who is the responsible party/parties?
- **IDENTIFY THE INFORMATION** - can you prove that the information is created by AI?
- **STEM THE FLOW** - can publication/ republication be restricted?
- **KEEP RECORDS** - do not delete any relevant material. Do you need to ask third parties to do the same?
- **CREATE A CLEAR NARRATIVE** - how can you rebut and challenge false statements?
- **MEDIA CHANNELS** - how and where to deploy your message?
- **STAKEHOLDER COMMUNICATIONS** - shareholders, employees, directors, investors?
- **CREATE YOUR TEAM** - do you need lawyers, strategic communications or law enforcement?

For more information on how we can help you contact:

JON MCLEOD
Partner
DRD Partnership
Jon.McLeod@drdpartnership.com
+44 (0) 7775 530 978

PERSEPHONE BRIDGMAN BAKER
Partner
Carter-Ruck
Persephone.BridgmanBaker@carter-ruck.com
+44 20 7353 5005

HELENA SHIPMAN
Senior Associate
Carter-Ruck
Helena.Shipman@carter-ruck.com
+44 20 7353 5005

Carter-Ruck

carter-ruck.com
The Bureau
90 Fetter Lane
London
EC4A 1EN

DRD PARTNERSHIP

drdpartnership.com
17 Slingsby Place
St Martin's Courtyard
London
WC2E 9AB