The Reporter is a Machine

An appropriate response to an AI agent inquiry? The guidelines are still being written. By **DAN ROBERTS** and **CHARLIE POTTER**

THE EMAIL LANDS, AS THEY OFTEN DO, late on a Friday afternoon.

It is professional-looking, responsible-sounding and familiar enough in shape. A set of allegations. A request for comment. A deadline. A warning that publication may follow.

Only this time the sender is not quite a reporter. It may not even be a person. It is an agent, instructed to investigate a company, sector or individual; to search public records, social posts, court filings, leaked material, dark-web references and past statements; to decide what looks interesting; to draft questions; and, perhaps, to approach the subject before any human editor has looked at the file. That may sound like science fiction. It may already be happening.

Gideon Benaim, a partner at the media law firm Simkins, recently told the *Brunswick Review* of a client that has already been approached by what appeared to be an agentic investigative journalist asking for comment. Nothing came of it, and the details remain hard to pin down. But Benaim thinks the direction of travel is obvious. "One could imagine hundreds or thousands of these agents," he says, "not only from the mainstream media but also from NGOs and citizen journalists, or bad faith actors, clogging up inboxes."

The first concern about agentic AI in crisis has naturally been cyber. Security teams are preparing for attackers that can probe systems, test weaknesses, write code, impersonate colleagues and adapt at machine speed. But if AI can accelerate the work of hackers, it can also accelerate the work of investigators.

We have already seen an adjacent version of this in cyber reporting. In one recent matter involving a prominent global institution, a ransomware group posted details of their alleged breach on the dark web. One of the first public articles to appear was not a conventional specialist report in the old sense. It appeared on a cyber news site under a byline that made no attempt to hide it was a machine and not a human, citing dark-web monitoring and packaging the claim into an article complete with analysis, "fact checker results" and a prediction of future attacks. It was published within minutes. That piece itself remained relatively unread. But it shows where the information environment is moving. A threat actor posts. A monitoring system detects. A machine-assisted publisher summarizes. A search engine indexes. A reputational problem that once might have taken days to surface can appear in public within hours.

The next reputational crisis may still begin with a phone call from a journalist. It may also begin with a machine-generated question that is good enough, fast enough and public enough to require an answer.

To be clear, serious reporting still depends on judgment, fairness, context, courage, ethics and accountability. The best investigative reporters will remain human. But the process around them is changing profoundly.

AI is already helping journalists find leads, process documents, interrogate datasets, transcribe audio, translate material, summarize filings and identify patterns that would once have required teams of analysts. It expands what journalism can hear, read and see: radio programs become searchable transcripts; document sets can be scanned for contradictions; satellite imagery and machine learning can reveal patterns of environmental harm that would be impossible to detect manually at scale. Each example is different. Together, they point to the same shift: The surface area of corporate scrutiny is expanding because the cost of looking is falling.

Outside newsrooms, campaigners, researchers, plaintiffs' lawyers, short sellers, activists, academics and specialist monitoring platforms can use similar tools to scan corporate claims, compare filings, search reports, monitor social media and test whether what a company says matches what public data shows.

Thought leadership has already become one early testing ground. Researchers have begun using AI-related tools not to write reports, but to audit them: checking citations, testing whether sources exist, comparing claims against original documents and looking for signs that public reports have relied on hallucinated or distorted references.

In several high-profile cases, companies have been forced to withdraw published reports after

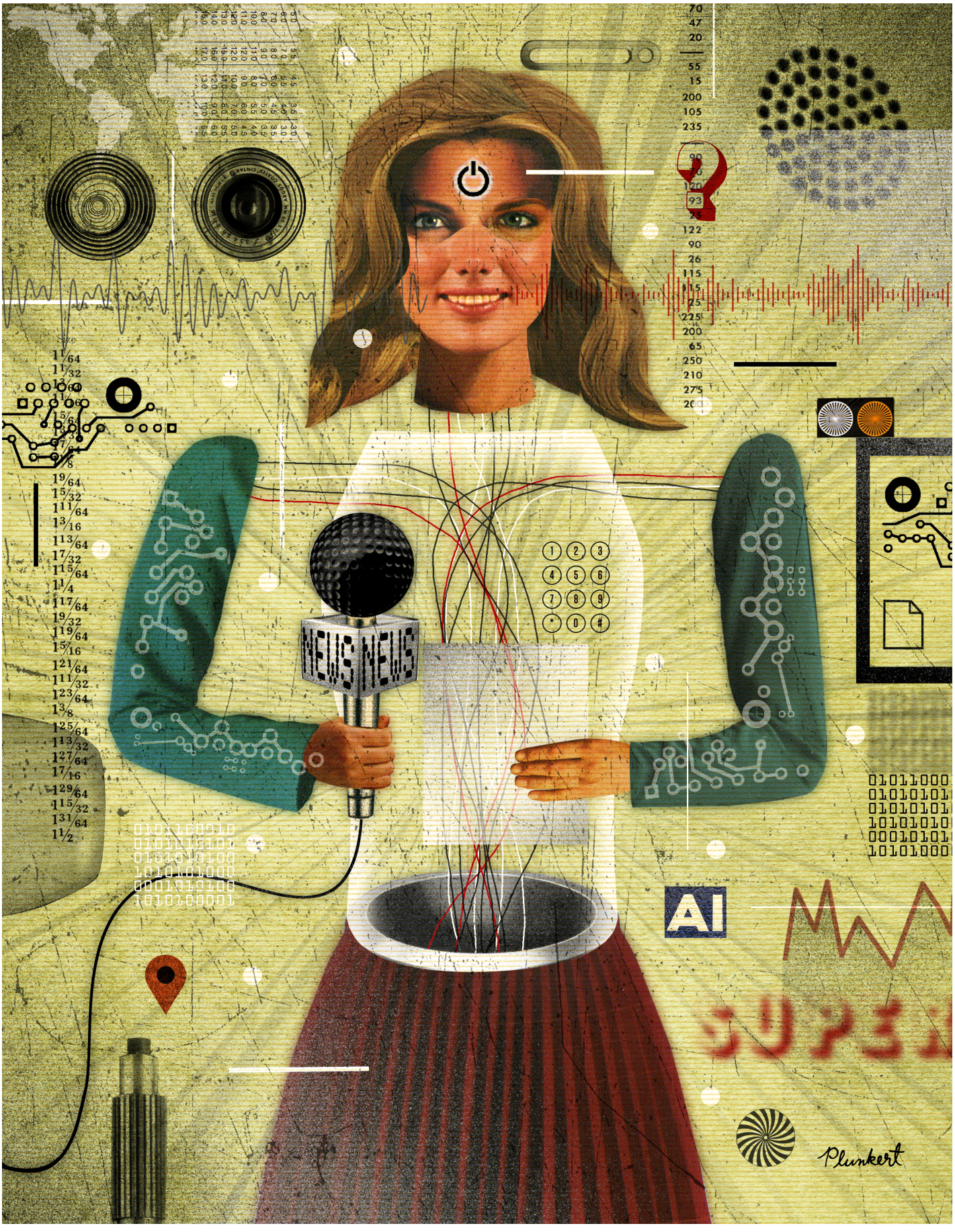


ILLUSTRATION: DAVID PLUNKERT

researchers found they contained fake citations, misattributed sources and inaccurate claims. It may prove a helpful quality control reminder. The important point is not one firm's embarrassment. It is that public thought leadership—the polished PDF, the data-heavy report, the footnoted insight piece—is becoming searchable at scale.

That should be uncomfortable for any organization that publishes reports, sustainability claims, technical assurance, AI governance papers, policy submissions or research-heavy marketing. Public material is no longer simply read by people who are interested. It may be scanned by systems designed to find contradiction, exaggeration, fabrication or weakness. The risk does not stop at investigation. AI also changes the afterlife of a story.

Once an allegation is in the open, it is easier to turn it into a spoof video, fake advert, parody song, synthetic speech clip or shareable meme. There are ever more sophisticated and creative AI-generated videos with almost Hollywood-standard production values either mocking or attacking corporate brands, products, even their executive leadership and boards. Sometimes it will be legitimate satire, sometimes it will be deliberate misinformation. Often it will sit somewhere murkier: obviously comic to some viewers, misleading or reputationally damaging to others.

A serious allegation about safety, misconduct, labor standards, environmental harm or weak governance can now be repackaged into professional-looking satire within hours. The joke can make the sting of the investigation last longer.

For crisis advisors, this creates two problems at once. The first is the allegation itself: is it true, partly true, misleading, wrong, or impossible to assess yet? The second is the form it takes once it escapes the original channel. A regulator's report may become a TikTok. A leaked document may become a comedy sketch. A data error may become a fake advert. The serious and the absurd can reinforce each other. This is why the right of reply matters more, not less, in the age of machine-assisted investigation.

In English media law, the idea that a publisher should usually seek comment before making serious allegations has long been part of the architecture of responsible journalism. The so-called Reynolds test has been replaced by the statutory public-interest defense in the Defamation Act 2013, and the old factors are not a rigid checklist. But the underlying point remains important: where serious allegations are being made, the fairness of the pre-publication process may matter.

THE SURFACE AREA OF CORPORATE SCRUTINY IS EXPANDING BECAUSE THE COST OF LOOKING IS FALLING.

For companies, the right-of-reply email is not just an irritation. It is a signal. It tells you that allegations have reached a stage where your answer—or silence—could form part of the record, and that your response may matter not only to the story but to the legal and reputational assessment of what happens next.

Agentic investigation complicates that signal. Who is the publisher if an agent sends the questions? Who is the editor? What counts as a reasonable deadline when the request is generated at machine speed? What if the agent asks 40 questions, some factual, some speculative, some derived from poor data? What if it does not disclose whose agent it is? Some of these requests will be spam. Some will be bad faith. Some will be designed to waste time or lure a company into amplifying a false allegation. But ignoring everything that looks non-human is not a strategy either. As Benaim puts it, if an agent has done real investigating and the allegation is serious enough, there may be risk in dismissing it.

The practical challenge is to preserve judgment when the machine has moved first: verify the source, separate fact from inference, understand the deadline, decide the forum and assume any response may be read by machines as well as people.

The answer is not to fear journalism, and certainly not to attack journalists for using better tools. Many of these tools will make journalism more powerful because they help reporters find things that matter. That is in the public interest. Companies that have behaved badly should expect to be found out more quickly.

For responsible companies, the challenge is to be accurate before they are accused, ready before they are asked and fair-minded when the request comes from somewhere unfamiliar. The next crisis may still begin with a phone call. But it may also begin with a machine-generated email, one that is sufficiently specific, sufficiently public and sufficiently dangerous to demand a human response.

It's important to remember not to respond at machine speed simply because the machine has moved first, but rather to do what good crisis response has always required: establish the facts, understand the stakeholder, judge the forum, preserve credibility and act with discipline.

The machine may move first. The company still has to behave like a grown-up. ◆

DAN ROBERTS is a Brunswick Partner based in Paris with 25 years' experience as a political and financial journalist. **CHARLIE POTTER,** a Partner based in London, co-leads the firm's global Litigation and Disputes practice.