

AI and the Future of Reputation Management (2025 Edition)

Key Takeaways

- **AI has redefined first impressions in reputation management:** With ChatGPT reaching 1 billion users and 30% of corporate marketing now AI-generated, LLMs have moved from novelty to necessity. They now function as information gatekeepers, synthesizing consolidated narratives about individuals and organizations that users trust implicitly and creating parallel optimization demands alongside traditional search.
- **Wikipedia remains a cornerstone of AI perception, but web browsing capabilities have amplified the importance of new sources:** Every LLM is trained on Wikipedia content, making it a critical reputation factor. And the integration of web browsing in major LLMs like ChatGPT and Claude now means real-time news and Wikipedia updates directly influence AI responses. Actively publishing and updating factual, balanced, and AI-optimized content can immediately impact how AI represents an individual or brand.
- **Direct publishing continues to increase its role in reputation building:** With Substack reaching 35 million readers by 2024 and 39% of young adults now getting news from TikTok, reputation management requires direct-to-audience strategies. Organizations must create structured, authoritative content specifically formatted for AI crawlers while maintaining consistent messaging across owned platforms.
- **AI content generation delivers substantial efficiency gains with improving quality:** Advanced LLMs have demonstrated dramatic improvements in general accuracy, industry-specific expertise, and the ability to write in nuanced styles. For reputation managers, this means content cycles can accelerate with proper human oversight, enabling faster crisis response while maintaining brand voice consistency across multiple formats and languages.
- **The risks of misinformation and deepfakes have intensified:** Organizations and individuals face unprecedented risks from both AI hallucinations (presenting false information with high confidence) and sophisticated deepfakes, which have evolved from novelties like the "Pope in a puffer jacket" to serious threats, such as a corporate fraud scheme that cost one engineering firm \$25 million through deepfake executive impersonation.

Table of Contents

1. Introduction
2. How AI Shapes Public Perception
3. LLMs as Discovery Tools
4. Comparing AI with Search Engine Traffic
5. Critical Signals for the Next Phase of AI Reputation Management
6. The Role of Wikipedia in Training LLMs
7. AI-Powered Sentiment Analysis Across Social Channels
8. Improvements in AI Content Generation
9. Deepfakes and AI-Generated Misinformation
 - Detection and Response
10. Legal and Regulatory Developments
 - Deepfake and AI-Content Laws
 - Defamation and AI Liability
 - Privacy and Content Takedown Policies
11. AI and the Transformed Media Ecosystem
12. AI-Optimized Reputation Management
 - Continuously Monitor Your AI "Reputation Profile"
 - Proactively Curate Authoritative Online Content
 - Integrate AI into Your Reputation Toolkit
 - Optimize for Generative Engine Optimization (GEO)
13. Conclusion

Introduction

In the two years since the publication of our 2023 white paper, advances in artificial intelligence have radically altered reputation management. AI has moved from novelty to necessity. It has become central to how information is created, shared, and perceived online.

ChatGPT reached 100 million users just two months after launch in 2023, the fastest-growing consumer application in history. That pace of adoption was not a fluke. In an appearance at TED 2025, OpenAI founder Sam Altman acknowledged that ChatGPT has since reached roughly 1 billion users. He said that "10% of the world" uses OpenAI's systems.¹

Large language models – the technology behind tools like ChatGPT, Google's Gemini, and Anthropic's Claude – now play a prominent role in search and social media. They have fundamentally changed how we create first impressions, just as search engines did decades ago.

¹ <https://www.forbes.com/sites/martineparis/2025/04/12/chatgpt-hits-1-billion-users-openai-ceo-says-doubled-in-weeks/>

Two years ago we felt that it was imperative to examine how AI would change our industry. That change has occurred, and it is ongoing. In this updated edition of our 2023 paper, we review recent developments in AI-related misinformation threats, analyze high-profile cases of reputational crises caused by AI, and survey AI tools and strategies for sentiment analysis and media monitoring. We also discuss emerging legal frameworks around AI accountability. Our goal is to provide practical recommendations for protecting and enhancing reputations in our newly AI-saturated world.

How AI Shapes Public Perception

Unlike traditional search engines that return a list of links, LLM-based chatbots like ChatGPT, Gemini, and Claude deliver answers in conversational form with a single synthesized response to a user's query. Sifting through pages of results is no longer necessary. Instead, AI provides a unified narrative that feels authoritative. As a result, single interactions with an AI chatbot are increasingly serving as first impressions of an individual or company. This means that there is now a need to proactively monitor and manage one's online presence to ensure that LLMs are drawing on the right information and not outputting misrepresentations. The critical importance once reserved for Google search rankings now extends to AI search results, creating parallel optimization demands.

The improved convenience and accuracy of AI has led to greater user trust in these tools, as well as an exponential increase in the volume of AI-generated content. A recent Gartner analysis estimated that 30% of outbound marketing messages from large companies are now AI-generated, up from less than 2% in 2022.² But the increase in AI content generation has not necessarily corresponded with a commitment to diligence. In a 2025 McKinsey survey, roughly a third of respondents from organizations that use generative AI said that less than 20% of their organization's AI content is reviewed by a human.³

The opportunities for utilizing AI-generated content to bolster an online reputation are plentiful, but LLMs also pose unique obstacles for reputation management. AI models can reflect the biases or inaccuracies present in their training. And an LLM's method of generating content is predictive; it creates content meant to fit a pattern represented by the content on which it was trained, meaning it is fundamentally unaware when that content is similar to the training pattern but nevertheless factually incorrect. This means any LLM is capable of "hallucinating" false information with great confidence. LLMs are black boxes. Even the designers of these systems

² <https://www.gartner.com/en/topics/generative-ai>

³ <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

can't tell you exactly how their algorithms arrive at answers.⁴ At the same time, every-day users may not realize an answer is AI-fabricated and might not double-check sources.

Consider the implications for an individual or company's reputation: a biased or erroneous summary from a chatbot could misrepresent an executive's track record or a company's products. It could aggregate negative reviews or news stories and present them as fact.

Ensuring AI outputs about your brand are accurate is now a key reputation task. This includes proactively feeding the digital ecosystem with correct data and addressing false or negative content that an AI might pick up. The explosion of AI-generated content, coupled with the risk of AI hallucinations, has amplified the need to proactively establish an online presence that is optimized to be AI-friendly and factually solid.

LLMs as Discovery Tools

Since 2023, LLMs have advanced from historically-trained autocomplete functions to sophisticated, real-time discovery tools.

What has changed is the integration of current information. When our first white paper was published, ChatGPT was limited to information available before its training cutoff date of January 2022.⁵ Now, major LLMs like ChatGPT, Claude, and Google's Gemini incorporate web browsing capabilities, allowing them to retrieve and synthesize current information. This enables these systems to function as up-to-date discovery tools that can report on recent events, including the latest news about individuals and companies.

LLMs now function as both information gatekeepers and organizers, determining which aspects of a person or company's online presence are most relevant to present to users. The critical factors that influence how an LLM portrays an entity remain similar to what we identified in 2023, but their importance has magnified:

1. **Quality and recency of content:** LLMs prioritize information from authoritative, frequently updated sources when generating responses. Organizations that maintain comprehensive, current digital footprints across multiple channels tend to be represented more accurately.
2. **Expert opinions and third-party validation:** External validation from recognized authorities in a field carries weight in how LLMs characterize individuals and

⁴ There are, however, promising and ongoing attempts to measure AI uncertainty. See <https://www.chicagobooth.edu/review/is-chatgpt-confident-about-its-answer-just-bluffing>. Tools like Claude's "Extended Thinking" and GPT's "Deep Research" also let users see inside of the process that goes into generating a response, but it is unlikely that casual users will seek out this kind of tool without an understanding of the potential for inaccurate AI predictions.

⁵ <https://community.openai.com/t/what-is-the-actual-cutoff-date-for-gpt-4/394750/4>

organizations. Positive mentions in industry publications, academic journals, or by thought leaders influence AI representations.⁶

3. **User reviews and sentiment analysis:** LLMs incorporate patterns from consumer reviews and social media sentiment, identifying recurring themes in feedback. This makes proactive management of review platforms and social channels increasingly important.⁷
4. **Consistency across platforms:** LLMs evaluate consistency of messaging and reputation across different online venues. Contradictory information or discrepancies between official statements and third-party reporting can trigger skeptical or qualified responses from AI systems.⁸
5. **Crisis handling and controversy framing:** How past controversies were handled and framed in online discourse impacts how LLMs present this information. The digital record of crisis response becomes part of the permanent reputational record that AIs draw upon.

Comparing AI with Search Engine Traffic

AI chatbot traffic has increased exponentially since the launch of ChatGPT. According to one recent two-year study, AI traffic experienced YoY growth of 80.92% from Apr 2024 to Mar 2025, totaling 55.2 billion visits.⁹ At the same time, Microsoft's Bing search engine, which incorporates the company's Bing chatbot and is powered by OpenAI's GPT models, has seen a small uptick in market share, going from about 3% to 3.9%, according to Statcounter.¹⁰ But despite the clear advancement of AI platforms as discovery tools, it's important to maintain perspective and monitor their current market position relative to traditional search engines. During that same April-March period, total AI traffic was only 1/34 the size of search engine traffic, and Google remains dominant with roughly 90% market share in search.

Competing narratives about AI's impact on search behavior emerged in May 2025, when Apple executive Eddy Cue testified that Google searches on Safari had declined for the first time in over two decades, attributing the shift to AI adoption.¹¹

But in a rare public statement, Google directly contradicted these claims. The company affirmed they "continue to see overall query growth in Search" including "an increase in total queries coming from Apple's devices and platforms."¹²

⁶ <https://www.sciencedirect.com/science/article/pii/S266729522400014X>

⁷ https://link.springer.com/chapter/10.1007/978-3-031-88531-0_9

⁸ <https://arxiv.org/abs/2501.08102>

⁹ <https://onelittleweb.com/ai-chatbots-vs-search-engines/>

¹⁰ <https://gs.statcounter.com/search-engine-market-share>

¹¹ <https://fortune.com/article/an-apple-exec-suggested-ai-chatbots-were-eroding-googles-search-business-sending-alphabet-shares-plummeting-the-truth-is-more-complicated/>

¹² <https://blog.google/products/search/statement-press-reports-about-search-traffic/>

Research conducted in March 2025 found that Google Search actually grew by over 20% in 2024, processing approximately 5 trillion searches (about 14 billion daily). The increase in searches is likely at least partially attributable to Google's own AI. The company recently claimed that its AI Overviews feature appears to increase search usage rather than cannibalize it.¹³

The apparent contradiction between Apple's testimony and Google's statement likely stems from measurement differences. While overall Google search volume continues growing, specific segments, like certain query types on Safari browsers, may indeed show declines related to increasing AI use.

We're still in the early stages of AI's impact on information discovery, but that impact is clearly expanding. There is an increasing need to optimize for both traditional search and emerging AI discovery tools, as information retrieval is fragmenting across multiple channels rather than simply shifting from one to another. The dual optimization challenge we identified in 2023 has only intensified, requiring reputation managers to develop sophisticated strategies that account for both search algorithms and LLM training patterns.

10 Critical Signals for the Next Phase of AI Reputation Management

1. The generative AI market was valued at \$67 billion in 2024 and is expected to grow annually by 24.4% through 2030, indicating significant industry expansion that reputation management strategies need to address.
2. ChatGPT's referral traffic to websites has grown dramatically—from sending traffic to less than 10,000 unique domains daily in early July to over 30,000 by November—showing its increasing influence in directing users across the internet.
3. About 54% of ChatGPT queries are handled without SearchGPT, while 46% utilize its web search feature—highlighting a split in how users interact with AI tools when seeking information.
4. 70% of ChatGPT prompts represent entirely new types of search intent, deviating from traditional search categories like navigational or transactional—indicating users are solving problems in novel ways through AI.
5. With SearchGPT enabled, user intent distribution more closely mirrors Google's, including increases in navigational, commercial, and transactional queries—vital insights for reputation managers optimizing for AI platforms.

¹³ <https://sparktoro.com/blog/new-research-google-search-grew-20-in-2024-receives-373x-more-searches-than-chatgpt/>

6. Academic publishers, educational platforms, and technical resources now receive significantly more traffic from ChatGPT than Bing, highlighting sectors where AI is becoming a dominant discovery tool.
7. ChatGPT's user base skews younger and more male than Google's, with stronger representation among students—valuable demographic data for refining messaging strategies in AI-influenced ecosystems.
8. Research suggests that LLMs are beginning to develop proprietary ranking and weighting systems distinct from traditional search algorithms, indicating that AI reputation optimization strategies will need to evolve beyond current SEO practices.
9. Analysis of ChatGPT responses shows that structured list-format content and comprehensive resource articles demonstrate higher retrieval rates for brand mentions, suggesting specific content formats that can improve visibility in AI-generated responses.
10. The concept of "Retrievability" is emerging as a critical factor, expanding reputation strategy beyond crawlability and rankability to include how easily core information about a brand is accessed and prioritized by LLMs.

The Role of Wikipedia in Training LLMs

Wikipedia continues to serve as a foundational training source for LLMs. In our 2023 paper, we noted a *Washington Post* and Allen Institute analysis of Google's C4 dataset that found Wikipedia was the second most prominent source of LMM training data after Google Patents, with patents likely used for more technical content not subject to as many casual user queries.¹⁴

This means that maintaining an accurate, balanced Wikipedia page is crucial to AI search results. Several developments since our initial paper have affected Wikipedia's role in AI reputation management:

1. **Expanded training corpora:** According to the Wikimedia Foundation, "every LLM is trained on Wikipedia content, and it is almost always the largest source of training data in their data sets."¹⁵ However, modern training now increasingly incorporates academic journals, selected news archives, and curated web content alongside Wikipedia.¹⁶ Wikipedia remains crucial but is not the sole authoritative source shaping AI perceptions of people and organizations.
2. **Citation validation:** LLMs have improved capabilities to cross-reference Wikipedia claims against other sources. However, a recent study found that, even for the most advanced

¹⁴ <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>

¹⁵ <https://wikimediafoundation.org/news/2023/07/12/wikipedias-value-in-the-age-of-generative-ai/>

¹⁶ <https://arxiv.org/abs/2401.02038>

models like GPT-4 with web search capabilities, "approximately 30% of individual statements are unsupported, and nearly half of its responses are not fully supported" by cited sources.¹⁷ While the focus of the study was medical citations, the same issue applies more broadly, and it has spurred development of automated verification platforms specifically designed to evaluate citation accuracy.¹⁸

- 3. Real-time updates:** With many LLMs now incorporating web browsing capabilities, they can retrieve the most current version of a Wikipedia page rather than relying solely on training-time snapshots, making ongoing Wikipedia monitoring even more essential.¹⁹

The potential reputation impact of Wikipedia inaccuracies has been magnified by AI proliferation. When an LLM encounters a query about a person or organization, it frequently draws key facts and framing from Wikipedia entries, often preserving the same emphasis and narrative structure. This means that problematic elements of a Wikipedia biography—excessive focus on controversies, misleading contextual framing, or outdated information—frequently reappear in AI-generated summaries.

This reinforces what has always been a key practice for effective online reputation management: Wikipedia monitoring and ethical engagement with the platform's editorial processes. Wikipedia remains the most visible and influential encyclopedia in the world, and its structured format makes it particularly valuable for AI training. While LLMs increasingly incorporate other sources, Wikipedia's significance as a reputation management priority has only increased with the rise of AI as an information intermediary.

AI-Powered Sentiment Analysis Across Social Channels

Sentiment analysis capabilities have advanced dramatically since our 2023 paper, with AI-powered systems now able to detect nuances, emotions, and cultural contexts that were previously challenging for algorithms.²⁰ These improvements have transformed how organizations monitor and manage public opinion.

Modern sentiment analysis tools do much more than count positive vs. negative mentions. They use AI to understand context and tone. Sarcasm, humor, and slang—which often confounded older algorithms—are now parsed more accurately by LLM-based systems. These systems can reference vast training data and implement natural language processing (NLP) to discern that a tweet like "Great, another product launch... 😬" is negative, despite the surface-level positive word. They also handle multilingual content with greater ease.²¹

¹⁷ <https://arxiv.org/abs/2402.02008>

¹⁸ <https://www.nature.com/articles/s41467-025-58551-6>

¹⁹ <https://www.sciencedirect.com/science/article/abs/pii/S0167923624001878>

²⁰ <https://link.springer.com/article/10.1007/s40547-024-00143-4>

²¹ <https://journals.sagepub.com/doi/abs/10.3233/IDT-240773>

Sentiment analysis platforms aggregate data from millions of sources and apply AI to filter noise and identify the conversations that matter. In practical terms, this means a brand can receive an automated alert like: "Mentions of [Your Company] are up 300% in the past 2 hours, with predominantly negative sentiment, driven by a Reddit thread in /r/technology." Immediately, you can click in to see the posts, which the AI might summarize for you. This is transformational for crisis response. Instead of finding out when a crisis becomes viral news, you can potentially catch the fire while it's still small.

Social listening AI can cluster similar mentions and boil down "key themes." For example, an AI might review 10,000 tweets about a company and report: "Top topics: customer service wait times (40% of negative sentiment), pricing changes (25% of negative), product quality (15% positive mentions about new feature)." These insights help pinpoint exactly what issue is hurting reputation, or what positive messages to amplify. The use of LLMs here allows more nuanced categorization than simple keyword spotting, because the AI "reads" the posts almost like a human analyst would.

Advanced tools can compare sentiment across different channels or demographics. You might discover that sentiment is +20 net positive on LinkedIn (perhaps your thought leadership is well-received), but -10 net negative on Twitter (maybe a separate audience is upset about something). Having this granularity guides you to deploy targeted messaging where needed. It also helps measure the impact of PR initiatives on specific audiences.

Beyond text, AI tools are now analyzing images and videos for reputation signals. Visual listening can identify, for example, your company's logo appearing in a viral meme or a YouTube video's comments expressing outrage at your brand.

It's worth noting that while these tools are powerful, human oversight remains crucial. AI analytics might occasionally mislabel sarcasm or miss cultural context. Also, bias in training data can influence automated sentiment scoring. Savvy PR teams use AI tools to augment, not replace, their judgment, treating the AI's findings as highly informative signals that still need human interpretation.

Improvements in AI Content Generation

When we published our original paper, it was novel to have AI draft marketing copy, blogs, or social posts. This has since become standard practice for many organizations. The shift to AI-based content generation requires careful understanding of where and how AI fits into content strategy and reputation management.

The core capabilities and limitations of AI-generated content that we identified in 2023 remain, but LLMs have since demonstrated markedly improved coherence with fewer instances of hallucination. LLMs' predictive method of content generation means that they will never be

immune to the risk of hallucination.²² However, as noted above, significant effort has gone into identifying, predicting, and eliminating hallucinations, and services such as ChatGPT and Claude are far more accurate and thorough than they were in 2023.

According to a 2025 report from The Stanford Institute for Human-Centered AI, LLM systems saw dramatic gains on newly introduced evaluation benchmarks last year. Scores increased by 67.3 percentage points on SWE-bench, 48.9 on GPQA, and 18.8 on MMMU—tests designed to assess reasoning, scientific knowledge, and software engineering. These rapid improvements reflect broader progress: language model agents have begun outperforming humans in certain time-constrained programming tasks, and AI-generated images, video, and even music have reached new levels of quality and coherence.²³

The Stanford report also found that the AI frontier is becoming more competitive and densely populated, with performance gaps tightening across top models. Over the past year, the Elo skill score—an adapted rating system that ranks models based on head-to-head comparisons, similar to chess—showed the gap between the top and 10th-ranked models shrink from 11.9% to 5.4%, with just 0.7% separating the top two. At the same time, model scale is accelerating: training compute doubles every five months, dataset sizes double every eight months, and power usage doubles roughly every year. While academia still produces the most highly cited research in AI, nearly 90% of leading AI models in 2024 originated from industry, a jump from 60% the year prior.

The most popular of these models, GPT-4o, achieved an accuracy rate of 86.5%²⁴ on the Massive Multitask Language Understanding (MMLU) benchmark, which assesses a model's proficiency across 57 diverse subjects such as mathematics, history, and law.²⁵ Similarly, Claude 3.5 Sonnet demonstrated significant improvements over its predecessor, Claude 2.1, with a twofold increase in accuracy on challenging, open-ended questions.²⁶

Technical advancements have extended to domain expertise as well. Fine-tuned models now handle specialized knowledge areas with greater precision, drawing on industry-specific training that allows for more technical content creation in fields ranging from healthcare to financial services. An April 2025 study conducted by MIT's Media Lab, the Brazilian university UFABC, and the pandemic prevention nonprofit SecureBio revealed that advanced AI models like GPT-4o and Gemini 2.5 Pro outperformed PhD-level virologists in practical lab troubleshooting tasks.²⁷

LLMs have also improved their ability to generate a variety of nuanced writing styles, a skill that is crucial for content generation related to a personal or corporate brand. LLMs' improved ability to maintain consistent brand voice across diverse content formats enables organizations to

²² <https://arxiv.org/abs/2401.11817>

²³ <https://hai.stanford.edu/ai-index/2025-ai-index-report>

²⁴ <https://openai.com/index/hello-gpt-4o/>

²⁵ <https://arxiv.org/abs/2009.03300>

²⁶ <https://www.anthropic.com/news/claude-3-family>

²⁷ <https://time.com/7279010/ai-virus-lab-biohazard-study>, <https://www.virologytest.ai/>

produce cohesive messaging at scale without the stylistic variations that previously betrayed AI authorship.

In addition, today's models support multilingual content creation with near-native fluency, detecting subtle idiomatic expressions and cultural references that earlier iterations missed. Content production workflows have also improved through better integration with media assets, where AI can now suggest appropriate images and videos that align with textual themes.²⁸

Pairing discipline-specific specialization with expanded language capabilities gives a skilled writer utilizing AI endless possibilities for generating informative, careful content at scale with a level of efficiency that was previously unimaginable.

While these are no doubt exciting advancements, and expectations for future improvements are justifiably high, persistent limitations warrant caution. Factual errors remain problematic, especially for time-sensitive information where models may confidently present outdated data as current. Subtle brand misalignments in tone or messaging can also occur when AI-generated content fails to capture the distinctive voice that organizations have carefully cultivated over time. And highly technical or regulated content in fields like law, medicine, and finance still benefit from expert review.

Human oversight remains important. Forward-thinking reputation management strategies utilize the exponential boost in scale that AI-generated content provides, while also maintaining the option of human-in-the-loop processes for content approval, particularly for high-stakes communications. AI can allow communications teams to scale content production while freeing human talent for higher-value tasks that leverage emotional intelligence, strategic thinking, and relationship building. Rather than diminishing the role of communications professionals, AI enables them to focus on the aspects of reputation management where human judgment adds the most value.

Deepfakes and AI-Generated Misinformation

Deepfakes—AI-generated synthetic videos or audio that mimic real people—have become more realistic and easier to produce since 2023. Likewise, the scale at which AI generates misinformation (fake news articles, bogus social media posts, impersonation emails) has increased.

AI misinformation is no longer a fringe concern; it's mainstream. One of the first examples occurred shortly before the publication of our first AI white paper, in March 2023: the "Pope in a puffer jacket" incident centered on an AI-crafted image of Pope Francis in a large white coat that went viral and fooled millions for a short time.²⁹

²⁸ <https://hai.stanford.edu/ai-index/2025-ai-index-report>

²⁹ <https://www.thomsonreuters.com/en-us/posts/government/deepfakes-federal-state-regulation/>

While that particular deepfake was light-hearted, it demonstrated how convincingly AI can distort reality. By early 2024, explicit deepfakes were causing real uproar. For instance, fake nude images of Taylor Swift were circulated on social media without consent.³⁰ Concerns surrounding the ease with which the public can create deepfake nudes have increased along with AI's technological improvements.

Scammers also used AI-generated deepfake videos of public figures such as Elon Musk, Tucker Carlson, and Mark Cuban to promote a fraudulent stock platform called "Quantum AI."³¹ Meanwhile, deepfake audio and video has been used to impersonate company executives in communications with employees or partners, a form of phishing that can lead to financial or data breaches and a loss of confidence in the targeted company's security. In February 2024, a finance employee at Arup, a British engineering firm known for projects like the Sydney Opera House, was defrauded of \$25 million through an elaborate deepfake scam in Hong Kong. The worker received an email supposedly from the firm's CFO requesting a "secret transaction," and though initially suspicious, was convinced after joining a video conference where all participants were in fact AI-generated deepfakes.³²

Such incidents illustrate the dual threat of deepfakes: on one hand, public figures or corporations may be blamed for things they never said or did, and on the other, audiences may start distrusting legitimate communications, suspecting they could be deepfakes.

Deepfake technology has advanced to the point that, with just a few minutes of audio or a handful of photos, realistic fake videos and audio can be produced with off-the-shelf AI apps. This democratization of deepfakes means any high-profile individual or brand could be targeted.

In addition to creating misinformation, AI can help spread falsehoods. In 2024, multinational dairy producer Arla became the target of a viral online misinformation campaign bolstered by AI.³³ The company had announced a pilot of Bovaer, a new feed additive for cows that can reduce methane emissions. In response, conspiracy theories exploded across social media claiming that the additive was unsafe for humans and was part of a nefarious "depopulation" plot linked to billionaire Bill Gates. These claims had no scientific basis – Bovaer had been approved by regulators and tested for years – yet the narrative caught fire among anti-GMO and anti-vaccine social media circles. The false story spread so widely and rapidly that Arla and its feed supplier DSM found themselves "scrambling to defend their products and reputations" against a tide of misinformation. In this case, AI likely played a role in accelerating the spread: algorithms amplified the most outrageous posts, and troll networks (potentially with automated

³⁰ <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box>

³¹ <https://www.rmit.edu.au/news/factlab-meta/elon-musk-used-in-fake-ai-videos-to-promote-financial-scam>

³² <https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>

³³ <https://agfundernews.com/the-arla-bovaer-case-highlights-a-new-frontier-for-agrifoodtech-combatting-misinformation>

bots) helped echo the conspiracy across platforms. The incident underscores how even a positive sustainability initiative can be spun into a reputational attack, especially when it touches on emotionally charged topics that can now be magnified via AI-curated echo chambers.

Deepfakes and AI falsehoods can inflict various harms: defamation, character assassination, fraud, or sowing general mistrust. A deepfake video of a CEO appearing to make a racist remark, for example, could ignite a global scandal in hours. Even if debunked later, the initial shock can permanently imprint on public consciousness. Likewise, AI-generated fake press releases or statements can move markets or damage stakeholder confidence.

There is now a "guilty until proven innocent" risk if a fake video, audio clip, social post, or news story surfaces. Immediate public reactions on social media can be harsh, and corrections often lag behind the viral falsehood.

Detection and Response

The tech community is not sitting idle. The same AI that creates deepfakes is also being used to detect them. Companies and academic labs have developed deepfake detection algorithms trained to spot inconsistencies or digital watermarks. Convolutional Neural Networks (CNNs) are commonly used to detect visual inconsistencies in images and videos, such as unnatural facial movements or irregular lighting. Recurrent Neural Networks (RNNs) and Transformers are employed to analyze temporal sequences in videos, identifying anomalies over time. These models are trained on large datasets containing both real and fake media to learn distinguishing features.³⁴

However, this has become an arms race. As detection improves, generation improves in tandem. This dynamic closely resembles what's seen in Generative Adversarial Networks (GANs), which were actually designed to improve generative AI systems based on this very principle. In GANs, two neural networks, a generator and a discriminator, are pitted against each other. The generator creates fake data, while the discriminator tries to distinguish real from fake. Through this competition, both networks improve: the generator gets better at creating convincing fakes, and the discriminator becomes more discerning. But in the deepfake context, this creates diminishing returns for detection as fakes become nearly indistinguishable from real media.

Detection tools have improved and can catch many deepfakes, but not instantly and not with 100% accuracy. Major social platforms have some automated detection (YouTube and Facebook scan for known deepfake hashes), but novel fakes can slip through. And casual consumers of online content are not vetting what they consume through these tools. This means a fake can spread widely before being identified.

³⁴ <https://opencv.org/blog/deepfake-detection-with-computer-vision/>

The Content Authenticity Initiative (backed by tech and media companies) is working on ways to cryptographically sign legitimate images and videos at the source, so viewers can verify authenticity.³⁵ Some news organizations now publish with such digital signatures. Meanwhile, companies like OpenAI and Google are researching watermarking AI-generated text and images. OpenAI released a classifier for AI-written text, but it had limited success. This area is evolving, and we anticipate more robust solutions in the next few years.

Reputation managers must treat visual/media content with skepticism and be ready to respond. Swift rebuttal is key: when a deepfake or false story emerges, having pre-established channels (official social media accounts, press contacts) to assert "This is fake, and here's how we know" can stop the spread.

Another strategy is pre-emptive inoculation: educating your audience and employees that "if you ever see or hear X outrageous thing supposedly from us, be aware deepfakes exist. Check our official statements first." While you can't predict every fake, simply raising awareness that "AI can fake videos and we'll never ask you to, say, transfer funds over a phone call" can prevent some harm.

Legal and Regulatory Developments

In 2024, U.S. federal agencies introduced 59 AI-related regulations—over twice the number from 2023, and spanning twice as many agencies. Internationally, AI references in legislation increased by 21.3% across 75 countries, a ninefold jump since 2016. Governments are backing this momentum with large-scale funding. Canada committed \$2.4 billion, China launched a \$47.5 billion chip fund, France pledged €109 billion, India allocated \$1.25 billion, and Saudi Arabia announced a \$100 billion initiative.³⁶

Lawmakers and regulators worldwide have awakened to the implications of AI on reputation and privacy. Between 2023 and 2025, we've seen a flurry of legislative proposals, new laws, and policy changes focused on AI accountability, deepfakes, data privacy, and content takedowns. For organizations concerned with reputation, it's critical to understand this new legal environment, as it creates both risks (new regulations to comply with) and remedies (new tools to combat malicious content).

Deepfake and AI-Content Laws

A number of jurisdictions have moved to curb malicious deepfakes. In the U.S., while there isn't yet a comprehensive federal law, Congress has been considering several bills. The DEEPFAKES Accountability Act seeks to provide legal recourse to victims of harmful deepfakes and potentially mandate certain disclosures.³⁷ Another proposal, the Protecting Consumers from

³⁵ <https://contentauthenticity.org/>

³⁶ <https://hai.stanford.edu/ai-index/2025-ai-index-report>

³⁷ <https://www.congress.gov/bill/118th-congress/house-bill/5586/text>

Deceptive AI Act, would require that AI-generated content is clearly labeled as such in many contexts.³⁸ By the end of 2024, at least 16 U.S. states had enacted laws addressing deepfakes, particularly in two areas—non-consensual explicit content and election interference.

For example, California and Texas make it a crime to create or distribute deepfake porn without consent, and they allow victims to sue. Texas also passed an election law that prohibits distributing deceptive deepfake videos of candidates within 30 days of an election, enforceable with civil penalties.³⁹ In Illinois, a proposed "Digital Forgeries Act" would give individuals depicted in any digital forgery the right to sue for damages if it caused harm.⁴⁰ These legal tools mean that if someone creates a damaging deepfake of a CEO or celebrity, the victim has stronger grounds to demand its removal and seek compensation. For companies, it's important to know the jurisdictions of these laws—e.g., if deepfake content is hosted on servers in a state with strict laws, that can aid takedown efforts.

The EU's Artificial Intelligence Act, which came into force on August 1, 2024, establishes a comprehensive legal framework for AI within the European Union. The Act classifies AI systems into four risk categories: unacceptable, high, limited, and minimal. Deepfakes, especially those used in contexts that can significantly impact individuals' rights or society (e.g., political manipulation, defamation), may be classified as high-risk and thus subject to stricter regulatory requirements. For high-risk AI systems, the Act mandates strict measures for data security and user privacy, ensuring that AI systems are designed and deployed with these considerations at the forefront. This includes rigorous requirements for how data is handled and protected, ensuring that users' personal information remains secure.⁴¹

China also took an aggressive early stance: as of January 2023, Chinese law mandates that any AI-generated media that could cause public misrecognition must be clearly marked with a watermark or label.⁴² By late 2024, China proposed even stricter rules to watermark all AI-generated images, videos, and audio, with requirements applying to platforms and creators alike. Singapore and India have also been studying similar measures.

Defamation and AI Liability

The past two years have seen groundbreaking legal cases regarding who bears responsibility when AI systems defame individuals. In *Starbuck v. Meta Platforms* (2025), political commentator Robby Starbuck sued Meta over claims that its AI chatbot (powered by Meta's Llama model) falsely accused him of criminal acts. According to the complaint, Meta's chatbot

³⁸ <https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation>

³⁹ <https://natlawreview.com/article/texas-did-it-first-texas-was-first-enact-state-legislation-use-deep-fakes-elections>

⁴⁰ <https://www.ilga.gov/legislation/billstatus.asp?DocNum=2123&DocTypeID=HB&GA=103&GAID=17&LegID=145586&SessionID=112>

⁴¹ https://en.wikipedia.org/wiki/Artificial_Intelligence_Act

⁴² <https://www.sixthtone.com/news/1015923/china-proposes-mandatory-labeling-of-ai-generated-content>

told users that Starbuck had participated in the January 6 U.S. Capitol riot—a fabricated allegation. Starbuck says he notified Meta in August 2024, but the bot continued to publish the lies for months, causing him serious reputational harm and even death threats. In April 2025, after these errors persisted in a new AI voice feature, Starbuck filed a defamation lawsuit seeking over \$5 million in damages. Meta acknowledged the issue as "unacceptable" and apologized, but the case raises fundamental issues of platform responsibility when an AI model acts as the publisher of false information.⁴³

Similarly, in January 2024, a Georgia judge denied OpenAI's motion to dismiss Mark Walters' defamation lawsuit (filed in 2023) after ChatGPT falsely claimed he had embezzled funds from a non-profit. This ruling signaled that courts are willing to scrutinize AI "hallucinations" under defamation law.⁴⁴

A central legal question is whether Section 230 of the Communications Decency Act (1996), which immunizes platforms from liability for third-party content, applies to AI-generated content. Traditionally, Section 230 shields internet companies when users post defamatory material, on the logic that the company is not the "speaker." But in the scenario of a large language model like ChatGPT or Meta's chatbot, the AI itself produces the language. This is uncharted territory. If an AI output is considered content developed by the platform's own tool, Section 230 might offer no shelter. So far, no statute or binding precedent explicitly extends Section 230 to AI-generated speech, and the courts in the Starbuck and Walters cases will likely grapple with this issue.⁴⁵

Privacy and Content Takedown Policies

Major tech platforms have updated their policies to address AI content. Google announced in mid-2023 and implemented by late 2024 a significant update: the search giant now makes it easier for people to request removal of AI-generated explicit or violent images of themselves, even if the image is fake.⁴⁶ In July 2024 Google changed its algorithm and policies to demote deepfake porn websites in search results and allow direct removal requests for non-consensual AI images.⁴⁷ This came after reports showed Google Search had been inadvertently driving traffic to such harmful content. Early evidence suggests this update is working—searches for celebrities known to be targeted by deepfakes now surface far fewer of those fake videos. For reputation defenders, this is a big win: it means if a client is victimized by AI deepfakes, we have a clearer precedent for getting that content scrubbed from search results, which is often the main way people would encounter it.

⁴³ <https://abcnews.go.com/Technology/wireStory/conservative-activist-roby-starbuck-sues-meta-ai-responses-121334771>

⁴⁴ <https://knowingmachines.org/knowing-legal-machines/legal-explainer/cases/walters-v-openai>

⁴⁵ <https://www.theverge.com/2023/6/9/23755057/openai-chatgpt-false-information-defamation-lawsuit>

⁴⁶ <https://www.theguardian.com/technology/2023/aug/03/google-to-launch-privacy-tools-which-remove-unwanted-personal-images>

⁴⁷ <https://techpolicy.press/big-tech-may-finally-be-taking-explicit-deepfakes-seriously>

Social media companies, under pressure, have also refined their stance. After an Oversight Board ruling in July 2024, Meta committed to permanently remove certain AI-manipulated images that sexualize public figures without consent. The Oversight Board made a strong statement that "Given the severity of harms, removing the content is the only effective way to protect people... Labeling manipulated content is not appropriate in this instance." In other words, some content is so egregious that just flagging it as "manipulated" isn't enough—it needs to come down entirely.⁴⁸ This is a shift from earlier, more lenient policies. Since 2020 Twitter (now X) has enforced a policy requiring people to label synthetic media and prohibiting harmful deepfakes (though enforcement under new ownership has been inconsistent).⁴⁹ TikTok and YouTube likewise prohibit deceptive deepfakes that cause harm, and they rely on a mix of AI detection and user reporting to catch these.⁵⁰

For those managing reputations, these legal developments offer both shields and guardrails. Individuals and companies have more avenues to fight harmful content: from invoking deepfake laws, to leveraging platform policies for quick removal, to citing privacy rights to remove personal data. On the other hand, companies deploying AI in their own operations must exercise caution: if you use an AI chatbot in customer service, be mindful of what it says; you could be on the hook if it gives libelous or biased responses.

AI and the Transformed Media Ecosystem

The new media environment we described in our 2023 paper—with traditional outlets struggling and direct publishing platforms gaining influence—has only been amplified in the past two years. The following five trends in AI have dramatically altered media in recent years.

- 1. LLMs as information gatekeepers:** Large language models have emerged as a new layer in the information ecosystem, functioning as both aggregators and repackagers of media content. When users query an AI about news or current events, they receive a synthesized version that may draw from multiple sources but presents a unified narrative. This "middleware" role means AI systems now influence which media sources receive attention and how their content is contextualized.
- 2. Direct publishing takes over:** The decline in traditional media that we noted in 2023 has continued, with further layoffs and consolidation in the news industry.⁵¹ Meanwhile, direct publishing platforms have expanded their reach. Substack reached 35 million active readers by 2024, Twitter (X) has maintained its position despite controversies, and platforms like podcasts, YouTube, and TikTok have become significant sources of news

⁴⁸ <https://www.oversightboard.com/news/new-decision-addresses-metas-rules-on-non-consensual-deep-fake-intimate-images/>

⁴⁹ <https://www.theverge.com/2020/2/4/21122661/twitter-deepfake-manipulated-media-policy-rollout-date>

⁵⁰ <https://apnews.com/article/tiktok-ai-label-creators-1ba5ee717832e41178e6a35cb0442e07>

⁵¹ <https://www.theguardian.com/media/2024/nov/18/associated-press-layoffs>

for younger demographics; in 2024, 39% of U.S. adults aged 18-29 reported regularly getting news from TikTok, up from 3% in 2020.⁵²

3. **Synthetic content at scale:** AI-generated media content has proliferated beyond what many anticipated. News outlets increasingly use AI for routine reporting (sports scores, financial results, weather), while synthetic voices read audio versions of articles. Some digital publications now openly attribute portions of their content to AI systems, while others use AI behind the scenes for research and drafting without explicit disclosure.
4. **Trust remains low but is redistributing:** Public trust in media has remained historically low. Gallup's 2024 polling showed only 31% of Americans expressing "a great deal" or "quite a lot" of confidence in newspapers, though this varied significantly by political affiliation.⁵³ Trust in AI-synthesized information and AI search is mixed, but is improving. One study found that participants attributed similar levels of credibility to content generated by AI and humans, with AI-generated content often rated as clearer and more engaging.⁵⁴ Additionally, a large-scale experiment indicated that while participants generally trust generative AI search less than traditional search, the inclusion of reference links and citations—even if incorrect—significantly increased trust in AI-generated content.⁵⁵
5. **LLM "grooming" and information manipulation:** A new phenomenon emerged in 2024-2025: coordinated efforts to influence AI training data to shape future responses. Researchers documented how state actors and special interest groups flooded the web with content designed to be ingested by AI crawlers, effectively manipulating how LLMs would later present certain topics. According to a March 2025 report, a Moscow-based disinformation network called "Pravda" flooded the web with 3.6 million AI-targeted articles in 2024—content pushing pro-Russian stories, especially about the war in Ukraine. The goal was not just to sway human readers, but to influence AI chatbot results.⁵⁶

These developments have profound implications for reputation management. Organizations and individuals must now consider not only how they appear in traditional and social media, but also how AI systems interpret and present that coverage. Several strategies have proven effective:

Direct-to-AI optimization: Forward-thinking organizations are developing content specifically designed to help AI systems accurately represent their positions. This includes creating

⁵² <https://www.pewresearch.org/short-reads/2024/09/17/more-americans-regularly-get-news-on-tiktok-especially-young-adults/>

⁵³ <https://news.gallup.com/poll/651977/americans-trust-media-remains-trend-low.aspx>

⁵⁴ <https://arxiv.org/abs/2309.02524>

⁵⁵ <https://arxiv.org/abs/2504.06435>

⁵⁶ <https://www.axios.com/2025/03/06/exclusive-russian-disinfo-floods-ai-chatbots-study-finds>

structured FAQ pages, maintaining up-to-date fact sheets, and publishing comprehensive position statements on key issues, all formatted in ways that are easily parsed by AI crawlers.

Omnichannel consistency: With AI drawing from diverse sources, maintaining consistent messaging across owned, earned, and social media has become even more critical. Contradictions between channels are more likely to be surfaced by AI systems, potentially undermining trust.

Authoritative content partnerships: Collaborating with trusted media outlets and established platforms on in-depth, factual content about you or your organization provides LLMs with high-quality sources to draw from when answering user queries.

Monitoring AI narratives: Regular auditing of how major AI systems represent your organization has become as essential as traditional media monitoring. This involves systematically querying systems like ChatGPT, Gemini, and Claude about your company, products, controversies, and executives to identify potential misrepresentations.

Correction protocols: When AI systems do misrepresent you or your organization, having established correction pathways is crucial. Many AI developers now have formal processes for organizations to submit factual corrections, though success varies across platforms.

The media ecosystem's transformation means reputation management must now encompass traditional journalism, social media, direct publishing, and AI-mediated information, each with distinct characteristics but all interconnected in shaping public perception.

AI-Optimized Reputation Management

It's no longer enough to do periodic Google audits and have a crisis PR plan on paper. Brands and high-profile individuals need to bake AI considerations into every aspect of their reputation strategy. This section provides future-forward recommendations to help you stay ahead of the curve and resilient against AI-related threats, while capitalizing on AI's benefits.

1. Continuously Monitor Your AI "Reputation Profile"

Just as you might track news and social mentions, you should keep tabs on what AI platforms "know" or say about you. This means regularly querying popular LLMs (ChatGPT, Gemini, Claude) with relevant prompts: "Who is [Your Name]?" or "What is [Your Company] known for?" and seeing what comes up. If the answers reveal inaccuracies or outdated info, that's a flag. Those tools derive their answers from the data available to them (training data or web results), so any error suggests something is amiss in the source information. You can also ask tools like Claude and ChatGPT to provide sources for any claim they make, which gives you an inside look into the kind of sources the model privileges and draws on. By monitoring these responses, you effectively audit your AI Search Engine Results Page (AI-SERP). If you find an error, address it at

the root (e.g., correct your Wikipedia page, publish clarifying content, etc.) because you can't directly "edit" the AI's answer. The goal is to ensure the first impression given by AI is accurate and positive, just as you do with Google SEO for traditional search.

2. Proactively Curate Authoritative Online Content

AI models and algorithms tend to prioritize information that is well-sourced, consistent, and comes from authoritative sites. To optimize for this, review the key sources that AI and search engines draw from when people seek info about you:

- **Wikipedia:** As noted earlier, Wikipedia is a top source for LLM training. Invest effort in maintaining a balanced, up-to-date Wikipedia page (if you meet notability standards). Ensure any negative incidents are contextualized and sourced correctly, and add noteworthy positive achievements. Given Wikipedia's guidelines, this often means working through impartial editors or providing facts on the Talk page. The payoff is huge: a clean Wikipedia article can influence AI outputs and signal credibility.
- **Search Engine Results:** Maintain a strong presence in Google's top results for your name or brand. This includes your official website (with an easy-to-scan "About" page), press releases, major media profiles, and professional platforms like LinkedIn. These high-ranking pages act as the "source of truth" that many AI systems will defer to. If you have old or negative results that are prominent, consider an ORM campaign to push them down—not just for human eyes, but because AI might ingest that negativity if it's highly visible.
- **Social Media and Direct Channels:** Increase your direct-to-audience communications. Regularly post on platforms where you have followers (Twitter/X, LinkedIn, a personal blog). LLMs might be trained on some of this content, but even if not, having a track record of authentic communication helps in a different way: if a controversy arises, you have an established channel to speak from, and people will look at your past posts (which hopefully convey transparency and goodwill) to gauge your character. Essentially, you're building a bank of goodwill and context that both humans and AIs can draw from.
- **Fact-Based Content and Thought Leadership:** Publish content that demonstrates expertise and positive contributions. This could be op-eds, white papers (like this one), conference talks on YouTube, etc. Not only do these bolster your reputation generally, but they often get picked up in the data that informs AI. For instance, an LLM might recall "[Your Name] has advocated for industry best practices in [Topic]" if you have enough content out there on that topic. It's a way of "training" the AI on your strengths.

3. Integrate AI into Your Reputation Toolkit (But Wisely)

To fight fire with fire, utilize AI tools to manage your reputation more effectively:

- **Social Listening AI:** As discussed, deploy advanced sentiment analysis to catch issues early. Set specific triggers for your known pressure points (CEO name + "scandal" or product name + "controversy", etc.). Use these tools to also identify positive testimonials or emerging advocates for your brand. Those can be amplified to strengthen your reputation.
- **AI Content Creation with Human Oversight:** Leverage AI for productivity in content creation. This could include drafting responses to reviews, generating first drafts of blog posts or press releases to announce positive news, translating content to reach a wider audience, etc. Using ChatGPT, Claude or similar LLMs can speed up content cycles, which is crucial in fast-moving situations. However, always have human review and fact-checking. An AI draft is a starting point; careful human oversight ensures the tone and facts are right. The advantage is you can respond to narratives or publish statements much faster, without sacrificing quality. In crises, that speed can contain a story before it spirals.

4. Optimize for Generative Engine Optimization (GEO)

The AI-focused optimization strategies we have discussed throughout this paper can be characterized as Generative Engine Optimization, or GEO. The term GEO was introduced in a 2023 paper published by researchers from Princeton University, Georgia Tech, the Allen Institute for AI, and IIT Delhi in late 2023.⁵⁷ Unlike traditional SEO, which targets page rankings in search results, GEO focuses specifically on how generative AI systems like ChatGPT, Claude, Perplexity, and Gemini cite and reference content when synthesizing responses to user queries.

The paper's findings, which were based on testing nine optimization methods across 10,000 search queries, revealed that effective GEO implementation can boost source visibility by up to 40% in AI-generated responses.

GEO strategies for reputation management center on content credibility and structure. Citations are a top performer, with content that includes authoritative sources seeing significant visibility gains. Data integration and expert quotations are also crucial, adding credibility signals that AI systems consistently recognize and value.

Generative engines reformulate complex user queries into simpler components, then use summarizing models rather than traditional ranking algorithms to synthesize responses. Content optimized for this process includes features such as clear headings, structured Q&A formats, and logical information hierarchies that facilitate AI parsing and extraction.

Measurement techniques must also adapt to GEO. Traditional metrics like click-through rates become less relevant when AI systems provide direct answers rather than driving traffic to source websites. Instead, reputation managers need to track reference rates: how frequently

⁵⁷ <https://arxiv.org/abs/2311.09735>

their content appears as source material in AI responses. This shift requires monitoring tools that can track brand mentions and sentiment across generative platforms, ensuring accurate representation in this new information ecosystem, where being cited matters more than being clicked.

Conclusion

As we've explored throughout this white paper, AI is reshaping the future of reputation management at a pace and scale that demands attention. The past two years alone have demonstrated both the promise and perils of AI in the reputation arena: from generative AI tools enabling richer engagement and faster analysis, to deepfake and misinformation threats that can upend trust in an instant. For corporations, executives, and public figures, successfully navigating this new reality is challenging, but success is achievable with the right knowledge and strategy.

A few key themes emerge from our 2025 analysis:

- **Adaptability is essential.** Just as the PR industry evolved with the rise of search engines in the early 2000s and social media in the 2010s, it must evolve again for AI in the 2020s. Those who update their tactics (e.g. optimizing for AI-driven search, responding to AI-fueled crises) will thrive, whereas those clinging to traditional methods risk blind spots. As we noted in 2023, the PR field is poised for transformation; indeed, we are now witnessing a "new breed" of professionals who blend communications with tech fluency. Ensuring your team, internal or via partners, is equipped with AI skills is no longer optional.
- **Proactivity and vigilance can prevent many issues.** By investing in monitoring (both what humans and AIs are saying) and strengthening your online information ecosystem, you can catch inaccuracies before they spread and inoculate your reputation against falsehoods. This means carving out time and budget for continuous reputation management, not just when a crisis strikes. It also means being proactive in telling your story and building goodwill, so that if something negative or fake emerges, there's a reservoir of positive truth to counterbalance it.
- **Misinformation and deepfakes are formidable, but not undefeatable.** We've seen legal frameworks start to form, and major platforms taking them more seriously (with improved policies and algorithms). Public awareness of these issues is also higher than ever, which can actually help. People are learning to double-check shocking claims. Preparing and responding swiftly enables us to mitigate the damage of deepfakes, even as they become more widespread. The reputational hits from AI-fueled incidents are not necessarily permanent if managed well; in fact, handling an AI crisis adeptly can showcase an individual's or company's integrity and competence.
- **Leverage AI's advantages for reputation enhancement.** The flip side of the coin is that AI offers powerful tools to build and boost reputation: deeper insights into audience

sentiment, innovative ways to engage (like personalized content), efficiency in content creation and distribution, and more. Incorporating these into a proactive strategy—from AI-guided content calendars to chatbots that improve customer satisfaction—can differentiate you positively. Being seen as technologically adept and forward-thinking can itself be a reputational asset, especially in industries where innovation is valued.

- **Fundamentals still apply.** Amid all the change, some principles remain constant. Honesty, consistency, and delivering on promises underpin a good reputation, whether spread by word of mouth or by algorithm. AI doesn't change ethical best practices; if anything, it makes them more important, as missteps can be amplified. A quote from our 2023 conclusion rings true today: "We believe there will always be a role for counselors to help brands and people manage their reputation... PR professionals will need to be proactive in monitoring and ensure that their complete content portfolio is optimized for AI-powered tools while helping clients create compelling content that reaches audiences directly."

Reputation strategy should be a living program that iterates with technological and societal shifts. We concluded our 2023 white paper by noting that, "In his 1859 book, *The Origin of Species*, Charles Darwin posited that the species that adapt best to their changing environment have the best chance of surviving, while those who do not adapt do not make it. The same can be said for those that work in the reputation management industry."

This remains as true as ever. At Status Labs, we have spent the last two years actively adapting to AI, a technology that we believe is driving a fundamental and permanent shift in the way the world engages with information. Like the advancement of AI itself, our adaptation process is ongoing but progressing quickly, and we are excited about the power of utilizing AI-forward strategies to manage and protect our clients' reputations.