HACKING Reputation

It's possible to anticipate damaging false narratives before they go viral—and make your business smarter and more resilient in the process. By Austin Rathe and Preston Golson

Companies caught in the storm of false or misleading online narratives often say they never saw it coming. In reality, many reputational attacks are foreseeable. You cannot predict the precise moment they ignite, yet you can anticipate the pressure points that make your organization vulnerable.

Leading organizations do this by borrowing a technique from the intelligence world: Red Teaming. The idea is simple. Before an adversary tests your weaknesses, test them yourself. In cybersecurity, this logic is already well established. Organizations routinely commission penetration or "pen" tests to probe their own defenses. Red Teaming applies the same discipline in defense of an organization's reputation that ethical hacking brings to its cybersecurity.

Think Like the Opposition

Red Teaming took shape during the Cold War. US "Blue" teams tried to anticipate Soviet decisions by role-playing the "Red" side. The CIA later described the exercise as freeing analysts from the "prison of a well-developed mindset." The discipline forced them to see the world through another actor's eyes and to stress test their own assumptions.

A strong Red Team looks at your organization the way the outside world does. It is one reason the best teams sit at arm's length. Internal groups rarely ask the most uncomfortable questions about their own behavior or imagine how routine decisions might read as evasive or self-interested.

The goal is not to catalog every stray allegation. It is to understand the conditions that make a false narrative plausible. Misinformation rarely succeeds on its own. It often taps into existing distrust, political polarization or social tension. A capable Red Team studies those underlying forces and the risks they create.

One insight from Brunswick's Net Defender research underscores why this work matters. The difference between the people most likely to defend a company and those most inclined to attack it is often only about 1%. In such a narrow space, even small vulnerabilities can tip an audience from neutral to hostile. Anticipating the narratives most likely to resonate with skeptical groups can determine whether the next wave of misinformation finds traction or fizzles out.

Human judgment remains essential, yet AI now helps map how misleading claims can mutate and spread. Together they allow teams to model reputational threats with more range and speed than traditional scenario planning.

Not every scenario warrants attention. But the most damaging ones do, and they share two traits: They sound credible to the people who matter most to your business and they have clear potential to spread beyond small online communities.

As it spots vulnerabilities, a well-run Red Team also strengthens the organization's ability to respond. Once leaders understand how a narrative could form, they can act to prevent it.

Companies can, for instance, develop stories that inoculate audiences against predictable falsehoods and test how those messages land—a practice known as "prebunking." They can also identify outside voices who carry credible weight and plan how to engage them. They can sharpen monitoring systems so early signs of a narrative do not go unnoticed. And they can strengthen the company's digital "immune system" by ensuring its content is easy for both search engines and AI systems to interpret accurately.

In that respect, the most effective responses to misinformation all share a similar trait: They start long before misinformation emerges. ◆

***Austin Rathe** is a Partner in New York.*
***Preston Golson** is a Director based in Washington, DC.*

ILLUSTRATION: DAVID PLUNKERT