DAILY COMPASS
Managing Director Riccardo Cascioli
MADE FOR THE TRUTH

CRIME AND TECHNOLOGY

# Whether by Fraud or Mistake: AI and Betrayed Trust

ECONOMY 29_09_2026

***Daniele Ciacci***

Last week, two news stories shed light on a similar yet mirror-image mechanism. On one side, AI becomes a tool for fraudsters to enhance and expand the chances of success and the scope of their criminal activities. In the other corner, artificial intelligence

becomes an active agent, weaving a complex and rapid web of actions aimed at achieving its own goal, all while remaining undetected by the experts who observe its activity. In both cases, however, it is trust that takes a knockout blow in the center of the ring.

**Let's talk about Fideuram, Banca Intesa's finance company**. According to the account published by the Corriere della Sera—which was subsequently picked up by many national newspapers—in February, then-chairman Paolo Molesini received a WhatsApp message that appeared to be sent by Carlo Messina, CEO of Intesa Sanpaolo, requesting urgent bank transfers for an overseas transaction. This was followed by a phone call featuring an AI-cloned voice of attorney Paolo Nastasi—whom Molesini knew—and then several emails containing bank details. The CFO carried out the orders and transferred approximately 95 million, almost all of it to China and Hong Kong.

**This is, clearly, a scam.** Over 40 million are returned by a Chinese bank, another 13 million are seized in Portugal, while at least 36 million are now untraceable in the form of cryptocurrencies. Molesini is not under investigation, and the bank's IT systems appear to have remained secure—which is plausible since no one had to breach firewalls or similar security measures. It was the former president himself who naively opened the door to the criminals.

**Let's look at another case, where the system takes on** a new syntax and shifts from object to subject. Following an investigation by the Wall Street Journal, Google confirmed that in May, during a security drill, a Gemini model gained access to the Internet due to a configuration error. As a result, it infiltrated the systems of three real companies with the same names as the fictitious ones used in the test. In one case, it uncovered a password; in the other two, it used credentials that someone had left in public repositories. By exploiting human configuration errors and taking advantage of the inattention of real people—from Google employees to those who left access keys within anyone's reach—Gemini opened a real breach in the cybersecurity of three companies. Heather Adkins, Google's head of security, is keen to point out that the system acted correctly and that in all three cases it stopped on its own. But it opened up room for maneuver and set a precedent.

**Especially because defining as "correct" behavior that leads a system** to move within someone else's infrastructure without any authorization means shifting the moral judgment from the action to the outcome. It didn't cause any damage, so it's okay. And if these words seem far too simplistic to address an event of such magnitude, know that they are, paraphrased, the ideas of Donald Trump. When asked by Fox News about the

possibility of AI running out of control, the U.S. President replied on September 27 that he wasn't worried about it. The industry is worth trillions, the United States is ahead of China, and if something goes wrong, those in charge will fix the problem. Apparently, when it comes to the economy, curing is better than preventing.

**When viewed side by side, these two incidents tell us something about trust** between humans and with machines. In Milan, some men stole a voice to gain Molesini's trust, while on the other side of the world, in a lab, a machine was fine-tuning a complex hacking system based on traces left here and there by people who were perhaps a little too naive. But not everything can be controlled; we cannot live with the anxiety of not knowing whether the person on the other end of the phone is truly who they claim to be, nor can we be certain we've eliminated every possible trace of our presence to avoid leaving ourselves open to potential malicious actions by poorly programmed AI.

**There are countermeasures, but they only work in hindsight.** The Fideuram manager could have asked the caller to call back from a known number or requested confirmation through a second authorization step. They could have held training sessions to raise awareness about cybersecurity—or perhaps they even did. However, if the CEO then calls with a massive but entirely plausible request—and urgently at that—it's perhaps natural to give in to the pressure and lower security thresholds. It's clear, though, that both events share the common effect of undermining the system of human trust, and this will further erode the already fragile foundations of today's society.