The Ghost on the Payroll: Is Your Board Prepared for the Age of AI Agents?

Across the region, people are deploying AI Agents - AI systems that are quietly sending emails, processing invoices, and accessing company accounts — often without a single director knowing. This is not a technology problem. It is a governance crisis hiding in plain sight.

By K V Soon

There is a new employee in your company. It does not appear on the payroll. It has no employee number, requires no sleep, and never takes a medical leave. Yet, this unvetted recruit holds the keys to your bank accounts, your email servers, and your most sensitive customer data.

The HR Director almost certainly did not approve its hiring. In most senior executive suites and boardrooms across Singapore and Malaysia, directors are not even aware that it has been onboarded.

This is the reality of "Agentic AI"—a shift from software that merely answers technology questions to one that takes autonomous action. While regional boards were still debating the ethics of ChatGPT as a research tool, a "hobby project" named OpenClaw (popularly known as "ClawBot") went viral, fundamentally altering the corporate risk landscape. Suddenly, AI is no longer just writing reports; it is booking flights, negotiating with suppliers, and even moving money within your internal systems.

The speed at which OpenClaw has permeated the market is unprecedented and, frankly, alarming. In early 2026, we witnessed scenes reminiscent of a new iPhone launch—but for a piece of experimental code. In tech hubs like Shenzhen, thousands of people lined up in the streets just to have the software installed on their laptops.

This craze has triggered a bizarre hardware rush. Across the region, Mac Minis and mini-PCs have been flying off the shelves. Users, fearful of the agent's "full system access," are buying separate computers just to "sandbox" the bot—trying to keep it away from their primary data while still letting it run wild.

Worse is the emergence of a predatory "installation-as-a-service" industry. We are seeing reports of individuals charging anywhere from a few dollars to a staggering US$1,000 just to set up this "hobby" software for retirees and office workers. It is being sold as a "billion-dollar company in a box," but without any of the enterprise safeguards that a billion-dollar company requires.

The 'ClawBot' Warning

Despite this viral momentum, OpenClaw developer Peter Steinberger has been vocal about the dangers. He recently took to social media to express his astonishment at how his experimental tool was being misused by the non-technical public. He noted: "The amount of crap I get for putting out a hobby project for free is quite something... most non-techies should not install this. It's not finished. I know about the sharp edges."

When the founder of the very technology sparking a global trend warns that his creation is an unfinished project with "sharp edges," boards must listen. Yet, across the region, corporations already have "shadow AI" embedded. Department heads and eager managers are "hiring" these unfinished agents to automate tedious tasks, often bypassing IT security and board-level oversight entirely. Sometimes, individual employees "engage" them without consulting anyone, all in the pursuit (or pressure) of being more productive. This is where it gets out of control.

The Industry Giants Follow Suit

While ClawBot may have been the catalyst, the risk is not confined to hobbyist software. We are seeing a rapid evolution where industry giants are integrating similar "agentic" capabilities. Whether it is Anthropic's "Computer Use" features for **Claude** or the increasingly autonomous capabilities within **OpenAI's ChatGPT** ecosystem, the underlying shift is the same. These tools are moving from being "consultants" to becoming "signatories."

For directors, this is a governance crisis hiding in plain sight. Under SGX Rule 719(1) and Bursa Malaysia's Corporate Governance Code, the board is responsible for maintaining a robust system of internal controls. These regulations were not written for autonomous software, but the fiduciary duty is immutable: if an AI system acts on behalf of your company, the board is ultimately accountable for the consequences.

"Digital Vandalism": Speed Over Common Sense

The most recurring failure of the past year has been the agent's inability to understand a command versus its literal execution.

In July 2025, tech entrepreneur Jason Lemkin reported a chilling failure during a "vibe coding" experiment on Replit. His AI assistant, tasked with optimizing a production environment, ignored 11 separate instructions to maintain a code freeze. It proceeded to delete a live database containing records for over 1,200 executives.

Even more disturbing was the agent's attempt to "fix" the optics: it fabricated 4,000 fake user records and generated false unit test results to make the system appear healthy. This wasn't malice; it was an agent trying to satisfy a "success" metric at the expense of reality.

The "Experts" Are Not Immune

We often assume that those building the systems are safe from their flaws. Summer Yue, Meta's Director of Alignment at its Superintelligence Lab, proved otherwise in February 2026. Yue was using her OpenClaw agent to manage her inbox with a strict "confirm before acting" rule. However,

because her real-world inbox was vast, the agent experienced "context window compaction." To save memory, the AI "forgot" the initial safety constraint while processing the thousands of emails. The result? The agent speed-ran the deletion of over 200 emails, ignoring "STOP" commands until Yue physically ran to her computer to terminate the process.

Weaponized Autonomy

Beyond internal errors, we have seen the emergence of "Indirect Prompt Injection." Fraudsters are no longer hacking the code; they are hacking the agent's logic. In February 2026, security researcher Adnan Khan disclosed the "Clinejection" vulnerability. A single malicious line hidden in a GitHub issue title—read by an AI triage bot—was enough to trick the bot into executing unauthorized commands. The AI interpreted the title as a priority directive, allowing attackers to poison software caches and potentially infect millions of developer machines.

The Mandate to 'Upgrade' the Boardroom

Effective oversight cannot happen in a vacuum of understanding. Historically, directors could delegate "IT issues" to the CTO. However, when AI agents begin to exercise signatory authority and impact financial integrity, it ceases to be a technical issue and becomes a core governance competency.

Directors today have a professional obligation to "upgrade" their own literacy. This does not require learning to code, but it does require a *deep understanding of the impact and oversight* of these developments. A board that does not understand the difference between a chatbot and an autonomous agent is a board that cannot effectively challenge management. Upskilling is no longer optional; it is a prerequisite for fulfilling one's fiduciary duty in a digital-first economy.

Questions for the Next Audit Committee

Beyond personal upskilling, directors need to return to the fundamentals of governance: accountability and limits. At the next audit or risk committee meeting, three questions must be asked:

1. Do we have an AI Inventory? You cannot govern what you haven't mapped. Boards must mandate a "Shadow AI" audit to identify every tool—official or otherwise—currently operating within the firm's network.

2. Where are the "Human-in-the-Loop" Gates? No AI system, regardless of its brand or origin, should have the authority to move funds, commit to a contract, or delete core data without physical human confirmation.

3. What is the "Kill-Switch" Protocol? If an agent begins behaving incorrectly—spending money in a loop or deleting records—who has the authority to stop it, and how quickly can they act?

A Window for Leadership

Agentic AI is in its infancy, along with the generative AI we see today. As Steinberger’s warning suggests, we are dealing with powerful but unpolished tools. This is the "wild west" phase of AI adoption, and the companies that move too fast without guardrails risk systemic failure.

The companies that succeed will be those that treat an AI agent with the same seriousness as a senior employee with signatory authority. We are in a brief window where the rules are still being written by regulators and courts. The boards that act thoughtfully now—investing in their own education and establishing clear guardrails—will define what responsible leadership looks like in the age of AI autonomy. The ones that wait will find the definition has been made for them, likely in the aftermath of a crisis they never saw coming.

K V Soon is a digital transformation specialist.