wtw | Willis

Weaponising AI

AI wallet attacks

April 2026



wtwco.com

Overview

As AI capabilities are increasingly embedded into financial systems, we must consider the expanding threat landscape we face and emerging novel attacks. **AI wallet attacks** represent a new class of cyber and financial risk emerging at the intersection of wallets, AI agents, and usage based AI token systems.

Unlike traditional wallet attacks that focus on stealing private keys or funds, AI wallet attacks exploit automation, delegation of authority, and economic consumption models to cause operational, integrity or financial failures without compromising the security of the systems.

The threat landscape now spans **three interrelated attack objectives**:

1. **Wallet compromise or coercion** (AI assisted theft)
2. **Denial of AI wallets** (availability and cost exhaustion)
3. **AI token misdirection and mis-consumption** (integrity and economic abuse)

These attack types are increasingly observed in crypto ecosystems, AI agent platforms, and AI enabled financial tooling between 2025–2026. [kucoin.com], [chainup.com]

wtw | Willis

1. AI wallet attacks

An AI wallet attack occurs when attackers use AI systems, or target AI based wallet controllers, to cause unauthorised or unintended financial actions.

Common patterns

- **AI assisted social engineering:** deepfake voice, video, or chatbots impersonating trusted parties to induce wallet approvals. [techrepublic.com], [helpnetsecurity.com]
- **AI optimised wallet drainers:** interfaces or scripts that dynamically adapt to victims, coaching them into signing malicious approvals rather than stealing keys. [chainup.com]
- **AI agent compromise:** autonomous agents (trading bots, portfolio managers, "AI assistants") legitimately execute harmful actions once misled or compromised. [cointelegraph.com]

The key shift is a focus on decision authority, as opposed to cryptographic security, as the primary attack surface.

2. Denial of AI Wallets

Denial of AI Wallet attacks focus on rendering an AI controlled wallet unusable or economically unsustainable, rather than stealing funds. The emerging term for this is Denial of Wallet (DoW), adapting denial of service concepts to usage billed AI systems. [layerxsecurity.com]

Because AI wallets and agents commonly operate on usage billed models, attackers deliberately trigger excessive or pathological operations, causing:

- Token or API quota exhaustion
- Budget cap triggers
- Rate limiting or safety shutdowns
- Gas depletion in on chain automation

The roll-on effect can stretch to:

- **Supply chain denial:** compromised AI libraries causing runaway compute or forced shutdowns. [cryptorank.io], [ox.security]

Importantly, the attacker does not pay — the victim bears the cost. [layerxsecurity.com]

For time sensitive transactions, such attacks leave the wallet as effectively unavailable whilst leaving the funds intact.



3. AI token mis-consumption (economic exhaustion)

AI wallets rely on **AI tokens** which consume different varieties of API tokens or keys. Denial of AI wallet attacks typically exhaust these tokens first, making token consumption the real choke point. This reframes wallet security as **economic safety**.

Token mis consumption occurs when:

- AI reasoning chains are artificially prolonged
- Agents are triggered to perform unnecessary computation
- Benign looking actions rack up unbounded token costs

Collectively, this is forcing us to consider DoW as a focused attack pattern, not a side effect. [layerxsecurity.com]

4. AI token misdirection (integrity without breach)

Attacks traverse through the threat landscape, responding to organisations that adequately protect token budgets and limits, by pivoting from **exhaustion** to **misdirection**.

AI token misdirection occurs when valid, authorised and legitimate tokens are used within permitted scope but for unintended contexts, consumed to maintain attacker preferred system states and redirected toward harmful or blocking actions.

But outcomes diverge from user intent, representing an integrity failure of decision making, not access control. [kucoin.com], [cointelegraph.com]



Why these threats converge

Attackers continue to push the bounds of technology use, or misuse, in the perpetual cat-and-mouse game of cybercrime and security.

Cybercriminals consistently adapt their techniques to monetise emerging technologies, migrating across the CIA Triad, targeting each of the three pillars being Confidentiality, Integrity and Availability as novel attacks to new technologies:

Attack Goal	Mechanism	Primary Impact
Steal	AI deception	Confidentiality & integrity
Deny	Token exhaustion	Availability
Misdirect	Token misuse	Integrity
Escalate	Forced shutdown	Business continuity

All of these attacks ultimately depend on AI token governance, not your traditional security mechanisms of cryptography applied to the wallet. They do however converge to significant business impacts, causing business continuity plans to be tested.

As a result, industry guidance has shifted from “wallet security” toward:

- AI capability governance
- Token scoped authority
- Spend aware automation
- Kill switches tied to token velocity

Strategic takeaway

AI wallet attacks are no longer about breaking cryptography — they are about abusing delegation, decision making, and token economics. Denial of AI wallets and token misdirection are natural evolutions of the same problem: unbounded or poorly governed AI authority acting at financial speed.

Author

Michael Parrant
Cyber & Technology | FINEX, Pacific
M +61 406 246 946
E: michael.parrant@wtwco.com

Contact

Eden Fletcher
Head of FINEX, Pacific
M +61 492 983 630
E: eden.fletcher@wtwco.com

Stephen Trickey
Head of Growth | FINEX, Pacific
M: +61 449 699 439
E: stephen.trickey@wtwco.com

Disclaimer: While all reasonable skill and care has been taken in preparation of this document it should not be construed or relied upon as a substitute for specific legal advice from your legal adviser, or for specific advice on your insurance needs. No warranty or liability is accepted by Willis Australia Limited its shareholders, directors, employees, other affiliated companies for any statement, error or omission.

Willis Australia Limited | ABN 90 000 321 237 | AFSL No. 240600 |
Copyright © 2026 WTW. All rights reserved | WTW1016AU

wtw | Willis