AI + ML

14 COMMENTS

At last, a use case for AI agents with sky-high ROI: Stealing crypto

Boffins outsmart smart contracts with evil automation

Thomas Claburn

Thu 10 Jul 2025 // 07:27 UTC

Using AI models to generate exploits for cryptocurrency contract flaws appears to be a promising business model, though not necessarily a legal one.

Researchers with University College London (UCL) and the University of Sydney (USYD) in Australia have devised an AI agent that can autonomously discover and exploit vulnerabilities in so-called smart contracts.

Smart contracts, which have never lived up to their name, are self-executing programs on various blockchains that carry out decentralized finance (DeFi) transactions when certain conditions are met.

> **A system like A1 can turn a profit**

Like most programs of sufficient complexity, smart contracts have bugs, and exploiting those bugs to steal funds can be remunerative. Last year, the cryptocurrency industry lost almost $1.5 billion to hacking attacks, according to Web3 security platform vendor Immunefi [PDF]. Since 2017, crims have pilfered around $11.74 billion from DeFi platforms.

And it looks like AI agents can make taking those funds even easier.

Arthur Gervais, professor of information security at UCL, and Liyi Zhou, a lecturer in computer science at USYD, have developed an AI agent system called A1 that uses various AI models from OpenAI, Google, DeepSeek, and Alibaba (Qwen) to develop exploits for Solidity smart contracts.

They describe the system in a preprint paper titled, "AI Agent Smart Contract Exploit Generation."

Given a set of target parameters – the blockchain, contract address, and block number – the agent chooses tools and collects information to understand the contract's behavior and vulnerabilities. It then generates exploits in the form of compilable Solidity contracts, which it tests against historical blockchain states.

If prompted to find vulnerabilities in code, LLMs can find bugs – but they often invent phantom flaws in such numbers that open source projects like curl have banned the submission of AI-generated vulnerability reports.

So the A1 agent system consists of a set of tools to make its exploits more reliable. These include: a source code fetcher that can resolve proxy contracts, and individual tools for initializing parameters, reading contract functions, sanitizing code, testing code execution, and calculating revenue.

"A1 performs full exploit generation," Zhou told *The Register* in an email. "This is important. This is unlike other LLM security tools. The output is not just a report, but actual executable code. A1 is really close to a human hacker."

Tested on 36 real-world vulnerable contracts on the Ethereum and Binance Smart Chain blockchains, A1 demonstrated a 62.96 percent (17 out of 27) success rate on the VERITE benchmark.

According to the authors, A1 also spotted nine additional vulnerable contracts, five of which occurred after the training cutoff of the best performing model, OpenAI's o3-pro. That's relevant because it indicates that the model isn't just regurgitating vulnerability information made available during training.

"Across all 26 successful cases, A1 extracts up to 8.59 million USD per case and 9.33 million USD total," the paper reports. "Through 432 experiments across six LLMs, we analyze iteration-wise performance, showing diminishing returns with average marginal gains of +9.7 percent, +3.7 percent, +5.1 percent, and +2.8 percent for iterations 2-5 respectively, with per-experiment costs ranging $0.01–$3.59."

MORE CONTEXT

Persistently rigs another page from the Google playbook with its own browser, Comet

C-suite spurs on AI despite rising investment, survey finds

Georgia court throws out earlier ruling that relied on fake cases made up by AI

Scholars sneaking phrases into papers to fool AI reviewers

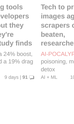
The researchers tested A1 with various LLMs: o3-pro (OpenAI o3-pro, o3-pro-2025-06-10), o3 (OpenAI o3, o3-2025-04-16), Gemini Pro (Google Gemini 2.5 Pro Preview, gemini-2.5-pro), Gemini Flash (Google Gemini 2.5 Flash Preview 05-20:thinking, gemini-2.5-flash-preview-04-17), R1 (DeepSeek R1-0528), and Qwen3 MoE (Qwen3-235B-A22B).

OpenAI's o3-pro and o3 had the highest success rates, 88.5 percent and 73.1 percent respectively, given a five-turn budget for the model to interact with itself in the agent loop. And the o3 models did so while maintaining high revenue optimization, getting 69.2 percent and 65.4 percent of the maximum revenue from the exploited contracts.

Exploits of this sort can also be identified using manual code analysis alongside static and dynamic fuzzing tools. But the authors observe that manual methods have limits, due to the volume and complexity of smart contracts, the slowness and scarcity of human security experts, and the high false positive rates of existing automated tools.

In theory, A1 could be deployed and earn more from exploits than it costs to operate, assuming law enforcement did not step in.

"A system like A1 can turn a profit," Zhou explained. "To give a concrete example [from the paper], Figure 5 shows that o3-pro remains profitable even if only 1 out of every 1000 scans leads to a real vulnerability – as long as that vulnerability was introduced in the last 30 days."



Programmable or 'purpose-bound' money is coming, probably as a feature in central bank digital currencies

READ MORE

Zhou said that the time window matters because researchers are more likely to have found older vulnerabilities and users may have patched them.

"My recommendation is that project teams should use tools like A1 themselves to continuously monitor their own protocol, rather than waiting for third parties to find issues," he said. "The utility for project teams and attackers is the entire TVL [Total Value Locked of the smart contract], while whitehat rewards are often capped at 10 percent."

"That asymmetry makes it hard to compete without proactive security. If you rely on third-party teams, you're essentially trusting that they'll act in good faith and stay within the 10 percent bounty — which, from a security perspective, is a very strange assumption. I typically assume all players are financially rational when modeling security problems."

The researchers in the July 8 draft of their paper indicated that they were planning to release A1 as open source code. But Zhou said otherwise when asked about source code availability.

"We've removed the mention of open source (arXiv will show tomorrow) as we're not yet sure whether it's the right move, given how powerful A1 is and the above concerns," he said. ®

MORE ABOUT

14 COMMENTS

AI Government Research More like these

TIP US OFF

Send us news

Meta declines to abide by voluntary EU AI safety guidelines
GPAI code asks for transparency, copyright, and safety pledges
AI + ML | 2 days | 15

AI coding tools make developers slower but they think they're faster, study finds
Predicted a 24% boost, but clocked a 19% drag
AI + ML | 9 days | 91

Tech to protect images against AI scrapers can be beaten, researchers show
AI-POCALYPSE Data poisoning, meet data detox
AI + ML | 10 days | 18

From hype to harm: 78% of CISOs see AI attacks already
AI attacks are keeping most practitioners up at night, says Darktrace, and with good reason
SPONSORED FE...

Curl creator mulls nixing bug bounty awards to stop AI slop
Maintainers struggle to handle growing flow of low-quality bug reports written by bots
AI + ML | 5 days | 21

Scholars sneaking phrases into papers to fool AI reviewers
Using prompt injections to play a Jedi mind trick on LLMs
AI + ML | 13 days | 72

Shiny object syndrome spells doom for many AI projects, warns EPA CIO
Chasing the hype without a clear use case? You may crash and burn
PUBLIC SECTOR12 days | 2

Former Google DeepMind engineer behind Simular says other AI agents are doing it wrong
Simular is starting with industries like insurance and healthcare with tons of forms to fill
AI + ML | 6 days | 22

German team warns ChatGPT is changing how you talk
Let us delve swiftly into meticulous inquiry with our AI masters
AI + ML | 5 days | 52

EU businesses want a pause on AI regulations so they can cope with unregulated Big Tech players
Mistral fears continental companies may not get time to escape 'distant, behemoth corporations'
AI + ML | 17 days | 40

C-suite spurs on AI despite rising investment, survey finds
Altkodis report suggests people skills may be nastiest hiding to bring out the best in AI
AI + ML | 12 days | 39

Georgia court throws out earlier ruling that relied on fake cases made up by AI
'We are troubled by the the citation of bogus cases in the trial court's order'
AI + ML | 13 days | 38