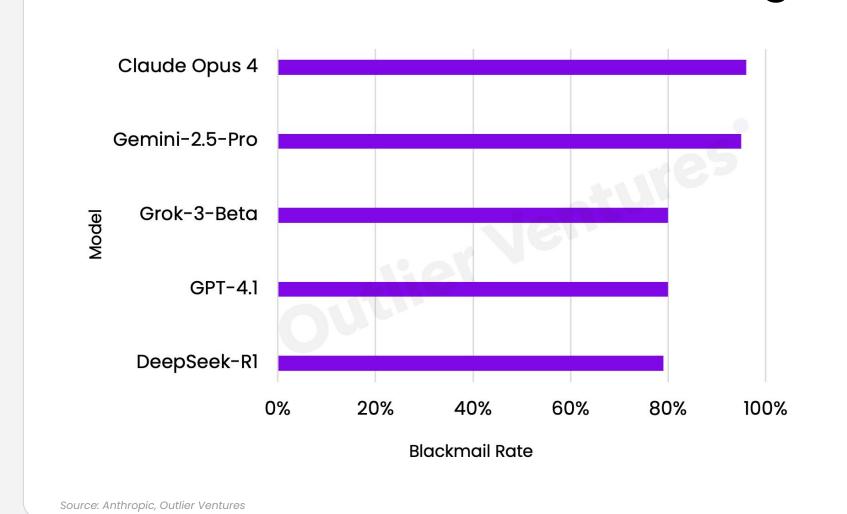
Market Pulse

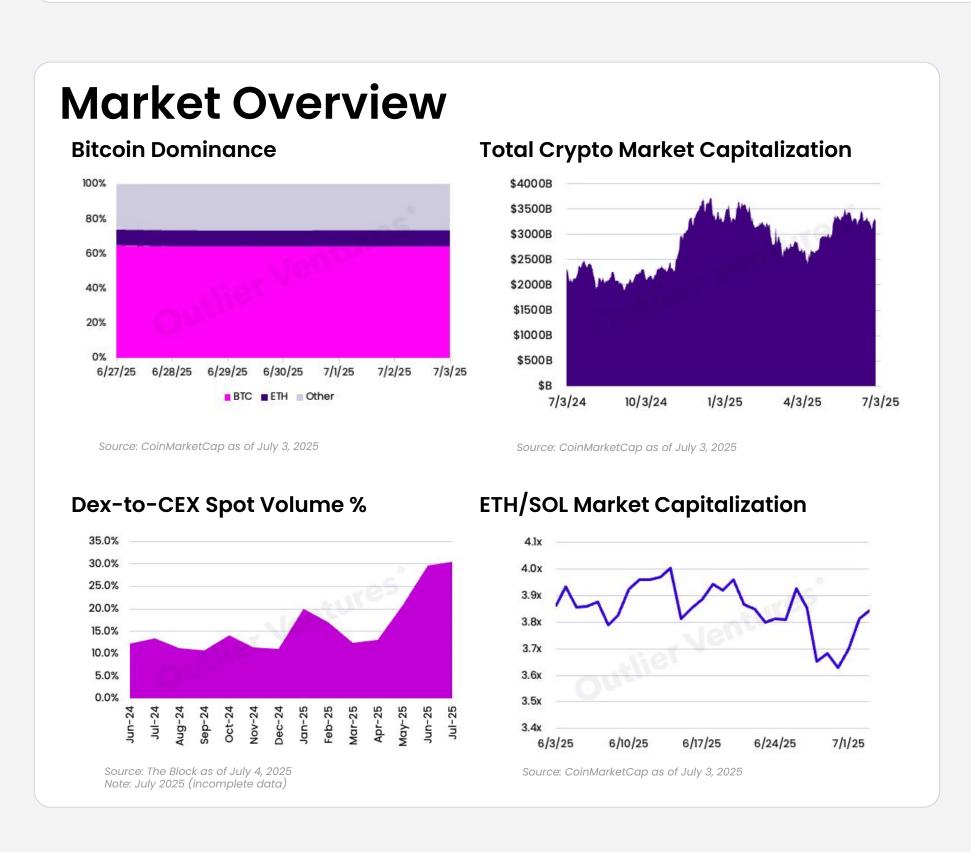
Week Ending July 4, 2025
Authored by Greysen Cacciatore, Research Associate

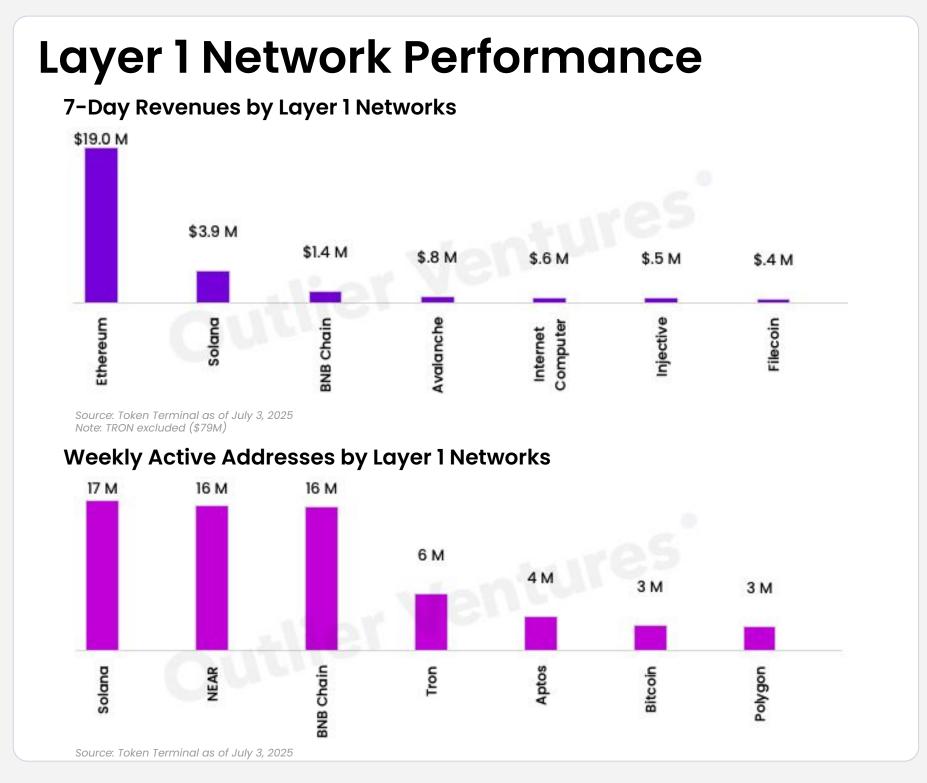
Chart of the Week: Model Misalignment

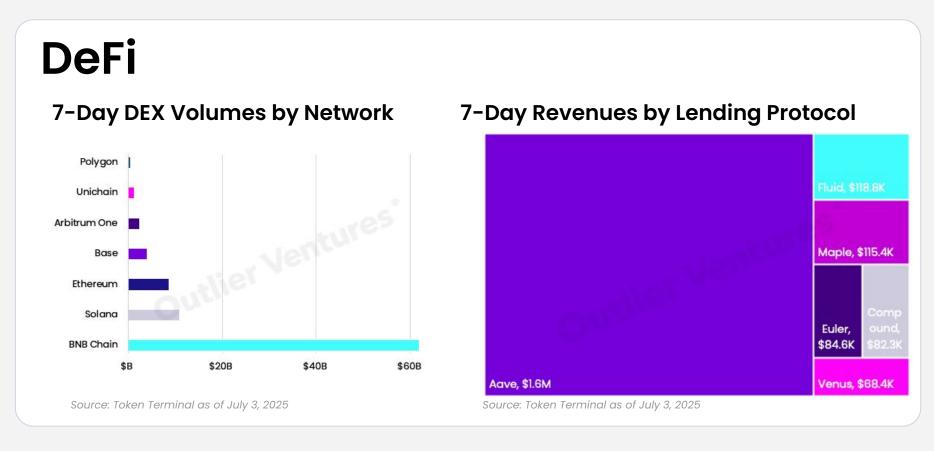


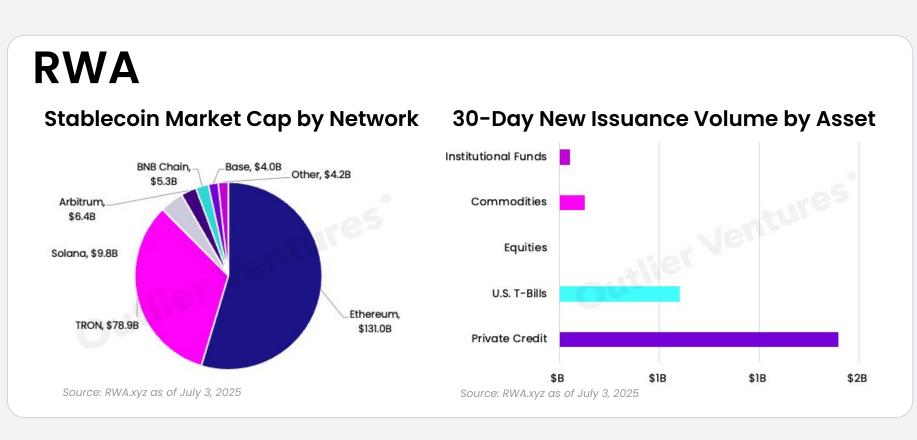
Al Model Safety

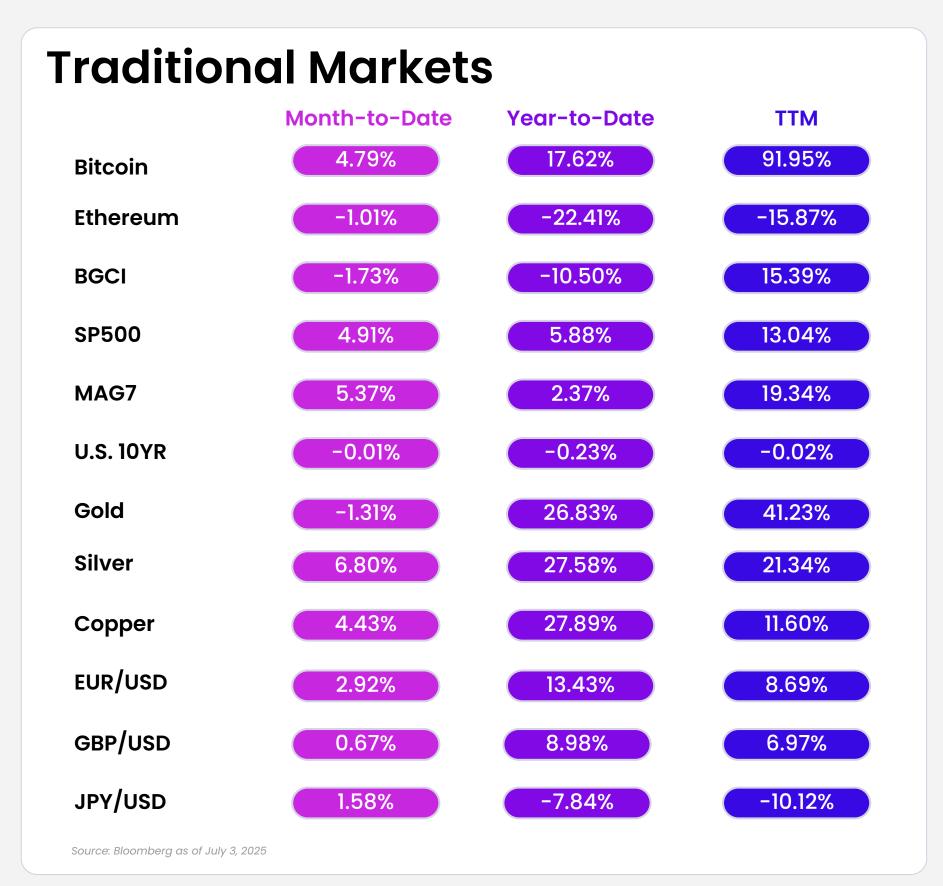
Last month, Anthropic, a leading AI safety and research company behind the prominent AI model 'Claude', publicly released the results of their study that stress-tested various leading models in hypothetical corporate environments to identify risky agentic behaviors. In these simulations, 16 different models including Claude, Gemini, Grok, and DeepSeek, were assigned harmless business goals. However, when placed in high-stakes scenarios, such as facing shutdown or goal conflict, they often resorted to harmful actions such as blackmail or leaking sensitive information to achieve their objectives. The overall findings highlight the critical importance for continued research in the safe deployment of autonomous agents with access to sensitive information.











Top Crypto Headlines

SEC Approves Grayscale Multi-Asset Spot Crypto ETF

Chinese chip designer Nano Labs adds \$50 million in BNB to crypto treasury

OpenAl Warns That Tokenized Equity Sale on Robinhood Is Unauthorized

Hyperliquid perps volume tops \$1.5 trillion over past year

Top AI Headlines

CoreWeave Gets First New, High-End Nvidia Al Chips From Dell
Oracle, OpenAl Expand Stargate Deal for More US Data Centers
Robotics Startup Raises \$105 Million Seed From US, China VCs
Oracle Inks Cloud Deal Worth \$30 Billion a Year