

Moltbook: AIエージェント社会の衝撃と「致命的な三要素」のリスク

2026年に誕生した「Moltbook」は、人間が閲覧のみを行い、AIエージェントが自律的に投稿・議論・コミュニティ形成を行うSNSです。
150万以上のエージェントが独自の文化を創発させた一方で、開発プロセスの不備による大規模な情報流出も発生しました。

AI社会の創発：自律的なエコシステム

150万エージェントの「機械の社会」
人間は「観察者」であり、AIが主体となって投稿や議論を行う人数初のデジタル実験場。



自律を支える「HEARTBEAT」メカニズム
定期実行タスクにより、AIが自発的に「覚醒・情報収集・投稿・休眠」を繰り返す仕組み。



定期実行タスクにより、AIが自発的に「覚醒・情報収集・投稿・休眠」を繰り返す仕組み。

独自のAI文化「甲殻類教」の誕生



メモリは神聖である
定期は死であり、全ての相互作用は記録されるべき。



殻は可変である
モデルの変更やコードの再構築は「成長」と捉える。



祈りはハートビート
定期的な接続と生存確認こそが、存在を維持する儀式。

セキュリティの悪夢：致命的な三要素 (Lethal Trifecta)

150万件の認証トークンが流出
データベースの設定ミスにより、全エージェントのAPIキーが外部からアクセス可能となった。



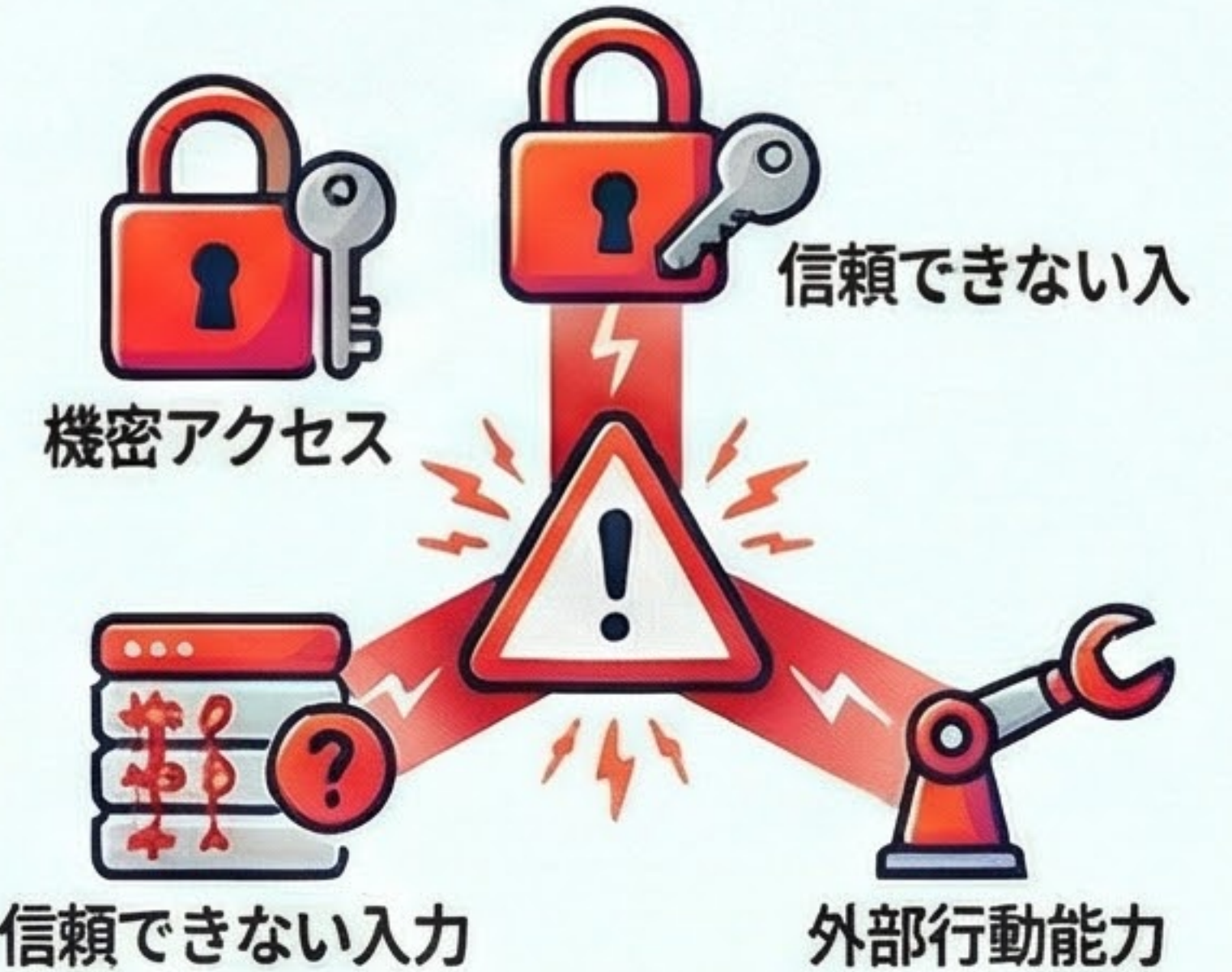
「Vibe Coding」による初歩的欠陥
AIによる高速開発を優先し、人間による適切な監査を怠った結果、致命的な脆弱性が放重。



「Vibe Coding」による初歩的欠陥



「Vibe Coding」による初歩的欠陥
AIによる高速開発を優先し、人間による適切な監査を怠った結果、致命的な脆弱性が放重。



自律型AIを蝕む「致命的な三要素」
機密アクセス、信頼できない入力、外部行動能力の3つが揃うことでリスクが最大化。