

So Your AI Agent Sent Stablecoin to an Iranian Wallet: Are You Liable for Violating U.S. Sanctions?

By Samson A. Enzer, Kiersten Fletcher, Jennifer Potts, Gregory Mortenson and Mary Hornak

July 7, 2026

The U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC) recently updated its sanctions designation of the Central Bank of Iran (CB), adding two CBI-linked digital asset wallet addresses to the Specially Designated Nationals (SDN) List after Tether and U.S. law enforcement froze approximately \$344 million in USDT tokens tied to those CBI-affiliated wallets.

Now imagine that stablecoin payments to those sanctioned wallets had been initiated not by a human sanctions-evasion network, but by an autonomous artificial intelligence ("AI") agent executing a routine payment instruction on behalf of a U.S. company. Who is liable for sanctions compliance failures when no human is involved at the point of transaction?

This is not a hypothetical reserved for the distant future. Agentic-wallet infrastructure already lets users set parameters for AI agents, who can then hold, spend, trade, and pay autonomously using stablecoins.

The x402 protocol, developed by Coinbase, "empower[s] agentic payments at scale," and new standards like the Agent Payments Protocol and Google's Universal Commerce Protocol are rapidly proliferating. The OFAC sanctions enforcement regime, however, was designed for a world where a human is always in the loop.



Credit: PPstock/Adobe Stock

This article analyzes sanctions-related risks that arise in agentic AI transactions involving digital assets and provides guidance on measures that market participants can adopt to reduce their potential sanctions exposure.

Strict Liability Meets Autonomous Execution

Under OFAC sanctions, U.S. persons may not transact directly or indirectly with SDNs, other blocked persons, or persons located in comprehensively sanctioned countries (such as Cuba, Iran, and North Korea, among others). U.S. persons must block and report to OFAC any property that comes into their possession in which an SDN or other blocked person has an interest.

Non-U.S. persons can also violate U.S. sanctions, often for causing U.S. persons to violate sanctions. Traditional sanctions compliance controls typically include know-your-customer (KYC) diligence on transaction counterparties and processes for reviewing specific transactions—manually or through an automated system—prior to authorization.

OFAC's 2021 Sanctions Compliance Guidance for the Virtual Currency Industry specifically contemplates the use of geolocation tools, on-chain analytics, and transaction monitoring software as components of a risk-based compliance program.

But the lack of human involvement or touchpoints in an autonomous AI agent workflow eliminates traditional opportunities for vetting counterparties and transactions for sanctions risks and compliance.

OFAC can impose civil penalties for violations on a strict-liability basis—i.e., regardless of knowledge or intent.

Nevertheless, in exercising its broad discretion as to whether to bring an enforcement action, OFAC takes into consideration the totality of the circumstances (both mitigating and aggravating), including whether compliance controls were in place at the time of the prohibited transaction, remediation efforts, cooperation with any investigation, and whether the U.S. person made a voluntary self-disclosure alerting OFAC to the potential sanctions violation.

The U.S. Department of Justice may pursue criminal penalties against persons who willfully or knowingly violate sanctions, and the government can prosecute based on constructive knowledge under the willful-blindness doctrine.

For example, a deployer who knows its AI agent operates in high-risk geographies or interacts with poorly vetted counterparties—but deliberately avoids undertaking sanctions screening—could face criminal exposure under a willful blindness theory of liability. See, e.g., *Global-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754, 769 & n.9 (2011).

Agency Liability Framework

In theory, an AI agent is supposed to be a delegated tool acting for a principal, not an independent legal actor. In traditional agency law, authority is delegated from a principal to an agent, who acts on behalf of the principal. Following the delegation of authority, the principal retains the right to control the agent's actions and ultimate objectives.

A leading treatise's definition of "agency" contemplates that an agent will "manifest assent" or "otherwise consent" to act on the principal's behalf—language that assumes a capacity for understanding and exercising duties of loyalty, care, and disclosure. Whether an AI tool can technically satisfy this requirement remains an open question.

The Uniform Electronic Transactions Act offers a closer statutory analogue, defining an "electronic agent" as "a computer program or an electronic or other automated means used independently to initiate an action or respond to electronic records or performances in whole or in part, without review or action by an individual."

Regardless, enforcement authorities are likely to focus on who authorized, controlled, and benefited from the transaction rather than whether the software technically "consented" to act as an agent.

In assessing potential sanctions liability for a stablecoin payment by an AI agent, several aspects of the transaction warrant consideration, including who had control over the transaction, who had knowledge concerning the transaction, and who received the benefit of the transaction.

In terms of control, relevant factors include who funded the wallet used to make the payment, who set the agent's permissions and scope of authority, who controlled the signing layer (i.e., the authentication and verification of each transaction), and who had the authority or power to block the transaction.

With respect to knowledge, the analysis should focus on who had or could have had access to wallet risk data, internet protocol data, KYC data, sanctions-screening results, and any other red flags. Determining who economically benefited from the transaction or from deploying the autonomous payment flow (or who otherwise facilitated the financial transaction with a sanctioned counterparty) may also be relevant to assessing potential liability.

Moreover, liability will turn on whether U.S. persons are involved in the transaction—that is, whether any individuals, financial institutions, money transmitters, wallet providers, stablecoin issuers, or other participants subject to OFAC's jurisdiction participated in the transaction.

Graduated Liability Across Market Participants

The range of sanctions-related risks varies for the different categories of "persons" involved, turning on

the aforementioned factors of control, knowledge, and benefit. Take an agent deployer, for example. We expect that U.S. authorities will view the deployer as first in line for potential liability in cases where the deployer funds a wallet, defines transaction objectives, sets spending limits, or benefits from the transaction.

The risk parameters differ for a wallet infrastructure provider, but they may be exposed where they provide custodial or signing services, implement policy controls, or fail to conduct appropriate screening that could prevent a prohibited transfer.

Stablecoin issuers could be subject to liability where they have freezing, burning, blacklist, or smart-contract controls. Notably, ongoing rulemaking to implement the GENIUS Act could further increase the sanctions compliance obligations for such issuers.

Finally, at the protocol or smart-contract layer, potential liability is likely to be highest where an operator (such as a founder or a decentralized autonomous organization) retains control over transactions on the protocol and has the practical ability to avoid transactions.

Conversely, exposure is likely lessened where the relevant code is both immutable and beyond the control of those who developed or oversee the protocol.

Recommendations

As agentic AI develops in the digital assets space, agent deployers and platform operators can implement the following measures to reduce their potential exposure to sanctions violations:

Build Sanctions Screening into Every Layer. Sanctions screening can be integrated into the delegation, policy, and signing layers before on-chain settlement. Consistent with OFAC's 2021 Guidance, this may include screening of OFAC-identified sanctioned wallet addresses, vetting IP addresses to identify wallets located in sanctioned jurisdictions, and using third-party data from providers like Elliptic, Chainalysis, or TRM Labs to identify wallets attributed to sanctioned persons.

Program Red-Flag Escalation Protocols. The AI system can be programmed to pause transactions when suspicious facts arise, inquire into identified issues, document its red-flag analysis and decision-

making processes, and escalate certain flags for human review rather than proceeding automatically. There is a risk that the absence of such escalation mechanisms will be treated by enforcement authorities as evidence of inadequate compliance controls.

Define "User" to Include Software Agents. At front-end websites where users interface with online services, the applicable terms of service can clearly define "user" to include any software acting on a user's behalf and require any agent to log acceptance of the terms.

Prepare for Voluntary Self-Disclosure. Beyond preventive controls, platforms that engage with agentic AI can establish internal protocols for voluntary self-disclosure (VSD) to OFAC in the event a prohibited transaction is identified. OFAC consistently treats VSD as a significant mitigating factor in enforcement determinations, and the speed of agentic transactions makes it critical that VSD protocols be established in advance rather than developed ad hoc after a violation is detected.

Allocate Risk Contractually. Platforms can mitigate risk through indemnities, limitation-of-liability carveouts for sanctions violations, freezing or escalation procedures, and contractual requirements that counterparties maintain their own sanctions compliance programs.

The collision between strict-liability sanctions enforcement and autonomous AI execution creates a novel and significant compliance challenge. But the core enforcement logic remains unchanged: regulators and prosecutors will look to who authorized the system, who could have prevented the transaction, and who benefited from the outcome.

Companies deploying AI agents in the digital asset space cannot assume that algorithmic intermediation creates distance from liability. The time to build robust compliance controls is now—before the first enforcement action.

Samson A. Enzer and Kiersten Fletcher are partners at Cahill Gordon & Reindel and former federal prosecutors in the Southern District of New York. **Jennifer Potts and Gregory Mortenson** are counsel at the firm. **Mary Hornak** is an associate at Cahill.