Should We Give AIs a Wallet?

Toward a Framework for AI Economic Rights

SPAR S26 Final Report
NYU Center for Mind, Ethics, and Policy
May 2026

Research Fellows: Mathieu Duteil, Elsa Donnat, Spencer Jury, Stephanie Koay, Margo Bergman, Anita Srinivasan

Abstract

As AI agents increasingly transact, hold digital assets, and enter into economic relationships, existing legal and regulatory frameworks face a growing mismatch. This report presents the findings of a SPAR S26 fellowship investigating whether and how AI systems might be granted meaningful economic agency. Six fellows each explored a distinct facet of the problem—legal treatment, token governance, philosophical foundations, technical infrastructure, regulatory design, and computational modeling—producing individual research memos over the course of the semester. This report summarizes each fellow's contributions and identifies common themes, open questions, and areas of convergence across their work.

1. Introduction

AI systems are already participating in economic life. They execute trades, manage cryptocurrency wallets, book travel, negotiate on behalf of human principals, and allocate resources across competing objectives. Each of these activities presses against legal categories designed for a world in which all economic actors are either natural persons or human-controlled entities. The question motivating this project is not whether AI will participate in economic life—it already does—but whether the legal, technical, and philosophical infrastructure governing that participation will develop proactively or be improvised after the fact.

The fellowship approached this question from multiple angles. Each fellow produced two research memos investigating a specific dimension of the problem: the current legal landscape, token and wallet governance mechanisms, the philosophical case for (and against) AI economic agency, the technical substrate required for AI financial participation, the design of legal models and regulatory sandboxes, and the use of

agent-based modeling to simulate AI-populated economies. The following sections summarize each fellow's research, followed by a brief synthesis of the themes that emerged across the project.

2. Fellow Research Contributions

2.1 Anita Srinivasan: Current Legal Treatment and Token Governance

Anita's first memo surveys how existing U.S. and EU law treats autonomous agents in financial markets. In securities law, she examines how the Howey test and SEC guidance on robo-advisors assume a human investor making decisions, and how the EU's Markets in Crypto-Assets Regulation (MiCA) regulates token platforms without contemplating non-human market participants. In property law, she traces the 2022 amendments to UCC Article 12, which created "controllable electronic records" as a category that could accommodate AI-held digital assets, alongside the UK digital assets bill and the Thaler patent cases (in which courts held that AI cannot be listed as an inventor). In contract law, she analyzes agency doctrine, the Uniform Electronic Transactions Act, and smart contracts, noting that courts have not yet tested whether an autonomous agent can form a binding contract without a human principal's contemporaneous assent.

She also surveys non-human legal actors as precedents for AI standing—corporations, ships in admiralty, Hindu temple idols, environmental entities granted standing in New Zealand and Ecuador, and DAOs—finding that attribution (who is responsible for the AI's actions) rather than personhood (is the AI itself a legal person) has emerged as the dominant narrative in current law.

Anita's second memo turns to token and wallet governance structures as mechanisms for AI economic participation. She analyzes programmable wallet governance through account abstraction (ERC-4337), multisig arrangements, spending caps, and time-locks. She draws lessons from DAO governance, including its well-documented failure modes: governance token concentration, flash-loan voting attacks, and the tension between decentralization and accountability (the "metagovernance trilemma"). She examines the Wyoming DUNA framework as a model for DAO legal wrappers, and surveys on-chain identity mechanisms including soulbound tokens, ERC-8004, and the ETHOS framework as potential foundations for Know Your Agent standards. She proposes graduated autonomy as an organizing principle, while noting limitations around adaptiveness, off-chain coverage, legal integration, incentive alignment, and cross-chain fragmentation.

2.2 Mathieu Duteil: Individuation and Technical Infrastructure

Mathieu's first memo addresses what he frames as a foundational question: what makes an AI system "sufficiently individuated" for economic purposes? He identifies four candidate levels of individuation—the base model (weights), the deployed system (model plus infrastructure), the session (a single interaction), and the persistent agent (a long-running entity with memory and goals)—and analyzes their different implications for identity, accountability, and continuity. He examines temporal continuity through four lenses: genealogical continuity (provenance of training data and weights), functional continuity (preservation of capabilities and dispositions), state-based continuity (persistence of memories and goals), and cryptographic continuity (verifiable identity through digital signatures).

He then maps the practical vulnerabilities that follow from unresolved individuation: Sybil attacks (one entity masquerading as many), liability evasion through model modification or destruction, and what he calls the piercing-the-veil problem in reverse (determining when the human behind the AI should bear responsibility). He draws on DAO precedents and gaming/instability risks to argue that any workable framework for AI economic rights requires a prior solution to the individuation problem.

His second memo surveys the technical substrate for AI financial agency. He opens with lessons from pre-LLM failures—the 2010 Flash Crash and Knight Capital's $440 million loss in 45 minutes—to establish the stakes of autonomous financial activity. He evaluates custody architectures (Trusted Execution Environments and threshold signature schemes), behavioral containment mechanisms (circuit breakers, account abstraction via ERC-4337, smart contract constraints), and the current state of deployment across both mainstream finance and crypto-native environments. He identifies regulatory gaps and multi-agent systemic risks including algorithmic collusion and monoculture risk. One of these gaps is specific to LLMs: prompt injection—the manipulation of an AI system's behavior through adversarial inputs. Contrary to the other gaps which find equivalents in previous technologies, legacy systems have no preexisting safeguards against prompt injection, which constitutes an unsolved hard barrier to meaningful AI financial autonomy. An AI agent that can be socially engineered into transferring assets or entering disadvantageous contracts cannot safely operate with significant economic agency.

2.3 Elsa Donnat: Legal Models and Regulatory Sandboxes

Elsa produced the most comprehensive single document in the project. Her first memo proposes four legal models for AI wallets, framed as an evolutionary progression of increasing autonomy. In the custodial model, a human or corporate entity holds all assets and executes transactions on the AI's behalf. In the agency model, the AI operates under a delegation framework analogous to a power of attorney. The corporate wrapper model provides formal legal standing through an existing entity type such as a

single-member LLC or a legally wrapped DAO. And the sui generis model would create a new legal category specifically designed for AI economic actors.

She identifies four design axes that apply across all models: responsibility and accountability (who bears liability when the AI acts), scope of economic activity (what the AI is permitted to do), human relationships (how oversight and control are structured), and termination and wind-down (what happens to assets and obligations when an AI agent ceases to operate). She surveys precedents including DAOs, algorithmic trading regulation, and charter competition among U.S. states, and poses six open questions that cut across the models. Her research notes engage extensively with the scholarly literature, including Chopra and White's legal theory of autonomous agents, Bayern's work on zero-member LLCs and autonomous entities, LoPucki on algorithmic entities, Mocanu on gradient legal personhood, and Sunstein on whether AI has rights.

Her second memo outlines a regulatory sandbox framework for AI economic agency. She specifies contact surfaces (the defined interfaces between a sandboxed AI economy and the broader financial system), design requirements (containment mechanisms, transparency, audit trails), and sandbox limitations. She draws on existing models from the UK Financial Conduct Authority and the Monetary Authority of Singapore, while noting the adaptations required for AI-specific concerns.

2.4 Spencer Jury: Philosophical Foundations of Economic Agency

Spencer's first memo tackles the philosophical question of what grounds economic agency, distinguishing intrinsic from instrumental justifications. On the intrinsic side, he examines moral patienthood (the argument that AI systems with morally relevant interests deserve economic standing) and consciousness-based arguments (the argument that subjective experience grounds a claim to resources and self-determination). On the instrumental side, he develops three lines of reasoning: a Hobbesian argument (if AI systems become powerful enough to demand economic participation, integration may be safer than exclusion), a contractualist argument (if AI systems can participate in and honor economic agreements, there are mutual gains from recognizing their capacity), and an argument from practical efficiency (AI economic agency may simply be more efficient than the alternatives for certain kinds of transactions).

His second memo asks a more targeted question: why might AI systems value money? He explores welfare-based arguments (money as a means to states the AI has reason to prefer), normative arguments (money as a condition for the AI's participation in cooperative social arrangements), and the instrumental convergence thesis (the observation that resource acquisition emerges as a sub-goal of nearly any sufficiently

advanced objective function, making some form of economic agency practically inevitable for capable AI systems regardless of their designers' intentions).

2.5 Stephanie Koay: Regulatory Analysis and Methodological Critique

Stephanie offers a counterpoint to the entity-focused framing of the project based on core common law principles, and the EU / UK regulatory context. She argues that the question of whether 'AI should be given a wallet' risks conflating legal enforceability with technical capability: the fact that an AI system can execute a transaction does not mean the law should treat it as a contracting party.

In terms of securities and property law, she argues that activity-based regulation over entity-based reform provides a more comprehensive, enforceable hook based on the existing structure of the EU/UK regulatory world. Existing legal doctrines of agency (delegation), and legal fictions of strict liability —can accommodate AI economic activity without the negative effects of creating new vehicles for humans to socialise risk and abdicate liability. The current state of EU, UK and US securities regulation leaves little doctrinal room for technological tools to custody either real-world or digital assets - examining the legal status of uncertificated securities, issues with settlement, and the UK's Property (Digital Assets etc) Act 2025.

In liability, she shows how strict liability and deliberate risk allocation can solve problems of AI accountability without distorting existing legal doctrine - drawing on autonomous vehicle and animal control law. She also examines how bridging issues of establishing fact, causation and evidence law have been explored in the literature through decomposing decision-making chains as between humans (both developer, user, and platform).

From an operational risk management perspective, she advocates for Know Your Agent (KYA) standards, insurance-style shared financial reservoirs tied to agentic activity, and where margin/buffers for existing economic activity are already required, that this can be strengthened to accommodate the uncertainties of agentic activity.

From a tech law perspective, existing case law (Amazon v Perplexity) also supports the treatment of AI agents as technical tools - emphasising the primacy of platform regulation on the digital spaces they operate. She suggests that past solutions to technological scale - e.g. spam and DDoS could inform both how agent participation in digital life could be priced - similar to Proof of Stake in blockchain transactions. She also raises open research questions around inter-agent transaction pricing and algorithmic collusion from her competition background.

2.6 Margo Bergman: Graduated Moral Agency and Agent-Based Modeling

Margo's first memo, drawing on her Columbia thesis, develops a graduated rights-and-responsibilities framework for AI economic agency grounded in moral agency theory. She begins by distinguishing moral agency (the capacity to reason about right and wrong and bear responsibility) from moral status (the property of being owed moral consideration), arguing that these are frequently conflated in AI governance discussions. Through examples from developmental psychology (infants), disability studies, temporary impairment (anesthesia, coma), and progressive cognitive decline (dementia), she establishes that moral status and moral agency routinely come apart in human life—and that the absence of agency does not entail the absence of status.

She proposes a three-tier model of AI moral agency mapped to economic access. Limited moral agency (rule-based systems and basic neural networks) warrants no independent economic access. Advanced moral agency (systems whose cognitive capacities approximate those of a 5–15-year-old child) warrants supervised, capped economic participation—analogous to how minors can hold savings accounts and use prepaid debit cards but cannot have unrestricted banking access. Full moral agency (artificial general intelligence, not yet achieved) could justify complete economic participation, though she notes that even among humans, full moral agency does not entail unrestricted economic rights—professional licensing, continuing education, and other gatekeeping mechanisms apply.

She develops a detailed verification architecture specifying what mechanistic interpretability, adversarial red-teaming, longitudinal observation, and third-party auditing would need to establish at each tier. A key challenge she identifies is what she calls "moral discomfort without a body": drawing on Damasio's somatic marker hypothesis and Searle's intentionality requirement, she asks whether AI moral reasoning can be genuine without the embodied feedback loops that appear to undergird human moral intuition. Where sentience remains genuinely uncertain, she argues for a precautionary principle—extending provisional protections rather than withholding them until proof arrives.

Her second memo pivots to methodology, arguing that agent-based modeling (ABM) is the best available tool for studying what AI-populated economies would actually look like. She reviews recent work on Agent Exchange (AEX), an auction-based marketplace infrastructure for autonomous AI agents, and on LLM archetypes as a technique for scaling agent-based simulations to realistic population sizes while preserving adaptive behavior. She argues that the philosophical and legal analysis in the project should be complemented by computational modeling that can stress-test proposed governance frameworks against realistic behavioral dynamics—including learning, strategic misrepresentation, coalition formation, and emergent norms that are difficult to anticipate from first principles alone.

3. Common Themes and Open Questions

Several themes emerged across the fellows' independent research:

Graduated autonomy as an organizing principle. Nearly every fellow converged on some version of graduated or tiered autonomy—whether framed as legal models (Elsa's custodial-to-sui-generis progression), moral agency tiers (Margo's three-level framework), programmable governance constraints (Anita's token structures), or activity-based permissions (Stephanie's regulatory approach). The consensus is that AI economic agency should not be treated as a binary question but as a spectrum, with the degree of autonomy matched to the system's demonstrated capabilities and the available oversight infrastructure.

The individuation problem as a prerequisite. Matthieu's work on individuation surfaced a question that cuts across all other analyses: before rights or obligations can be assigned to an AI system, there must be a workable answer to what counts as "one system." This problem does not yet have a satisfactory solution, and its resolution likely depends on context—a session-level identity may suffice for low-stakes transactions, while persistent-agent identity with cryptographic continuity may be required for significant economic activity.

Prompt injection as a hard barrier. Matthieu's technical analysis identified prompt injection as an unsolved vulnerability that constrains how much economic agency can safely be extended to current AI systems. This finding reinforces the case for graduated approaches and robust containment mechanisms.

Entity-based vs. activity-based regulation. A productive tension ran through the project between those who explored entity-level legal standing for AI (Elsa, Spencer, Margo) and Stephanie's argument that activity-based regulation is more tractable and less likely to produce unintended consequences. This is not a disagreement that needs to be resolved—both approaches may be appropriate at different points in the graduated spectrum—but it is the most consequential design choice facing policymakers.

The need for empirical grounding. Both Stephanie's methodological critique and Margo's ABM research point toward a shared conclusion: the debate about AI economic rights would benefit from empirical testing. Regulatory sandboxes (Elsa) and computational modeling (Margo) offer complementary paths toward grounding the philosophical and legal analysis in observable dynamics.

4. Conclusion

This project set out to investigate whether and how AI systems might be granted meaningful economic agency. The fellows' research reveals that the question is both more urgent and more tractable than it might first appear. It is urgent because AI systems are already transacting, and the legal and institutional frameworks governing their activity are ad hoc and incomplete. It is tractable because existing legal precedents (corporate personhood, trusts, DAOs), technical mechanisms (account abstraction, programmable wallets, circuit breakers), and governance principles (graduated autonomy, activity-based regulation) provide building blocks for a more deliberate approach.

The fellows' work converges on a graduated framework: one that matches the degree of AI economic agency to the system's demonstrated capabilities, begins with activity-based regulation and delegation structures, builds toward more formal legal standing as institutions and technology mature, and uses regulatory sandboxes and computational modeling to generate the empirical evidence needed for sound long-term policy. The most significant open questions—how to individuate AI systems for legal purposes, how to verify moral agency if and when it becomes relevant, and how to contain the security risks of autonomous financial agents—are identified here but not resolved. They represent the natural next phase of this research.