

INDUSTRY REPORT · AGENTIC AI × FINANCE

When your agent holds the *wallet*

AI fluency, stablecoins, and the new architecture of accountability in machine-speed finance. What the convergence of agentic commerce, regulated digital dollars, and explainable AI means for financial institutions.

SERIES

Responsible AI in Financial Services

FOCUS

Governance · Payments · Regulation

PUBLICATION

Clarium · 2026



AI agents are about to transact on your rails. *The control framework hasn't shipped yet.*

In under eighteen months, the industry built the plumbing for AI agents to hold money: Visa opened its network to agentic payments, Coinbase revived HTTP status 402 for machine-to-machine stablecoin settlement, Google standardized agent-driven checkout, and the United States enacted its first federal stablecoin framework. What has not been built at the same pace is the judgment layer: who delegates, who verifies, and who answers when an agent is wrong. This report argues that a human-capabilities framework, the 4Ds of AI fluency, maps with unusual precision onto the governance gap financial institutions now face.

2025	402	2	4
The year agentic-commerce standards converged: Visa Intelligent Commerce (Apr), GENIUS Act (Jul), Trusted Agent Protocol (Oct)	The dormant HTTP status code repurposed by x402 for native stablecoin payments between AI agents and services	AI use cases the EU AI Act designates high-risk for finance: credit scoring and life and health insurance pricing	Disciplines of AI fluency: Delegation, Description, Discernment, Diligence. A working governance grammar for agentic finance

- 01 **The transaction stack is being rebuilt around agents, not humans.** Visa Intelligent Commerce, Coinbase's x402, Google's Universal Commerce Protocol, and Visa's Trusted Agent Protocol each solve a different layer of the problem, yet all assume software, not people, initiates payment.
- 02 **Stablecoins give agents a currency that moves at their speed.** The GENIUS Act (P.L. 119-27) created the first federal reserve and licensing framework for payment stablecoins, clearing the way for machine-to-machine money, and for its risks.
- 03 **Regulators are converging on a simple demand:** where AI decisions touch people's financial lives, systems must be explainable, documented, and attributable to an accountable human. Agent autonomy does not dilute this; it sharpens it.
- 04 **The missing control layer is a harness, and it is human.** Anthropic's own agent research shows frontier models fail without one: task lists, written state, mandatory verification. The 4Ds, Delegation, Description, Discernment, Diligence, are the human version of that harness.



Data is where every agentic decision begins. Photo: Unsplash

Four protocols, one assumption: *the initiator of payment is no longer a person*

Between April and October 2025, the largest names in payments and cloud independently shipped interoperating standards for agentic commerce. Read bottom-up, they form a stack: settlement, credentials, intent, identity.

EXHIBIT 1 The agentic commerce stack assembled itself in six months

Identity & trust VISA TRUSTED AGENT PROTOCOL	With 10+ partners including Akamai: cryptographic proof of an agent's identity and authorization to merchants, the digital equivalent of showing ID at the till. Mainstream adoption targeted for 2026.	OCT 2025
Intent & checkout GOOGLE UCP	Universal Commerce Protocol: open standard for agent-driven shopping, checkout and payment, compatible with Agent2Agent (A2A) and Agent Payments Protocol (AP2).	SEP 2025
Credentials VISA INTELLIGENT COMMERCE	Visa opens its network to AI developers: agent-usable payment credentials and spend controls, later an MCP Server integration layer.	APR 2025
Settlement X402 · STABLECOINS	Open-source protocol reviving HTTP "402 Payment Required": an agent pays any server in stablecoin within a single HTTP request. No account, no card, no human.	MAY 2025

Source: Visa press releases (Apr 30, 2025; Oct 2025); Coinbase developer platform, x402.org; White House GENIUS Act fact sheet (Jul 18, 2025); Google Merchant documentation; CRS IN12553.

Each protocol answers a question merchants and banks have asked for decades, and each relocates trust. Visa Intelligent Commerce makes card credentials agent-usable under human-set spending rules. x402 removes the merchant relationship entirely: any HTTP endpoint becomes a paid resource. Google's UCP standardizes how an agent expresses purchase intent across retailers, while Trusted Agent Protocol lets an agent cryptographically prove who authorized it.

The direction of travel is unmistakable. When Visa's CTO frames agentic commerce as the next scale chapter of the network, and Mastercard ships its own Agent Pay in parallel, this is no longer a crypto-adjacent experiment. It is the core of the payment industry preparing for customers that never sleep, never mis-click, and never read the fine print.

SO WHAT

Financial institutions tend to evaluate these as discrete product integrations. The sharper question is strategic: if your customer's agent negotiates, selects, and pays programmatically, possibly in stablecoin and possibly off-card, what exactly is your brand, your margin, and your compliance perimeter attached to?

Stablecoins are the only currency that moves at agent speed, *and the GENIUS Act just made them institutional*

An AI agent does not carry a card. It holds keys. Stablecoins settle in seconds, run continuously, and cost fractions of a cent: the native money of machine-to-machine commerce. **Regulation has now caught up with the money, but not yet with the machine.**

“*The GENIUS Act regulated the **money**. It said nothing about the **machine** holding it.*”

Signed into law on July 18, 2025, the **GENIUS Act (P.L. 119-27)** establishes the first federal regulatory system for payment stablecoins: 1:1 high-quality liquid reserves, monthly public disclosures, issuer licensing regimes, and redemption rights. Law firms and congressional researchers alike read it as the moment stablecoins moved from gray zone to regulated payment instrument, deliberately framed as infrastructure for internet-native and agentic payments.

For financial institutions, this unlocks a concrete design space: treasury operations that run continuously, programmable settlement between counterparties, and micropayments small enough that humans would never price them but agents will. x402-style exchanges, where an agent pays \$0.001 for an API call or \$0.05 for a dataset, only clear economically when settlement is instant, near-free, and programmable.

EXHIBIT 2 One law, two halves: what the GENIUS Act covers, and leaves open

 Now regulated THE MONEY	 Still unassigned THE MACHINE
● Reserve quality: 1:1 backing in cash, short-term Treasuries and similar high-quality liquid assets ● Issuer licensing: federal or qualifying state regimes, with capital and examination requirements ● Transparency: monthly reserve disclosures and attestation obligations ● Redemption: holders' right to convert stablecoins to fiat at par	? Agent authority: who or what may legally authorize a stablecoin transfer on a principal's behalf ? Reversibility: final settlement versus fraud recovery when an agent errs in seconds ? Custody standards: for agent-held wallets and programmatic spend mandates ? Cross-border agents: operating across fragmented global stablecoin regimes

Source: GENIUS Act, P.L. 119-27 (Jul 18, 2025); Congressional Research Service IN12553; Mayer Brown and Arnold & Porter client analyses (2025). Right panel reflects Clarium analysis of open governance questions.

SO WHAT

The compliance frameworks banks trust today, authorization logs, velocity limits, dispute windows, were designed for human cadence: signatures, card swipes, batch settlement. **Agents collapse those timeframes.** The institutions that win will define delegated monetary authority, programmatically bounded, fully logged, and human-owned, before their first agent incident, not after.

The 4Ds of AI fluency are emerging as *the de facto control grammar* for agentic work

Anthropic and its academic partners, Rick Dakan (Ringling College) and Joseph Feller (University College Cork), codified AI fluency as four disciplines. What began as an educational framework reads, in a financial context, **like a control checklist regulators and boards will recognize.**

EXHIBIT 3 Each D maps to an existing financial-governance reflex

D Delegation MANDATE DESIGN <i>Deciding when to use AI and when to do the work yourself, mapping goals before asking for help.</i> In finance: which decisions may an agent take autonomously (reconciliations, data gathering) and which stay human (credit approval, client advice)? This is outsourcing governance applied to software.	D Description INSTRUCTION DISCIPLINE <i>Communicating clearly: detailed context, explicit rules, exact format and tone.</i> In finance: ambiguity in a prompt is a control deficiency. An agent told to "maximize yield" without constraints will find interpretations a compliance officer would never sign. Written mandates, hard limits, auditable instructions.
D Discernment OUTPUT VERIFICATION <i>Evaluating AI output: checking facts and logic instead of trusting blindly.</i> In finance: verification is already the job. Model validation, reconciliation, four-eyes review. An agent's confident hallucination enters the same pipeline as any model output: tested against source data before it touches a client or a ledger.	D Diligence PERSONAL ACCOUNTABILITY <i>Taking personal responsibility: watching for bias, protecting private data, staying honest about AI use.</i> In finance: GDPR-class privacy duties, fair-lending law, and disclosure duties do not transfer to the agent. The signatory owns the outcome, and increasingly must disclose where AI was used.

Framework: Dakan & Feller, AI Fluency Framework (aifluencyframework.org), developed with Anthropic, CC BY-NC-SA. Financial-services mapping and sub-labels: Clarium analysis.

The framework matters at institutional scale because the machine side of the story points the same way. Anthropic's engineering research on long-running agents shows a frontier model failing repeatedly without a harness around it: an explicit task list in the environment, a written record of progress, and mandatory end-to-end testing before any work is called complete. Given that harness, the same model ships. **The model is not the control system. The harness is.**

The 4Ds are the human version of that harness. Delegation writes the task list. Description writes the instructions. Discernment is the test suite. Diligence is the commit signature. Institutions that deploy agents without building this harness, human and technical, are running the experiment Anthropic already ran, and failed.

SO WHAT

Most institutions will train staff on AI fluency. The leaders will go further: they will embed the 4Ds into role design, putting delegation matrices in job descriptions, discernment protocols in sign-off chains, and diligence clauses in accountability statements, so that human judgment scales as fast as agent autonomy.

Regulators are drawing the line now: where AI touches a financial life, *the system must explain itself*

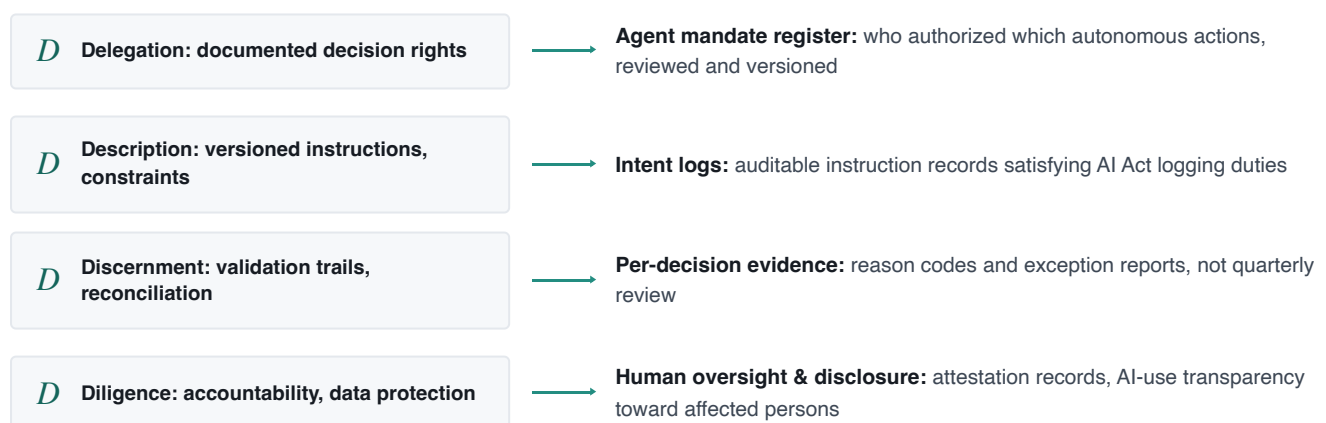
The EU AI Act places AI credit scoring and life-and-health insurance pricing in the high-risk category, triggering human oversight, logging, transparency and robustness obligations. For agentic finance, **explainability is no longer a model feature; it is a license condition.**

The **EU AI Act**, the most significant AI regulation enacted by any jurisdiction to date, classifies two financial use cases as high-risk: AI systems used to **evaluate creditworthiness**, and AI used for **risk assessment and pricing in life and health insurance**. Deployers face requirements for human oversight, automatic logging, data governance, and transparency toward affected persons.

Now extend that logic one step: an agentic system that pays, prices, or screens is not one model but a chain: a delegating human, an orchestrating agent, purchased data, a payment rail, settlement in stablecoin. Each link must carry its evidence forward. Trusted Agent Protocol shows the pattern regulators will generalize: cryptographic identity, declared intent, auditable trail. Expect the same grammar for agents' *decisions*, not merely their transactions.

This is where **explainable AI** becomes operational rather than aspirational: per-decision reason codes, retained feature-level evidence, reproducible runs, and a human who can reconstruct and defend what the agent did and why. What changes is granularity: from validating a model quarterly to instrumenting every autonomous decision continuously.

EXHIBIT 4 From the 4Ds to regulator-grade artifacts



Right column aligned to EU AI Act (Regulation (EU) 2024/1689) high-risk obligations: human oversight, automatic logging, transparency, data governance. Sources: European Commission "AI in finance" (2024); KPMG analysis for banks. Mapping: Clarium analysis.

SO WHAT

There is a tempting shortcut: keep agents in the low-risk bucket, or paper over autonomy with after-the-fact review. It will not survive contact with an audit, or with the first agent-caused loss that reaches a client. The durable position is the opposite: make every agentic decision **reconstructable by design**, so explainability is a property of the architecture, not a report generated under subpoena.

From framework to floor plan: *a supervised credit-screening agent* for a commercial lender

This is how Clarium thinks about AI in financial services, grounded in the work we already deliver: loan-origination platforms, core-banking and LOS integration, cloud migration, and data platforms for regulated lenders. The pattern: **agents where volume is high and stakes are structured; humans where judgment is the product; evidence everywhere.**

EXHIBIT 5 Control chain: agentic document intelligence in commercial credit

STEP 0 · HUMAN Credit officer, the signatory	D DILIGENCE · ACCOUNTABLE FROM DAY ONE The officer owns the credit decision before the agent is switched on. The agent drafts; the human decides, and signs. AI use is disclosed in the credit file.
STEP 1 · AGENT LAYER Screening & document intelligence	D DELEGATION · BOUNDED MANDATE D DESCRIPTION · VERSIONED INSTRUCTIONS Extracts financials from statements, tax forms and covenants; drafts the screening memo. No authority to approve, decline, or contact the borrower. Mandate, constraints and tone held in versioned instruction specs.
STEP 2 · VERIFICATION Reconciliation & exception queue	D DISCERNMENT · TRUST NOTHING UNVERIFIED Every extracted figure is reconciled against source systems before use. Discrepancies route to a human exception queue. The agent's confidence score is input, never verdict.
STEP 3 · SYSTEMS OF RECORD LOS · core · data platform	D DILIGENCE · EVIDENCE BY DESIGN Decisions write through the integration fabric to the loan-origination system and data platform, with per-decision reason codes and full input evidence retained for model-risk and EU AI Act-grade audit.

Reference architecture: Clarium. Delivery patterns based on Clarium engagements: portfolio management and asset-based lending platforms, PPP loan origination, middle-office portals for mortgage lending, MuleSoft integration of Abrigo LOS with Fiserv Premier core, and AWS cloud migration for financial services. All as reference design; not a depiction of any single client.

Data foundations Consolidated, quality-controlled lending data on Databricks and AWS. An agent is only as trustworthy as the data it reads.	Enterprise integration MuleSoft-pattern connectivity across LOS, core banking and third-party risk tools: the rails the agent reads from and writes to, governed at the API layer.	Security & monitoring Cybersecurity-first delivery with 24/7 security monitoring. Agent identities, secrets and tool permissions treated as first-class threat surface.	Governance by design Explainability instrumentation, decision logs and the 4D mandate register built into the architecture, not bolted on for the audit.
---	--	---	--

THE RESPONSIBLE-PARTNER PRINCIPLE

We don't sell autonomy. We sell **autonomy with a paper trail**: the model inside a harness of mandates, logs and verification, inspectable by your credit committee, your model-risk team and your regulator. That is the difference between an AI vendor and a responsible partner.

Five questions we cannot answer alone, *and don't think anyone can yet*

The pieces are all real: protocols, currency, fluency framework, regulation. What's uncharted is **how they compose**. These are the questions we put to clients first, and we don't pretend to have settled answers.

01 Who is your agent's "signatory"?

Law and custom assume a human behind every payment decision. When an agent settles in stablecoin at 3 a.m., **which person in your organization is legally and operationally accountable** for that instruction?

02 Does x402 economics break your business model, or fund it?

Micropayment-settled API access changes the price of data and services by orders of magnitude. If your competitors' agents can buy intelligence for fractions of a cent, **what does your firm sell to agents, and what does it refuse to sell them?**

03 How much "world model" does a financial agent actually have?

Research on world models asks whether AI builds predictive models of its environment or merely pattern-matches it. A trading or payments agent operating on pattern-matching, **in a market that adapts to it, is not a theoretical risk; it is a feedback loop.**

04 Is your model-risk framework sized for quarterly review, or ten thousand decisions a second?

EU AI Act-grade oversight assumes inspectable systems and competent human oversight. Continuous, per-decision explainability is an architecture decision. **Most current stacks were never built to provide it.**

05 What is your institution's 4D standard, and who signs it?

Delegation, Description, Discernment, Diligence can stay a training slide, or become the accountability spine of your agentic operations. **The difference is one signature at the top. Which will it be?**

Which of these five keeps you up at night?

Clarium works with financial institutions at exactly this intersection: AI architecture, data foundations, cybersecurity, and the governance to hold it together. Bring us your hardest answer.

Start the conversation · clarium.tech

Sources

- [1] Dakan, R. & Feller, J., "AI Fluency: Framework & Foundations," developed with Anthropic. aifluencyframework.org; Anthropic Academy (anthropic.skilljar.com). Framework released under CC BY-NC-SA 4.0. The four competencies, Delegation, Description, Discernment, Diligence, structure this report's Section 03.
- [2] Coinbase Developer Platform, "Introducing x402: a new standard for internet-native payments" (2025); x402.org. Open-source protocol using HTTP status 402 for native stablecoin payment between clients and servers.
- [3] Visa Inc., "Find and Buy with AI: Visa Unveils New Era of Commerce," press release, April 30, 2025 (Visa Intelligent Commerce, developed with AI industry partners); Visa Developer, "Trusted Agent Protocol" ([visa/trusted-agent-protocol](https://visa.com/trusted-agent-protocol), GitHub); Visa Investor Relations, "Visa and Partners Complete Secure AI Transactions," October 2025; Akamai and Visa press release on agentic commerce security.
- [4] Google, "New tech and tools for retailers to succeed in an agentic commerce era" (Universal Commerce Protocol launch, 2025); Google Developers, "Getting started with Universal Commerce Protocol"; Google Support, "About UCP and UCP-powered checkout," noting compatibility with Agent2Agent (A2A) and Agent Payments Protocol (AP2).
- [5] GENIUS Act, "Guiding and Establishing National Innovation for U.S. Stablecoins Act of 2025," Public Law 119-27, signed July 18, 2025; White House Fact Sheet (Jul 18, 2025); Congressional Research Service, "Stablecoin Legislation: An Overview of the GENIUS Act of 2025" (IN12553, updated 2026); Mayer Brown, "GENIUS Act Signed into Law" (2025); Arnold & Porter, "What You Need To Know About the New Stablecoin Legislation" (2025).
- [6] European Union, AI Act, Regulation (EU) 2024/1689, high-risk classification of AI credit scoring and life/health insurance pricing; European Commission, "AI in finance" (June 2024); KPMG, "Setting the ground rules: the EU AI Act"; Goodwin, "EU AI Act: Key Points for Financial Services Businesses" (2024); American Bankers Association, "Understanding the Impact of the EU AI Act."
- [7] Anthropic, "Effective harnesses for long-running agents," Anthropic Engineering (Nov 26, 2025), on incremental progress, structured state, and end-to-end verification across agent sessions; cited in Section 03.
- [8] On world models and agent decision-making: "Critiques of World Models," arXiv:2507.05169 (2025); LeCun, Y., JEPA and world-model architecture research (2024-2025). Referenced in Section 06, Question 03.

About this report. Prepared by Clarium as part of the Clarium Insights series on responsible AI in financial services. Clarium delivers AI/ML, cloud, data analytics, cybersecurity, and enterprise integration solutions for financial institutions and regulated industries. Clarium partners with AWS, Microsoft Azure, Google Cloud, Databricks, Salesforce, MuleSoft, and leading security providers. **Disclaimer.** This document is for general information only and does not constitute legal, regulatory, or investment advice. Statements reflect publicly available sources as of publication and may be subject to change as regulation evolves. · clarium.tech



The conversation continues. Photo: Unsplash