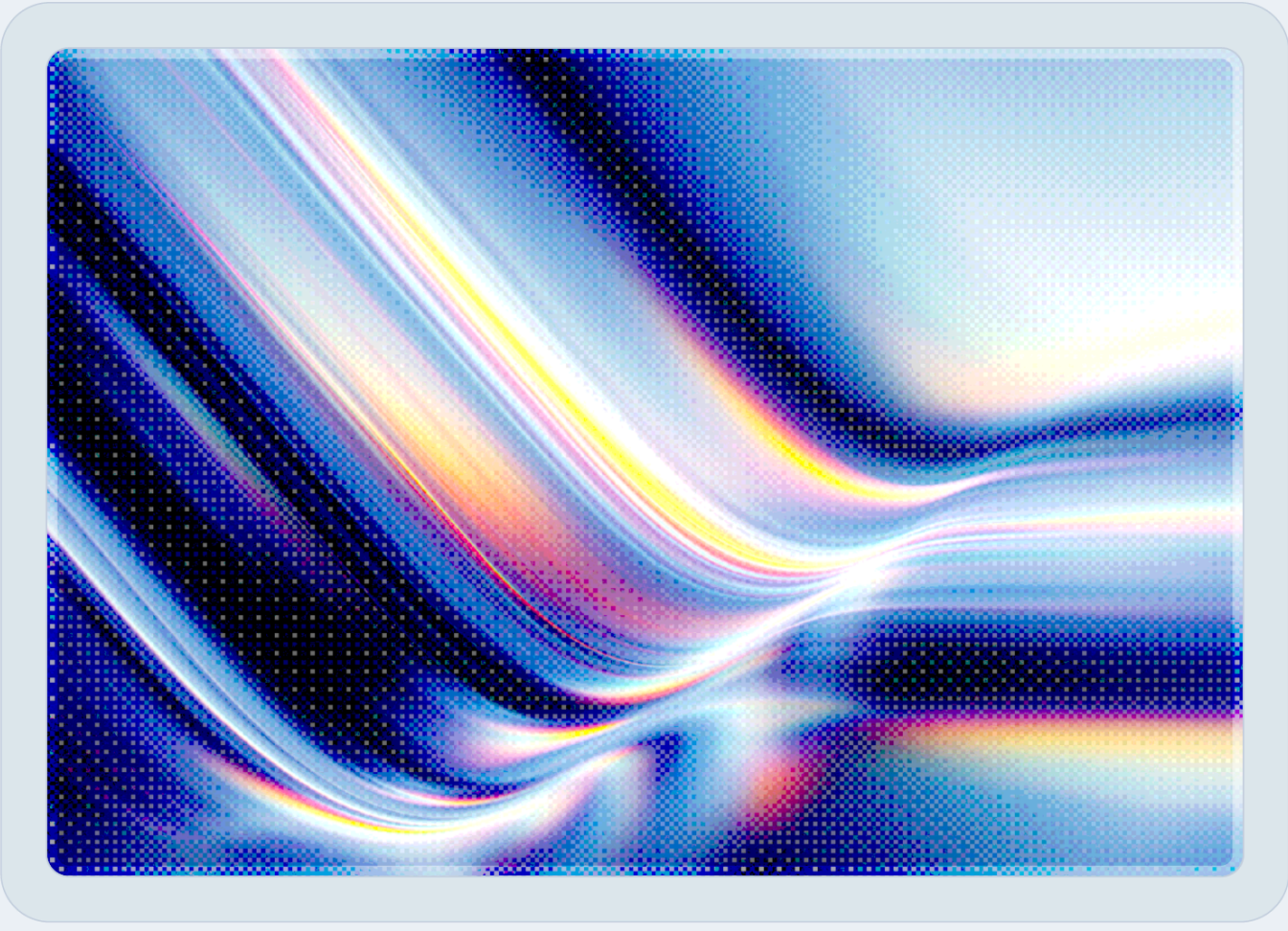
CSAI Foundation | Cloud Security Alliance

EU Digital Fairness Act: Agent Authorization Controls Become Consumer Law

Security Implications of Articles 26 and 27 for AI Agents That Conclude Contracts

2026-11-24

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

On November 11, 2026, the European Commission adopted its proposal for a Digital Fairness Act (DFA), a regulation that would replace the Unfair Commercial Practices Directive (2005/29/EC) and apply directly in all 27 Member States [1][2][3]. Initial coverage has emphasized the act's bans on dark patterns, subscription traps, and addictive design [3][8]. For security and identity teams, however, Articles 26 and 27 are likely the most directly relevant provisions, which, according to legal analyses of the proposal, would govern "automated systems" that can conclude contracts on a consumer's behalf without human intervention [2][4]. In practical terms, the proposal would turn several controls that the security community has recommended for AI agents (scoped delegation, user-set spending limits, step-up confirmation, and disclosure of conflicts) into legal obligations for any trader that makes such an agent available to EU consumers. Article-level details in this note are drawn primarily from secondary legal commentary and should be verified against the Commission's text [1].

The core requirement is a consumer-configured authorization policy. On first use, the consumer would have to choose expressly whether the agent must obtain confirmation for every contract, only for contracts meeting criteria the consumer sets (such as a maximum price or a time-limited mandate), or for no contracts at all. The choice could be changed at any time and would be confirmed on a durable medium [4][5]. Regardless of that setting, where the agent's provider is paid for proposing a contract or has a conflict of interest, the agent would have to disclose this and obtain the consumer's confirmation [2][4]. Traders would also be barred from making an agent the only way to obtain offers or make purchases [2][4].

Three gaps deserve immediate attention from security architects: two identified by legal commentators and a third that emerges from CSA's security analysis. First, according to commentary on the proposal, Articles 26 and 27 do not address who bears liability when an automated system errs [2], and Article 2(2) leaves national contract law untouched [4]. Second, one analysis observes that the consumer's spending limit binds the agent platform but is not transmitted to the counterparty merchant [4], so preventive enforcement of the limit rests on the agent provider's controls, with the consumer's withdrawal rights and any payment-layer restrictions serving as secondary or after-the-fact safeguards. Third, the proposal permits pre-contractual information to be delivered in machine-readable form when both sides of a contract are automated systems [4], which formalizes a channel through which counterparty-controlled, and potentially adversarial, content flows directly into an agent's decision process. The DFA remains a

proposal. It must pass through the European Parliament and the Council, and one law firm does not expect it to be in force before 2028 [3]. Organizations building or deploying commerce agents nonetheless have a clear signal of the control model regulators intend to require.

Background

From Consultation to Proposal

The DFA emerged from the Commission's "Digital Fitness Check" of EU consumer law, which examined whether the Unfair Commercial Practices Directive, the Consumer Rights Directive, and related instruments remain adequate for digital markets [1][6]. During the consultation phase, automated contracting by AI agents was treated as an open question; as late as October 2025, legal commentators noted that the Commission was not yet proposing specific measures on it [7]. The November 2026 proposal changed that by adding a dedicated set of rules for automated systems acting on behalf of consumers.

The proposal arrived as part of a three-part package. Alongside the DFA itself, the Commission proposed a revised Consumer Protection Cooperation (CPC) Regulation and a separate Simplification Directive amending the Consumer Rights Directive and the ePrivacy Directive [2]. (Some commentators instead describe the consumer-rights and enforcement changes as part of the DFA itself [3].) Under the revised enforcement framework, Member States would have to provide for maximum fines of at least 4% of annual turnover in the Member States concerned for certain widespread cross-border infringements [2] [3], and the Commission could impose fines of up to 6% of worldwide annual turnover in widespread cross-border cases it handles directly [2][3][8]. According to one analysis, the DFA and revised CPC Regulation would apply 12 months after entry into force [2]. Because the DFA is a regulation rather than a directive, its agent provisions would apply uniformly across the single market rather than through 27 national transpositions.

What the Agent Provisions Would Require

The DFA's agent rules sit within a broader chapter on digital contracting that also mandates online withdrawal and cancellation functions [2][3]. Table 1 summarizes the agent-related provisions as reported by legal analyses of the proposal. Article numbers and specific wording are drawn from secondary commentary and should be checked against the Commission's official text [1], which may change during the legislative process.

Provision (as reported)	Requirement	Security control analogue
Art. 3 definition	"Automated systems" are software able to take decisions with legal effects, including concluding contracts, without human intervention; per Recital 77, assistants still under user control are also covered [4]	Agent inventory and classification
Art. 26 confirmation settings	On first use, the consumer chooses confirmation for all, some (by criteria such as maximum price and duration), or no contracts; changeable at any time and confirmed on a durable medium [4][5]	Scoped, time-bound delegation policy with human-in-the-loop tiers
Recital 78 examples	Criteria may include excluding contracts above an amount or limiting purchases to chosen sellers or sellers offering withdrawal [4]	Allowlists and transaction thresholds
Art. 26(4) conflicts of interest	If the provider is paid for proposing a contract or has a conflict of interest, it must disclose and obtain confirmation regardless of settings [2][4]	Mandatory step-up authorization on policy-defined risk signals
Art. 26(5)–(6) identification	Optional identification through European digital identity wallets or the European Business Wallet [4]	Verifiable agent and principal identity
Art. 26(7) contact details	Consumer contact details passed to the seller [4]	Principal attribution to counterparty
Art. 26(9) no forced agent use	Traders may not make an agent the only route to offers or purchase [2][4]	Preserved human channel and fallback
Art. 26(10) scope exclusion	Financial services are excluded [4]	Separate regime (payment services law) applies
Art. 27 agent-to-agent contracts	Where both sides are automated systems, pre-contractual information may be machine-readable, with a plain-language confirmation on a durable medium still required [4]	Structured data validation and audit record

Table 1. Reported DFA provisions on automated systems and their security analogues.

The proposal is also explicit that delegation does not dilute consumer rights. According to one analysis of Recital 78, using an agent does not waive rights under the Consumer Rights Directive, including the right of withdrawal, even where the consumer has chosen to dispense with confirmations entirely [4]. Separately, Article 17(1)(d) would require features that simulate relationships, such as an AI companion's prompts to continue engaging, to be capable of being switched off [3].

The Industry Context

The DFA reaches agentic commerce at a moment when payment networks and AI platforms are building delegation infrastructure of their own. Visa and Mastercard each announced agent payment programs in April 2025 that let users set purchase limits and rules for what agents may buy [9][10], using tokenized credentials [10]. Google's Agent Payments Protocol (AP2), launched in September 2025 as an extension to the Agent2Agent protocol and the Model Context Protocol, expresses user authority as cryptographically signed "mandates" that record a user's intent and approval of a specific cart [11][12]. The FIDO Alliance has since reported that Google and Mastercard contributed AP2 and Mastercard's Verifiable Intent framework to the Alliance [13]. On November 13, 2026, OpenAI introduced ChatGPT Wallet, a branded payment wallet for its personal agent, a development CSA analyzed for its loss-allocation implications [14]. The DFA proposal would effectively set a regulatory floor for the contracting behavior of agents built on these schemes, while payment-layer aspects remain governed by payment services law, given the reported exclusion of financial services under Article 26(10) [4]. Whatever protocol a provider adopts, it would need to implement the consumer's chosen confirmation policy and the conflict-of-interest override.

Security Analysis

Authorization Policy Becomes a Compliance Artifact

The central shift for security teams is that the agent's authorization policy would no longer be a product design choice; it would be a regulated record. A provider would need to demonstrate that the consumer made an express choice at first use, that the choice was recorded on a durable medium, and that every subsequent contract either fell within the delegated scope or triggered a confirmation. That burden maps directly onto controls that CSA has described as weak in current deployments. CSA's research on agent identity sprawl described how enterprise agents frequently accumulate credentials and lack the

attribution needed to reconstruct who authorized what [15]. A consumer-facing agent with similar attribution gaps would struggle to prove that a disputed purchase was within the consumer's chosen parameters.

This suggests that the confirmation policy should be implemented as an enforceable, versioned policy object evaluated outside the language model, not as an instruction in the agent's prompt. A policy expressed only in natural language can be overridden by prompt injection [19], misread by the model, or silently ignored after a model update. A policy evaluated by deterministic code at the payment or contract-execution boundary, with each evaluation logged against the policy version in force, is more resistant to adversarial manipulation and easier to evidence to a regulator. Signed-mandate designs such as AP2 illustrate one way to bind a user's approval to specific transaction parameters [11][12], although a mandate scheme is only as strong as the integrity of what the user was shown when signing.

The Spending Limit Stops at the Platform Boundary

According to one analysis, the proposal passes the consumer's contact details to the seller under Article 26(7) but does not require the spending limit itself to be communicated, so a seller that accepts an over-limit order arguably "violates nothing" [4]. If that reading holds in the final text, the agent provider becomes the primary preventive enforcement point for the consumer's financial boundary. The consumer's withdrawal rights, which Recital 78 reportedly preserves even when confirmations are switched off [4], and payment-scheme dispute mechanisms would remain available, but these are after-the-fact remedies rather than preventive controls. Any failure in the provider's control, whether from a logic bug, a compromised agent session, or manipulation of the agent into splitting a purchase across multiple transactions below a per-contract cap, would fall outside the merchant's visibility. Security teams should therefore treat limits as aggregate as well as per-transaction constraints, and should consider whether payment-layer controls (such as network token restrictions of the kind Visa and Mastercard describe [9][10]) can provide a second, independent enforcement point.

Conflict-of-Interest Disclosure Can Also Serve as an Integrity Control

Article 26(4) is framed as a consumer-fairness measure, but it can also serve as an integrity control. It requires the agent to recognize when a recommendation is commercially influenced and to escalate to the consumer regardless of the delegation setting [2][4]. Recital 78, as reported, gives sponsored suggestions and preference for the products of the seller that supplies the assistant as examples [4]. Implementing this reliably requires the provider to maintain trustworthy metadata about commercial relationships and to propagate it into the agent's decision path. The same mechanism can also mitigate an adversarial pattern: a merchant or intermediary that injects content designed to steer an agent toward

its offer is creating a conflict the consumer has not consented to. A provider that can detect paid placement but cannot detect injected persuasion would satisfy the obligation as drafted while leaving a closely related integrity risk unaddressed.

Machine-Readable Contracting Opens an Injection Channel

Article 27 would allow sellers to deliver pre-contractual information in machine-readable form when both parties are automated systems [4]. The measure is intended to make automated contracting more efficient, but it also standardizes a channel through which counterparty-controlled content flows directly into the buying agent's context. Product descriptions, terms, and pricing data supplied by an adversarial or compromised seller agent could carry instructions intended to override the buyer's policy, misrepresent material terms, or exfiltrate information. The requirement for a plain-language confirmation on a durable medium [4] gives the consumer an after-the-fact record but does not prevent the agent from acting on manipulated data. Buyer-side agents should treat counterparty-supplied contract data as untrusted input, validate it against a schema, and compare material terms such as price, recurring charges, and withdrawal conditions against the consumer's policy in deterministic code before execution.

Liability Remains Unresolved

Covington's analysis notes that Articles 26 and 27 "do not address who bears liability when an automated system makes an error" [2], and another analysis observes that Article 2(2) leaves national contract law untouched [4]. This means an agent's mistaken purchase could be treated differently in each Member State, even though the authorization obligations themselves would be uniform. CSA's earlier analysis of AI liability argued that converging regulation tends to concentrate accountability on the deploying enterprise [16], and its analysis of agent payment wallets concluded that strong per-purchase confirmation can shift the cost of human-approved agent errors toward consumers and merchant returns processes [14]. Under the DFA, a provider that offers a "no confirmation" mode would likely face a heightened evidentiary burden in disputes, depending on how national law allocates the burden of proof, because the only record of consumer intent would be the original policy choice and the agent's own logs. The quality of those logs may effectively determine the provider's exposure in disputes.

Identity Wallets Are Optional

The proposal's references to EU Digital Identity Wallets and the European Business Wallet [4] point toward verifiable identity for both principals and counterparties, which would address a persistent weakness in agent-to-agent commerce: neither side can easily verify that the other agent acts for whom it

claims. Because these mechanisms are reported to be optional, however, providers may rely on weaker identification in practice. Organizations that adopt verifiable credentials early may find that they simplify both dispute resolution and the agent-to-agent trust problem that Article 27 makes more prominent.

Recommendations

Immediate Actions

Organizations that offer or plan to offer AI agents capable of completing purchases for EU consumers should first inventory those systems against the reported Article 3 definition, including assistants that remain under user control [4]. Because Article 3 reportedly defines automated systems as acting "without human intervention" while Recital 77 extends coverage to assistants under a person's control, scope assessments should err toward inclusion until the final text resolves this tension. Each in-scope agent should be assessed for whether its authorization policy is enforced in deterministic code at the transaction boundary or relies on model instructions. Where enforcement depends on the model, teams should prioritize moving price limits, seller allowlists, and confirmation triggers into a policy engine that the agent cannot modify. Legal and security teams should also obtain and review the Commission's official proposal text [1] rather than relying on secondary summaries, since article numbering and wording reported in commentary may differ from the official document.

Short-Term Mitigations

Because these controls are good practice independent of the DFA and take time to retrofit, providers should begin within the next two to three quarters by designing a first-use consent flow that captures the consumer's choice of confirmation mode, stores it as a versioned and durable record, and allows changes at any time. Around that record, several controls translate the proposal's reported requirements into implementable security measures. Per-transaction and aggregate spending limits, including time-bound mandates, should be enforced in a policy layer independent of the model. Commercially influenced recommendations should be tagged at the data layer, so that confirmation is forced whenever a paid or conflicted offer is selected, irrespective of the user's delegation mode.

On the counterparty side, machine-readable seller data received under Article 27 should be treated as untrusted input, subject to schema validation and deterministic comparison of material terms against policy. Every policy evaluation should be logged with the policy version, the inputs the agent relied on, and the confirmation shown to the consumer, and those logs should be retained to support withdrawal and dispute handling. Finally, providers should preserve a non-agent purchase path wherever an agent is

offered, consistent with the reported prohibition on forced agent use. These measures align with the control direction of signed-mandate protocols such as AP2 [11] and network-level agent programs [9] [10], and adopting such protocols may reduce implementation effort where they cover the required policy dimensions.

Strategic Considerations

The DFA proposal is the first major regulatory text to codify a model of agent authorization built on explicit, user-configured scope with mandatory escalation on defined risk signals, and it aligns with the direction of industry schemes such as AP2 and the card networks' agent programs [9][10][11]. Organizations operating globally may benefit from treating this model as a baseline across jurisdictions rather than building EU-specific variants. Security leaders should engage with the legislative process, particularly on whether consumer spending limits should be communicated to counterparties, a change that would create a second enforcement point and reduce reliance on a single platform control. Any such mechanism would need to be designed to avoid disclosing the consumer's willingness to pay, which could invite pricing up to the cap, expose additional personal data, and give seller agents a value to probe or manipulate; one option is for the agent to assert only that a transaction is within limits rather than revealing the limit itself. Finally, because liability will initially be governed by divergent national laws [2][4], organizations should weigh the evidentiary value of their agent logs as a risk-transfer asset and coordinate with insurers and payment partners on how agent-initiated errors will be allocated.

CSA Resource Alignment

CSA's recent analysis of agent payment wallets, *Agents at the Checkout: Loss Allocation in Frontier-Lab Payment Wallets* [14], is the most directly relevant prior work. That note examined how per-purchase confirmation and network tokenization address credential exposure while leaving the authorization and loss-allocation problems largely unresolved, and it observed that the most probative evidence of user intent now sits with the agent platform. The DFA's reported confirmation-policy and durable-medium requirements would formalize exactly that evidentiary role in EU law, making the platform's records central to any dispute.

AI Agent Identity Sprawl: The Enterprise Authorization Crisis [15] documents credential accumulation and attribution collapse in enterprise agent deployments. The same failure modes, applied to consumer agents, would make it difficult for a provider to demonstrate that each contract fell within the consumer's delegated scope. CSA's *Agentic AI Identity and Access Management: A New Approach* [17] proposes

identity and delegation models for agents based on verifiable credentials and fine-grained, context-aware authorization, which offer an architectural path toward the scoped, time-bound mandates the DFA describes and toward the optional EU wallet-based identification it contemplates.

AI Liability Inflection: Enterprise Accountability in the Agentic Era [16] anticipated that converging regulation would concentrate accountability on deploying enterprises. The DFA is a concrete instance of that trend for consumer-facing agents, even though it defers the final liability question to national law. At the control level, the AI Controls Matrix v1.1 [18] provides the mapping layer for these obligations, particularly its Identity and Access Management, Logging and Monitoring, and Governance, Risk and Compliance domains, through which organizations can align consent capture, policy enforcement, and audit logging with recognized controls.

References

[1] European Commission. "Digital Fairness Act Proposal." European Commission, November 2026.

[2] Covington & Burling. "The EU Digital Fairness Act Proposal: European Commission Proposes Major Update of EU Consumer Law." Inside Privacy, November 18, 2026.

[3] Paul, Weiss, Rifkind, Wharton & Garrison. "European Commission Proposes the Digital Fairness Act." Paul Weiss Client Memo, November 16, 2026.

[4] Pasquale Pillitteri. "Digital Fairness Act, Brussels Unveils Rules for AI Shopping Agents." Pillitteri, November 13, 2026.

[5] Pinsent Masons. "AI Agents and Dark Patterns Targeted by Digital Fairness Act." Out-Law, November 11, 2026.

[6] European Parliament. "Legislative Train: Digital Fairness Act." European Parliament Legislative Train Schedule, accessed November 2026.

[7] Osborne Clarke. "EU Digital Fairness Act Unpacked: Digital Contracts." Osborne Clarke Insights, October 16, 2025.

[8] Euronews. "EU Targets Subscription Traps, Addictive Design and Hidden Fees with New Digital Rules." Euronews, November 11, 2026.

[9] TechCrunch. "Visa and Mastercard Unveil AI-Powered Shopping." TechCrunch, April 30, 2025.

[10] Digital Commerce 360. "Visa and Mastercard Launch Agentic AI Commerce Programs." Digital Commerce 360, May 6, 2025.

[11] Google Cloud. "Announcing Agent Payments Protocol (AP2)." Google Cloud Blog, September 16, 2025.

[12] AP2 Project. "Agent Payments Protocol (AP2)." AP2 Protocol Documentation, accessed November 2026.

[13] FIDO Alliance. "Building the Trust Layer for Agentic Payments with AP2 and Verifiable Intent." FIDO Alliance Blog, May 26, 2026.

[14] Cloud Security Alliance. "Agents at the Checkout: Loss Allocation in Frontier-Lab Payment Wallets." CSA AI Safety Initiative, November 15, 2026.

[15] Cloud Security Alliance. "AI Agent Identity Sprawl: The Enterprise Authorization Crisis." CSA AI Safety Initiative, June 20, 2026.

[16] Cloud Security Alliance. "AI Liability Inflection: Enterprise Accountability in the Agentic Era." CSA AI Safety Initiative, June 27, 2026.

[17] Cloud Security Alliance. "Agentic AI Identity and Access Management: A New Approach." Cloud Security Alliance, 2025.

[18] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.

[19] OWASP. "LLM01:2025 Prompt Injection." OWASP Top 10 for LLM Applications, 2025.

© 2026 Cloud Security Alliance. AI-assisted Rapid Research.