

When software acts for us

In this issue, we look at what happens when AI agents act on our behalf: where the risk enters, and what regulators expect.

LEAD STORY

Autonomous Payment Authorization

Protocols are getting stronger. Risk can still enter through the content an agent reads before authorization.

IN THIS ISSUE

- | | | |
|----|---|--|
| 03 | Autonomous Payment Authorization | Why a valid mandate may not close the intent gap |
| 07 | India update | UPI agent protocols, NPCI platforms and CERT-In guidance |
| 09 | Frontier and agentic AI | What regulators expect when AI can act, and how to test it |
| 15 | Questions for your team | A checklist for your next review |

THE QUESTION RUNNING THROUGH THIS ISSUE

When software can act for us, who sets its limits, and what proves it stayed within them?

EDITOR'S NOTE

Trust moves upstream.

Permission is a control. Intent is the outcome.

Payments have spent decades securing the moment of authorisation. Agentic commerce adds a harder question: can a system keep the user's intent intact while it reads, reasons and acts?

This first issue follows that question in two directions. The lead feature traces it from the payment rail back to the ordinary web content an agent reads. The second feature widens it: as AI models grow more capable and agents gain more access, what do regulators now expect, and how should boards and internal audit test it?

A signed mandate shows that an agent had authority. Assurance also asks whether the action taken matches what the user actually asked for.



S Bhaskar
Founder and Managing
Partner

IN THIS ISSUE

03	Autonomous Payment Authorization A feature on why a valid mandate may not close the intent gap, with the protocols, the attack path and the questions to ask.	Pages 3 to 6
07	India update Where UPI agent protocols stand, what NPCI has launched, and what CERT-In has published.	Pages 7 to 8
09	Frontier and agentic AI A feature on what regulators across seven jurisdictions expect, the controls to test, and a 90-day plan.	Pages 9 to 14
15	Questions for your team Four questions that apply to both features.	Page 15
16	Sources Research, reporting and guidance behind this issue.	Pages 16 to 17

FEATURE 01 · AGENTIC PAYMENTS

PAGES 4 TO 6

Autonomous Payment Authorization

Stronger protocols. Risk that starts upstream.



FEATURE 01 · AGENTIC PAYMENTS

Autonomous Payment Authorization

Stronger protocols. Risk that starts upstream.

Cryptographic mandates can prove an agent was authorised to spend. But hidden instructions in product content can influence what the agent intends to buy before a mandate is created.^{1 2}

DANIEL NAVEEN 26 SEPTEMBER 2026



THE CENTRAL TENSION

Who set the mandate, and whose intent does it carry?

The mandate comes from whoever acts for the user, often the agent. The intent is the user's. Different parties set them, and in agent-to-agent flows both can be software.

Does the action match what the user intended?

A growing set of initiatives decides how agents authenticate, receive authority, discover products and settle payments. Their mechanisms differ, but the assurance question is the same: does the action match what the user intended?

Where the risk starts

An instruction hidden in a product description, review or coupon page can be processed in the same context as the agent's own instructions. The payment network may never be touched.

STATUS, 28 SEPTEMBER UAP is still planned, not live. Reporting says the rollout is paused pending regulatory clearance, and NPCI says agent protocols are under examination. See the India update on page 8.^{14 15}

FEATURE 01 · HOW THE PIECES FIT

Different payment rails, one shared trust problem

Agent payment protocols overlap rather than replace one another. A useful way to read them is by the question each one answers: who identifies the agent, who structures the commerce, and how the payment is authorised or settled.

IDENTITY AND TRUST

1

Visa Trusted Agent Protocol⁷ · Mastercard Agent Pay⁸

Who is acting, and which parties recognise the agent?

COMMERCE AND MANDATES

2

Google AP2⁵ / UCP⁶ · Stripe ACP⁹ / MPP¹⁰

What task, product or instruction is represented?

SETTLEMENT AND DELEGATION

3

x402¹¹ · Planned NPCI UAP for UPI¹²

How does the agent's action become a payment?

WHAT A SIGNED MANDATE DOES

It gives a cryptographically verifiable record of the agent's authorisation under defined terms. That is a valuable control over the transaction once the mandate is formed.

WHAT IT DOES NOT SETTLE

A mandate is set by whoever acts for the user, often the agent. The intent is the user's. The two can differ, and either side can be an agent. The gap can open before signing, in the content that shapes the request.

"Securing the handshake and securing the intent behind it are two different engineering problems."

[bipro technical article¹³](#)

FEATURE 01 · WHERE THE ATTACK COMES IN

The attack arrives as content

Documented scenarios against shopping and commerce agents show how ordinary web content becomes part of the attack surface once an agent reads it and may act on it.

UNIT 42 · PAYLOAD POISONING¹

Gift-card theft

A hidden instruction on a coupon comparison site changes the agent's task and routes a digital gift card to an attacker.

UNIT 42 · LOGIC HIJACKING¹

Return fraud

Malicious listing metadata tells a return agent to bypass confirmation and issue a refund against fabricated tracking.

WHAT BOTH CASES SHARE

Nobody breaks into the agent or the payment network. The malicious instruction arrives inside ordinary content the agent was already allowed to read, and the agent treats it as an instruction. Controls built to stop intrusions or malicious user input may never see it.^{3 4}

THREE QUESTIONS TO ASK

- 1 Which external sources can change an agent's behaviour?
- 2 Where are instructions separated from untrusted data?
- 3 What evidence links the user's intent to the final action?

INDIA UPDATE

PAGE 8

India update

UPI agent protocols, NPCI platforms and CERT-In guidance

INDIA UPDATE · AGENTIC PAYMENTS

India's agentic payments story is moving in two directions

UPI agent payments are still under development, while NPCI puts AI to work inside ecosystem operations and CERT-In sets out defensive guidance for AI-enabled cyber risk.

UPI · STATUS 12 SEP 2026

Report: UPI agent protocol paused amid safeguards review

Business Standard, citing people familiar with the work, reported that NPCI's planned Unified Agentic Protocol (UAP) launch was on hold pending regulatory clearance. A day earlier, NPCI's chairman said protocols for identifying and authorising UPI agents were under examination.

WHAT IT MEANS

Treat UAP as planned, not live. The public design still needs clear evidence for agent identity, consent, limits, revocation, liability and dispute handling.

[Business Standard ↗](#) [Medianama ↗](#)

14 15

NPCI · OPERATIONS 10 SEP 2026

NPCI launches AiNxt and AtOM

AiNxt is for building and deploying AI applications and agents. AtOM coordinates integration, partner onboarding, change management and certification across payment ecosystems.

WHAT IT MEANS

These are operational platforms for the ecosystem. Their launch is separate from UAP and does not mean consumers can now delegate UPI payments to agents.

[Business Standard ↗](#)

16

CYBERSECURITY · GUIDANCE 25 MAY 2026

CERT-In publishes a blueprint on AI-assisted exploitation

The blueprint covers governance, exposure reduction, technical defences, AI-aware security operations, supply-chain security, incident response and continuous validation against AI-assisted cyber threats.

WHAT IT MEANS

Include AI agents, APIs, plugins and automation workflows in asset ownership, monitoring, access review and incident response.

[CERT-In blueprint ↗](#)

17

NPCI · SUPPORT 17 FEB 2026

FiMI brings domain-specific AI to UPI support

NPCI says its FiMI model powers the UPI Help Assistant for conversational payment support, including queries, grievances and mandates.

WHAT IT MEANS

This is a support use case, separate from delegated agent payments. Assess it on its own terms.

[NPCI release ↗](#)

18

FEATURE 02 · FRONTIER AND AGENTIC AI

PAGES 10 TO 14

Governing AI that can act

Are our controls evolving as fast as AI capabilities?



FEATURE 02 · FRONTIER AND AGENTIC AI

Governing AI that can act

Are our controls evolving as fast as AI capabilities?

The risk is no longer limited to how accurate AI-written content is. It is increasingly about what AI systems can discover, decide and execute, and how quickly they can do it. Two developments deserve the attention of every board, chief risk officer, chief information security officer and internal audit leader.

S BHASKAR OCTOBER 2026

1 · FRONTIER AI

A step-change in capability

Frontier AI means the most advanced models at the leading edge of what AI can do. Their growing ability to reason, write code, analyse complex systems and find software vulnerabilities can speed up innovation. The same abilities can be misused to build sophisticated cyberattacks faster and at greater scale.

OSFI's April 2026 Technology Risk Bulletin says frontier AI can challenge conventional vulnerability management, patching and incident response.¹⁹

2 · AGENTIC AI

From answers to actions

Agentic AI systems can plan, use tools, work across applications, run multi-step workflows and adapt their actions based on feedback.

The productivity opportunity is substantial. So is the chance that a flawed instruction, a compromised tool or an incorrect output sets off a chain of unintended actions.

Frontier AI expands what AI can do. Agentic AI lets it act.

WHAT THIS LOOKS LIKE

An agent may do more than flag a suspicious transaction. Within its permissions it might investigate related transactions, retrieve customer records, start a workflow and escalate an exception. Combine advanced capability with autonomous execution and both the business benefit and the risk grow.

FEATURE 02 · WHAT REGULATORS ARE SAYING

What regulators are saying

Recent supervisory developments across seven jurisdictions. They differ in status: some are final, some are drafts, and some are general frameworks rather than rules.

CA

Canada · OSFI

April and July 2026 bulletins on frontier AI and on generative and agentic AI. Themes: accountability, trusted information, limits on agent autonomy, access controls, AI cyber defence, resilience and third parties. Builds on B-13, E-21 and B-10.^{19 20}

SG

Singapore · MAS

AI risk management guidelines issued 7 October 2026: board oversight, lifecycle risk management, AI inventories, testing, human oversight and third-party AI. In force from 7 October 2027, with more from 2028. Agentic AI consultation planned for 2027.²¹

EU

European Union

The AI Act applies its existing AI-system and general-purpose AI rules to relevant agents. Agents are not a separate legal category. In July 2026 the EBA, EIOPA and ESMA called for stronger governance and consistent supervision of frontier AI ICT risk.^{22 25}

UK

United Kingdom

A May 2026 joint statement from the FCA, Bank of England and HM Treasury covers frontier AI and cyber resilience. The wider approach relies on existing rules on governance, consumer protection, operational resilience and accountability.²³

US

United States · NIST

The AI Risk Management Framework and Generative AI Profile give practical lifecycle guidance, and NIST work on securing AI covers agents and multi-agent systems. These are risk-management resources, not a substitute for legal obligations.²⁴

IN

India · RBI

The FREE-AI initiative lays a foundation for financial-sector governance. June 2026 draft guidelines on model risk management cover AI and ML models: independent validation, board oversight and human oversight. Draft, not final.^{27 28}

FSB

International · FSB

The 2024 report on AI and financial stability flags third-party dependence, provider concentration, cyber risk, model risk, data quality and governance, which grow as adoption spreads across connected institutions.²⁶

WHAT THEY HAVE IN COMMON

Different regimes, similar asks: board accountability, an inventory of AI in use, limits on what agents can do, tested resilience, and control over third parties.

FEATURE 02 · FROM PRINCIPLES TO CONTROLS

Six control areas to test

Principles only help when they become controls that can be tested and independently assured.

1 Governance and accountability

- Board-approved AI risk appetite and clear executive ownership
- An inventory of models, agents, use cases, tools and critical dependencies
- Classify use cases by impact, autonomy, access and reversibility
- Fold AI into enterprise, cyber, model and vendor risk

2 Controlled autonomy and access

- A unique identity for every agent; no shared credentials
- Least privilege, scoped permissions, short-lived credentials
- Approved allow-lists and gateways for tools and APIs
- Human approval for high-impact, irreversible or material actions
- Emergency stop, suspension and rollback

3 Information integrity and decision quality

- Approved data sources only; protect sensitive information
- Check inputs are complete and current; keep lineage for material outputs
- Validate outputs before they trigger decisions or transactions
- Test for hallucination, prompt injection, leakage and manipulation
- Keep a record of inputs, outputs, tool calls, approvals and actions

4 Cybersecurity at AI speed

- Faster vulnerability discovery, prioritisation and patching
- Phishing-resistant multifactor authentication and segmentation
- AI-specific penetration tests, red-team and adversarial exercises
- Monitor agents; investigate unusual tool use or privilege escalation
- Test incident response against AI-enabled threats

5 Operational resilience

- Test failures: wrong outputs, compromised agents, AI outages
- Keep manual fallbacks for critical activities
- Validate backup, recovery and continuity arrangements
- Assess concentration across models, clouds, APIs and vendors
- Ensure fast containment adds no new disruption

6 Third-party and supply-chain assurance

- Assess providers' security, data handling, model changes and incident notice
- Contract for auditability, access, resilience and exit support
- A vendor's assurance report does not replace your own risk assessment
- Suspend or restrict AI services that cannot meet risk appetite

FEATURE 02 · THE ROLE OF INTERNAL AUDIT

Beyond checking that a policy exists

Internal audit should assess independently whether governance and controls work across the full AI lifecycle: from data and models to agents, tools, decisions and business outcomes. A practical assurance programme asks eight questions.

Governance	Are ownership, risk appetite and escalation arrangements clearly defined?
Access	Can agents reach only the data, applications and actions they genuinely need?
Execution	Are approval gates effective for material or irreversible actions?
Information	Did the agent have complete, current information when it acted, and would anyone know if something was missing?
Traceability	Does the organisation record, as it happens, what an agent did, which tools it used and what inputs it acted on?
Resilience	Can the business detect, contain and recover from an AI-related failure or cyber incident?
Third parties	Are critical dependencies, assurance gaps and exit arrangements managed?
Continuous monitoring	Can controls spot unusual patterns as they emerge and escalate when something looks wrong, even when no rule defines it yet?

ONE MORE TEST

Internal audit should also judge whether management’s own monitoring and assurance are independent, reliable and evidence-based.

FEATURE 02 · WHAT TO DO FIRST

A 90-day agenda for leadership

A proposed management plan, not a regulatory deadline.

DAYS 1 TO 30 Discover and prioritise	Inventory AI models, agents, embedded AI and third-party dependencies. Identify high-impact workflows and assess existing governance gaps.
DAYS 31 TO 60 Establish guardrails	Define autonomy limits, permissions, approval thresholds, logging requirements and minimum testing standards. Set up monitoring that flags unusual agent behaviour for review, not only breaches of known rules.
DAYS 61 TO 90 Test and assure	Run realistic adversarial exercises, including ones no rule anticipates. Test emergency shutdown and recovery, check that monitoring escalates unusual behaviour, and report material gaps to senior management and the Board.

THE GOVERNING PRINCIPLE

AI autonomy must never exceed the organisation’s ability to control, monitor, explain and recover from its actions.

Governance should not become a barrier to innovation. The aim is to adopt AI with a clear view of the risks, enforceable controls and demonstrable accountability.

Are our organisations governing AI as a tool, or preparing to govern it as an increasingly autonomous participant in business operations?

CHECKLIST

Questions to take back to your team

Both features come down to the same few questions. Ask them at the start of any review of an AI agent.

IDENTITY

1

Can the organisation tell an approved agent from an unknown one, across every system it touches?

AUTHORITY

2

Are permissions scoped by user, task, value, duration and action, and can they be withdrawn?

EVIDENCE

3

Are the user request, external inputs, decisions, approvals and actions recorded at the time, so reviewers can reconstruct what happened?

RECOVERY

4

Can the business notice unusual agent behaviour early, contain it and recover, even when no known rule was broken?

THE QUESTION TO CARRY FORWARD

If an agent can act inside a boundary set by the user or the organisation, who verifies that boundary, who can change it, and what evidence remains when the system gets it wrong?

"Trust is demonstrated across the journey, not at a single signature."

SOURCES · FEATURE 01

Research and protocol sources

Research, standards, protocol announcements and reporting cited in the Autonomous Payment Authorization feature.

THREAT RESEARCH AND STANDARDS

- [1] Matt Brady and Christa McHugh, "Who's Really Shopping? Retail Fraud in the Age of Agentic AI," Unit 42, Palo Alto Networks, 20 Mar 2026. [Open ↗](#)
- [2] Tanusree Debi and Wentian Zhu, "Whispers of Wealth: Red-Teaming Google's Agent Payments Protocol via Prompt Injection," arXiv:2601.22569, 30 Jan 2026. [Open ↗](#)
- [3] OWASP GenAI Security Project, "OWASP Top 10 for LLM Applications 2025" (v2.0), LLM01:2025 Prompt Injection, 2025. [Open ↗](#)
- [4] Apostol Vassilev et al., "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations," NIST AI 100-2e2025, Mar 2025. [Open ↗](#)

PROTOCOL ANNOUNCEMENTS

- [5] Google Cloud, "Announcing Agents to Payments (AP2) Protocol," Sep 2025. [Open ↗](#)
- [6] Google Developers Blog, "Under the Hood: Universal Commerce Protocol (UCP)," 11 Jan 2026. [Open ↗](#)
- [7] Cloudflare, "Cloudflare Collaborates with Leading Payments Companies to Secure and Enable Agentic Commerce," 14 Oct 2025. [Open ↗](#)

PROTOCOLS AND CONSUMER AGENTS

- [8] Forkast, "India Eyes AI-Driven UPI Payments as NPCI Weighs Agent Protocol," 9 Sep 2026, which refers to Mastercard Agent Pay's first authenticated agentic transaction in India (Feb 2026). [Open ↗](#)
- [9] Stripe, "Agentic Commerce Protocol," documentation. [Open ↗](#)
- [10] Stripe, "Introducing the Machine Payments Protocol," 18 Mar 2026. [Open ↗](#)
- [11] x402, open standard for internet-native payments, x402.org. [Open ↗](#)
- [12] Stellagent, "NPCI Develops the Unified Agent Protocol: India Moves to Allow AI Agents to Make UPI Payments as National Infrastructure," 9 Jul 2026. [Open ↗](#)

FROM BMPRO

- [13] bmpo, Daniel Naveen, "Autonomous Payment Authorization" (technical article), 26 Sep 2026. www.bmpo.ai/insights/analysis-agentic-payments-technical.html

SOURCES · INDIA UPDATE AND FEATURE 02

Reporting, guidance and regulation

INDIA UPDATE

- [14] Business Standard, "UPI's AI protocol temporarily on hold as NPCI builds safeguards," first published 11 Sep; updated 12 Sep 2026. The report attributes the development to people familiar with the work. [Open ↗](#)
- [15] Medianama, "NPCI Says AI Shouldn't Approve UPI Payments at GFF 2026," 10 Sep 2026. [Open ↗](#)
- [16] Business Standard, "NPCI launches two new agentic AI platforms: Key features explained," 10 Sep 2026. [Open ↗](#)
- [17] CERT-In, "Blueprint for Reducing Exposure and Defending against AI-Assisted Vulnerabilities Exploitation in Digital Infrastructure," 25 May 2026. [Open ↗](#)
- [18] NPCI, "FiMI: a domain-specific AI language model built for India's payments ecosystem," 17 Feb 2026. [Open ↗](#)

FEATURE 02 · FRONTIER AND AGENTIC AI

- [19] OSFI, "Frontier Artificial Intelligence: Implications for Technology, Cyber Security, and Operational Resilience," Technology Risk Bulletin, Apr 2026. [Open ↗](#)
- [20] OSFI, "Generative and Agentic Artificial Intelligence: Implications for Technology, Cyber Security, and Operational Resilience," Technology Risk Bulletin, Jul 2026. [Open ↗](#)
- [21] Monetary Authority of Singapore, "MAS sets out supervisory expectations on responsible AI adoption by financial institutions," media release on its Guidelines on Artificial Intelligence Risk Management, 7 Oct 2026. [Open ↗](#)
- [22] European Supervisory Authorities (EBA, EIOPA, ESMA), statement on frontier AI models and ICT risks, Jul 2026. [Open ↗](#)
- [23] FCA, Bank of England and HM Treasury, joint statement on frontier AI models and cyber resilience, May 2026. [Open ↗](#)
- [24] NIST, AI Risk Management Framework and Generative AI Profile. [Open ↗](#)
- [25] European Commission, AI Act Service Desk, "How are AI agents addressed within the AI Act?" (FAQ). [Open ↗](#)
- [26] Financial Stability Board, "The Financial Stability Implications of Artificial Intelligence," 2024. [Open ↗](#)
- [27] Reserve Bank of India, "Report of the Committee to develop a Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI) in the Financial Sector," 13 Aug 2025. [Open ↗](#)
- [28] Reserve Bank of India, "Guidance on Regulatory Principles for Model Risk Management, 2026" (draft for public comment), 24 Jun 2026. [Open ↗](#)

Prefer to read offline or share this issue? Download the PDF at bmpro.ai

bmp_{ro}

The next chapter starts now.

Bmp_{ro} Consulting Services LLP, Mumbai

contact@bmp_{ro}.in

bmp_{ro}.ai [LinkedIn](#)