



BRIEFING NOTE FOR SENIOR LEADERS

Know Your Agent

How public services and established businesses should prepare for the arrival of AI agents

Author: Tom Loosemore, Founding Partner, Public Digital
tom@public.digital

10 August 2026

Executive Summary

People are starting to send software, not themselves, to deal with your organisation.

If that sounds slightly odd, it won't for long.

An **AI agent** is software you give a job to in plain English. It gets on with it: finding, booking, buying or appealing on a user's behalf. Big tech and thousands of developers are racing to build these right now.

Agents are already researching, filling in forms and complaining. They're beginning to transact and pay autonomously, albeit at the margins. But there may not be one recognisable new front door. Some agents will arrive through signed machine channels. Others will use the customer's ordinary browser, app or device session and may be indistinguishable from the customer.

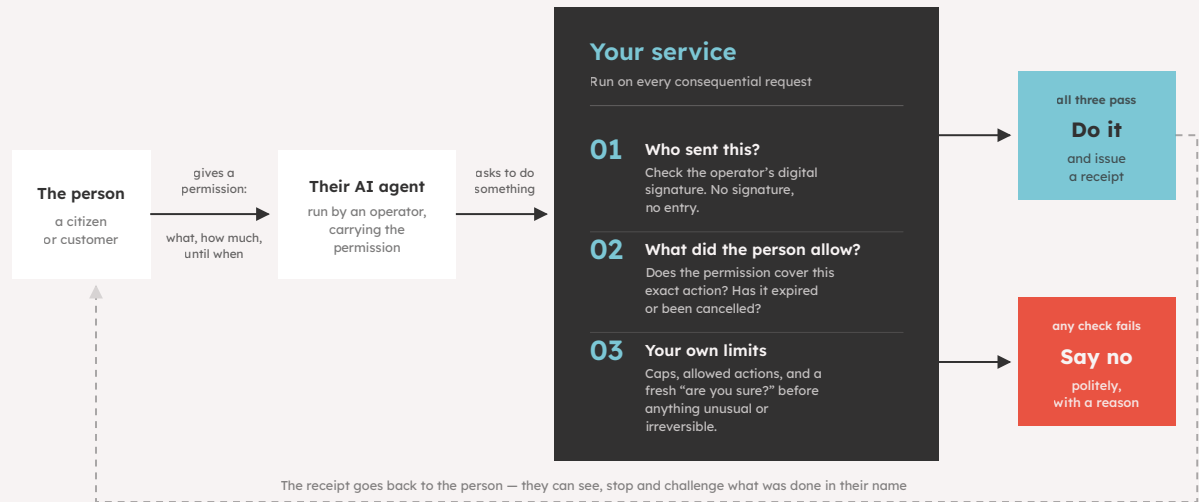
That creates two related problems. If software declares itself, can you verify who runs it and what the person allowed it to do? If it doesn't declare itself, how will you control consequential actions when you can't tell whether a human or an agent is at the controls? In either case, what evidence will you keep and how can the person challenge or undo the result?

I call the answer **Know Your Agent (KYA)**. It's not a detector, and it's not a single product you can buy. It's two things working together. When software arrives carrying evidence of who it is and what its owner allowed it to do, check that evidence and keep a record. And before anything that really matters, put limits on the action itself: a spending cap, a fresh "are you sure?" back to the person, a receipt, a way to undo. That second half is what keeps you safe when the software shows up carrying no evidence at all.

I'm mostly concerned here with agents acting for individual citizens and consumers, though much of it holds when one business sends an agent to deal with another. This is a strategic and technical briefing. It isn't legal advice, and I'm not a lawyer.

When an AI agent knocks on your service's door

The checks that make it safe to say yes



A well-behaved agent stays within its permission anyway. A hijacked one won't. Your checks protect you either way.

If you read nothing else

People may soon reach your services through an AI agent rather than your website or app. Shopping traffic arriving via AI services is growing fast. The payment networks are pouring money in. The analysts reckon agents will become a new front door to services. Nobody knows how quickly, me included. This feels like smartphones circa 2008: the direction is obvious, the volumes aren't. Prepare anyway.

Don't assume agents will arrive through a separate channel. Some cloud agents now sign their traffic. Others act inside the user's browser, phone or app session. A service may see an ordinary logged-in customer even when software is clicking. Detectability is therefore a design uncertainty, not a prerequisite you can bank on.

No single 'badge' will make an agent trustworthy, and many agents may present no badge at all. Where evidence is available, you need several checks: who runs the agent, whose authority it carries, what it may do, whether this action stays within those limits, what record is kept and how the person can challenge or undo the result. Where it isn't, consequential actions still need proportionate limits, fresh confirmation and redress.

Build on the identity systems you already use. The plumbing that lets people sign in and gives software limited access will probably remain the foundation. But treat a valid session as evidence of account control, not proof that a human performed the action. The genuinely new problem is how to turn a person's instruction into clear limits that another organisation can check and enforce, without demanding to know more about the person than it needs to.

Payments show that this can work, not that the problem is solved. Early payment schemes show how a signed instruction, a tightly restricted means of payment and a proper record might fit together. Benefits, health, housing and complaints are rather different. I don't yet know whether the same approach will travel.

AI-assisted demand pressure has begun, although autonomous-agent flooding hasn't. UK cases show AI adding volume or complexity, mostly through existing human, campaign and professional channels, and the same pattern is now being reported in other countries. Know Your Agent won't deal with all of this. Organisations need both delegated-authority controls and better ways to measure and handle AI-assisted submissions.



Part 1: Why prepare now?

1.1 Another way into your services

Think about how most online services work today. A person signs in with a password, an app or a government login. They read the questions, fill in the form and press the button themselves. The service assumes that the person using the account is also the person making the request.

There are exceptions. Solicitors, carers, accountants and other representatives already act for people. Some business systems also talk directly to one another. But these are specialist routes. The normal online journey still expects a human being at the keyboard.

AI agents break that assumption. A person can ask software to find a better energy tariff, challenge a parking fine, apply for SEN support for their child, pursue a housing repair, cancel a subscription or assemble a benefits application. The agent can cross several organisations, gather evidence, write emails and come back when it needs a decision.

Lawyers and technologists call the person or organisation giving the instruction the **principal**. The software is the **agent**. I use those terms where they help. They don't make the software a legal person, nor do they settle who pays when it goes wrong.

Nor is there a tidy line between an agent and the bots your organisation already deals with. A web scraper might be acting for one user. An agent might act for thousands. Looking at the traffic alone may not tell you much. What matters is the evidence travelling with each request:

- who runs the agent;
- which particular agent is asking;
- who asked it to act;
- what they allowed it to do;
- whether this request stays within those limits; and
- how you stop it, inspect what happened and put things right.

Today's bot defences generally can't answer those questions. They can spot suspicious or unusually busy traffic. They can't tell whether a legitimate customer asked the software to cancel a contract on their behalf.

The central uncertainty in this briefing isn't whether agents can use services. They already can. It's whether the service can know that an agent is at the controls.

Some agents run in cloud browsers and can identify themselves with a signed request. Others operate the user's own browser, call functions inside mobile apps or use the operating system's existing permissions. In those cases the external service may receive the same cookies, account identity, device signals and app traffic it would receive from the person. Behavioural bot detection may sometimes infer automation; it can't reliably establish delegated authority.

The rest of this briefing therefore considers two futures at once. In the cooperative future, agents declare themselves and KYA can govern the channel. In the messier future, many agents are practically indistinguishable from humans and KYA covers only the agents that volunteer (or are required) to use it. Your organisation should prepare for both eventualities.

1.2 What the evidence does (and doesn't) show

The case for preparing rests on five different flavours of evidence. Keeping them separate matters. Otherwise a forecast quietly turns into a fact and a pilot becomes "industry adoption" by the time it reaches the board paper.

Evidence	What I found
There's some observed consumer behaviour	Traffic reaching US shops after somebody followed a link from a generative-AI service grew 4,700% year on year in July 2025, from a tiny starting point. Most of this was people using AI to research shopping, not agents buying things by themselves. These visits were still less likely to end in a purchase than other visits. ¹
AI-derived demand pressure is real	AI-assisted material is adding volume or complexity in some UK services, and abroad: one Dutch municipality saw objections to property-tax valuations rise by 400% after a commercial AI platform launched to file them. The published cases concern people, professional representatives and campaigns using AI. They aren't autonomous consumer agents moving through services under delegated authority. ^{34, 35, 37}
There are early working approaches to KYA	Payment and technology companies have launched ways to identify agents, limit what they may do and keep records. Visa has reported hundreds of controlled, real-world transactions started by agents. ^{6, 8, 9, 10}
The analysts are excited about agents	Gartner reckons that by 2028 a third of people's interactions with digital services will start via an agent, rather than via the organisation's own app or website. A forecast, note. Not a measurement. ²
And plenty of it is still hypothesis	That autonomous consumer agents will strip enough effort out of applications, appeals, complaints and switching to cause a general surge. I think the mechanism is sound. I have no idea about the scale or the timing, and nor does anybody else.

The Competition and Markets Authority supplies a useful bucket of cold water. Its March 2026 research finds the technology still at an early stage, with most consumer-facing agents kept on a short leash and referred back to a human for anything that matters.⁴

So no, this isn't the web arriving while you weren't looking. But capable agents exist, the plumbing is being built, and big organisations take five years to change. The adoption curve won't wait for you. Wait for the volumes to surge before you act and you'll spend the following decade retrofitting.

1.3 AI-driven demand: the first evidence is here

Many services are hard work to use. AI can keep records, draft submissions and persist through several rounds of correspondence. That may widen access. It may also increase demand. **Agentic flooding** is the name given to the possible surge once AI removes much of the effort that previously put people off.³¹

The first evidence is now visible, but it needs reading carefully. Chris Schmitz's database lists 50 UK cases: four classed as confirmed, ten as likely, ten as suspected or confounded and 26 with no evidence of AI involvement.³⁴ These are useful warning signals, and worth watching. They're not 50 confirmed examples of autonomous agents, and that difference matters when the database gets quoted in board papers.

The Financial Ombudsman reports more AI-assisted complaints, including very long submissions from professional representatives.³⁵ The Bar Standards Board says AI is contributing to the volume and complexity of reports; NHS Employers and Bedford Borough Council describe similar pressures in recruitment and Freedom of Information requests.^{37, 39, 40} An AI-assisted Manchester campaign generated at least 1,588 objections to one planning application.^{34, 38} None of this demonstrates autonomous consumer agents acting across services from beginning to end. Nor should a rising caseload automatically be blamed on AI: benefit appeals, for example, have risen substantially for several possible reasons.^{29, 30}

None of this is only a British problem. The Economist reported in August 2026 that Schmitz has a forthcoming paper counting 93 examples across 11 countries, and that the EU Ombudsman, the Canadian privacy regulator and Australia's work tribunals all put a sharp rise in claims down to AI. Schmitz' suggestion is that Britain may be unusually exposed, because so many of our services accept open-text submissions and open text is exactly what AI is good at drafting. The same piece has the Information Commissioner's Office attributing more complex freedom-of-information requests to AI, and describes commercial services that will draft a planning objection or a parking-fine appeal for a fixed fee.⁴⁷

I'm quoted in that piece making the point I'd make here. An agent won't stop at telling you your home might be in the wrong council tax band. It will go on and try to get the band changed. I know the first half of that is easy, because I built it. OverBanded, a weekend experiment of mine, combines the Energy Performance Certificate API with council tax bands and past sale prices to guess which nearby homes might be banded wrongly. It took me a train journey. The second half, actually filing the challenges, is the

part an agent will happily do next, and the part your service will feel. That is the pattern to plan for: software that reads the small print, fills in the forms and, unlike a person, never gets bored enough to give up.⁴⁷

The distinction matters, so it's vital to be eyes-wide-open about what KYA does and doesn't get you. It kicks in whenever an AI agent declares itself, or whenever your service insists on an agent-aware route. It will do nothing about AI-written text pasted into an ordinary form. It may do nothing about an agent quietly driving a person's own browser session. So you need two things running side by side. KYA, for the agents you can attribute. And a second set of controls that remain in place whether or not you can attribute the request to an agent: structured intake, evidence requirements, duplicate and campaign detection, rate and value limits, a fresh confirmation before anything consequential, triage, and a real route to redress.

Leaders should measure what is happening and model what may happen next:

- What happens if the effort required to apply or appeal falls by, say, 25%, 75% or even almost zero?
- Which services currently rely, deliberately or otherwise, on users giving up?
- What demand is legitimate unmet need, which reflects ambiguity in policy and which is fraudulent?
- How are submission length, duplication, evidence quality and staff handling time changing, as well as headline volume?
- Can structured machine channels reduce the cost without (further) privileging people who can afford the best agents?

Not every form of friction is futile. Targeted charges appear to have changed the behaviour of professional representatives using the Financial Ombudsman Service.³⁶ Fees, rate limits, evidence requirements and additional checks may sometimes be justified, notably in a professional context. They should be lawful, transparent and proportionate, and shouldn't shut out people with valid claims. That's quite different from making a universal public service harder to use in the hope that ordinary people will give up.

1.4 Inclusion cuts both ways

Agents could become important assistive technology. Think of a carer who can't spend a morning on hold, a tenant with limited English, a disabled resident facing an inaccessible application or a parent trying to navigate several agencies at once.

They could also deepen inequality:

- capable agents may initially be concentrated among confident or affluent users;
- identity-heavy verification could exclude people without conventional documents;
- fee-charging verification services could favour large platforms over free or self-hosted agents; and
- blanket bot blocking could shut out legitimate assistive use.

GOV.UK's own work on digital identity acknowledges that not everyone can easily prove who they are online or use a digital route.²⁵ KYA should therefore ask for no more personal information than the service genuinely needs. It should also preserve a usable human route. "Computer says no" was never a good service design. Adding AI doesn't improve it.

Part 2: The checks you will need

2.1 The badge and the letter

Don't get hand-wavy about 'trust'. A signed or declared agent presents two distinct questions. An invisible browser- or operating-system-level agent may present neither; in that case, move straight to controls based on the risk of the action. (And don't assume the first question below is always answerable.)

Question one: who sent this request?

Suppose a piece of software arrives claiming to be a ChatGPT agent, or an agent run by a bank. Is it genuine, or is somebody merely using the name? A digital signature can help answer that. Think of it as a seal that shows which organisation sent the request and whether anybody has altered it on the way. It's the courier's ID badge.

Question two: what authority does it carry?

Did the customer or citizen ask the agent to do this? What exactly did they allow? How much could it spend, which information could it share and when should the permission expire? That's the letter carried by the courier.

The badge doesn't prove what is written in the letter. The letter doesn't prove that the courier is competent, safe or still following instructions. Both checks matter. Neither is enough on its own.

2.2 Six questions that services should ask AI agents

KYA is best understood as six questions, preceded by a gate: has the request arrived with agent evidence at all? Some of the technical names are included below because suppliers will use them. Senior leaders don't need to memorise them.

Check	The plain-English question	Some of the technical names you may hear
1. Where did this come from?	Is the request really from the company it claims to be from, and has it been altered?	HTTP Message Signatures, Web Bot Auth, TAP-style request signing
2. Which agent is it?	Which piece of software is acting, and which running copy is making this request now?	OAuth, OpenID Connect, SPIFFE/WIMSE, Microsoft Entra Agent ID, Google Agent Identity
3. What did the person permit?	Who asked the agent to act, what may it do, within which limits and until when?	Limited access tokens, action-bound approvals, AP2-style mandates
4. Did it stay within the limits?	Did this action match the permission, which limits did your service enforce and what record remains?	Rules enforced by the receiving service, caps, confirmation, receipts, logs and cancellation
5. Which other claims can you believe?	Who says the agent is capable, tested or safe, and why should you believe them?	Agent Cards, registries, verifiable credentials, independent assessments
6. What happens when it goes wrong?	Who is told, who can stop it, who investigates and how does the person challenge or undo the result?	Notices, records, complaints, dispute handling, contracts and insurance

These six questions cover different jobs. A directory can tell you that an agent exists; it can't prove that a citizen sent it. A digital identity can prove which software is calling; it can't tell you what the customer meant. A signed certificate can carry somebody's claim that the agent passed a test; it can't tell you whether that test was any good. And none of the six can help if the agent arrives inside an ordinary session without presenting evidence.

This is why buying one "trusted agent" badge would be such a mistake. It squashes six separate judgements into one comforting tick. Beware vendors flogging one.

2.3 Knowing its name doesn't make it safe

AI agents have a particularly awkward security problem called **prompt injection**. An instruction hidden in a web page, email or document can try to hijack the agent. It might tell a claims agent to ignore its original task, reveal private information or take some other action its user never intended.

However, the hijacked agent may still be using entirely genuine passes and signatures. It really is the customer's agent. Its permission may also be genuine, but too broad. Knowing its identity helps you investigate the damage. It doesn't prevent the damage.

That's why identifying the agent, deciding what it may do, enforcing limits and watching its behaviour must remain separate jobs. NIST's work reaches the same conclusion.¹²

For consequential actions, the receiving service should enforce hard limits ('runtime limits') on its own side:

- permissions that expire quickly;
- short lists of things the agent is allowed to do;
- limits on frequency, quantity or spending;
- a fresh "are you sure?" from the person before anything unusual or irreversible;
- a quick way to stop the agent or cancel its permission;
- warnings when its behaviour looks odd; and
- clear notification and a practical way to put things right.

These rules must be enforced by your service, not left to the customer's agent to remember. A name helps after the accident. Brakes help before it.

2.4 "Which agent?" has three answers

Technical teams will need to distinguish three things:

1. **The continuing agent.** The equivalent of its name and employment record, which may last through many updates.
2. **The copy running now.** A short-lived identity for this particular installation or session.
3. **What it is made from.** Claims about its current software, AI model, tools, settings and testing.

One proof of identity shouldn't be expected to prove all three. Updating an agent doesn't necessarily turn it into a wholly new agent. It may, however, make last month's safety assessment worthless.

Part 3: Where the technology is now

3.1 Recognising who sent the request

Web Bot Auth is the clearest current attempt to make agent traffic verifiable. The operator signs each request and the receiving website checks the signature against a published key. OpenAI now says ChatGPT Work's Cloud browser signs outbound requests as `https://chatgpt.com`. Google publishes a Google-Agent user agent for agents hosted on its infrastructure and is experimenting with Web Bot Auth as `https://agent.bot.goog`. Google says only a subset of those requests are currently signed.^{6, 7, 41, 42}

Real progress. But it is a good deal narrower than 'OpenAI and Google agents identify themselves', which is how it tends to get reported. It covers particular cloud-hosted traffic and no more. Google's earlier Project Mariner capabilities have since moved into products including cloud-hosted Gemini Spark and Chrome auto browse, which run differently. Google's own Web Bot Auth documentation stops well short of promising the signal survives on every Gemini surface, Android app actions included.^{42, 46}

Anthropic's public documentation says Claude in Chrome reads, clicks and types through the user's own browser. It doesn't promise that the website receives separate agent attribution. The same uncertainty applies to local computer-use harnesses, Gemini in Chrome and operating-system actions. Until vendors document the network path, describe these as potentially opaque rather than definitely detectable or undetectable. That might actually suit the big tech players, given the liability risks in all this.⁴³

Test verified agent traffic at your internet front door, but also test the opposite scenario. Ask what happens when a logged-in account performs the same sequence with no agent signal. A signature can show where a request came from; it can't prove user authority. An absence of a signature proves even less: it may be a human, an opaque agent or ordinary automation. Don't use unverified behavioural detection as the basis for denying rights or services.

3.2 Give the agent its own pass

Large organisations already have systems for deciding which people and which pieces of software may use a service. The strongest practical approach is to extend those systems to agents, giving each agent its own pass rather than letting it borrow the user's password.

- Microsoft's Entra Agent ID gives supported agents their own identities and administrative controls.¹⁵
- Google Cloud offers a similar approach for agents running in supported Google services.¹⁶
- The Model Context Protocol, widely used to connect agents to tools, uses the familiar OAuth method for granting limited access.¹⁷
- The OpenID Foundation is mapping agents onto the standards already used for online identity.¹⁴
- NIST, the US standards body, has launched a programme covering agent identity, security and the ability of different systems to work together.¹³

An industry-authored proposal submitted to the Internet Engineering Task Force makes the same case: adapt familiar identity standards rather than invent an entirely new system.¹¹ The proposal is useful evidence of where the industry is heading, but it isn't yet an agreed internet standard.

For channels you control, the foundation is likely to be evolutionary rather than revolutionary. Treat declared agents as identifiable pieces of software. Give them their own short-lived passes. Make this governed route easier and more reliable than pretending to be the person. But don't assume every agent will accept its own pass.

That still leaves two hard bits. An agent's pass can show which agent made the request, but not what the citizen or customer meant it to do. And an opaque agent may use the person's existing browser or app session without presenting a separate pass at all. For important actions, require action-bound confirmation or a structured agent credential regardless of how the request arrived.

3.3 Capturing what the person actually allowed

Google's Agent Payments Protocol, AP2, gives a concrete example. Suppose you tell an agent: "Find me size-nine running shoes for less than £70." AP2 can keep a signed record of that instruction, called an **Intent Mandate**. When the agent finds the shoes, a second

record, the **Cart Mandate**, ties your approval to the exact shoes, seller, price and terms.⁸

Mastercard's Agent Pay follows the same broad pattern. It gives the agent a tightly restricted substitute for the card rather than the card number itself. If there is a dispute, the existing card-network process provides somewhere to take it.⁹

This matters, but the shiny terminology can make it sound more settled than it is. A signed record shows that a particular digital key approved particular words. How convincing that record is still depends on:

- how the key was linked to the person;
- who controlled the phone, computer or digital wallet;
- what the person was actually shown before approving;
- whether the words were understandable;
- how those words were turned into rules a machine could enforce;
- whether the person later cancelled the permission; and
- whether the service did the thing that was approved, rather than something merely similar.

The useful general idea is **proof tied to the action**: evidence connecting what the person approved to the exact thing the organisation is being asked to do. "Permission mandate" is a handy label for some versions of this. It isn't yet a universally agreed term.

Researchers have proposed extending today's online sign-in and permission systems to cover agents.¹⁹ But caution is warranted. A July 2026 survey of 21 proposed permission systems and five commercial agents found none that managed all three of these things at once: easy for people to use, precise enough to remove ambiguity and reliably enforceable by computers.¹⁸ Narrow permissions can work. Turning a loose human instruction into safe, exact rules remains hard.

3.4 Payments: the first serious proving ground

Payments have something most other services lack: every action has a price, fraud costs somebody real money and there is already a well-established process for arguing about who pays. That makes payments the natural first proving ground.

Visa reported hundreds of controlled, real-world agent-initiated transactions by December 2025, alongside more than 100 participating partners and a smaller number in active integration.¹⁰ Google launched AP2 with more than 60 collaborators.⁸

Mastercard has extended Agent Pay across its ecosystem.⁹

Serious money, and proof the machinery works. It's a long way from mass adoption or settled consumer behaviour, and I wouldn't assume any of it transfers to public services. A card payment has a price, a seller, a payment method and a chargeback process. A housing assessment, a school-admission appeal or a care decision is a rather different beast. Some invention, market-shaping and standards-setting is going to be needed by those of us running public services.

The right lesson is: **learn from payments, but test that what's working for payments will work for your services.**

3.5 Digital certificates: an envelope, not a verdict

So-called **Verifiable Credentials** are signed digital certificates. They provide a standard envelope for carrying claims between organisations: that somebody is over 18, for example, or that an agent passed a particular assessment.²⁰ They may eventually carry an agent's permissions and safety claims, too.

Checking the signature tells you who issued the certificate and whether it has been altered. It doesn't prove that the claim is true, still relevant or made by an organisation worth trusting. A beautifully signed certificate from Dodgy Agents Ltd remains a certificate from Dodgy Agents Ltd.

These certificates can be designed to reveal only what is needed. A person might prove that they are eligible without revealing their name. But that privacy doesn't happen by magic. Poorly designed certificates can contain permanent identifiers that let organisations track the same person across services. Ask for the smallest claim you need, then test what the whole process leaks.

3.6 Directories and Agent Cards: signed phone books

A2A Agent Cards are machine-readable directory entries. They say who provides an agent, where to find it and what it claims to do. The cards can be digitally signed, so another system can check who published the entry and that it hasn't been altered.²¹

Useful, certainly. But it is still a phone-book entry, albeit a signed one. It says nothing about whether the agent is any good, whether its claims are accurate, or whether a customer ever asked it to act. MCP registries, which are directories of the tools agents can use, sit in exactly the same place.²²

Directories can support contracts, review and takedown. “Listed” must never be allowed to become shorthand for “trusted for every purpose”.

3.7 Reputation scores and proving who is behind the agent

Blockchain-based schemes such as ERC-8004 try to create public registers and portable reputation scores for agents.²³ An independent 2026 study of the early system found few working service addresses and a great many apparently fake or co-ordinated reviewers. It’s the digital equivalent of stuffing a marketplace with invented five-star reviews.²⁴ Take that as a warning about today’s implementations. It doesn’t mean every decentralised approach is doomed, but it does mean you shouldn’t take a reputation score at face value.

At the other extreme, some schemes want the full verified identity of the person behind every agent request. Banks and other regulated services will sometimes need that. Demanding it for every delegated errand would be expensive, exclusionary and needlessly intrusive.

Ask who the person is only when you genuinely need to know. For many lower-risk services, it may be enough to know that an organisation you trust has checked the person is eligible and that this particular action is permitted. Their name and passport number can stay out of it.

Part 4: What to ask for

4.1 Write down the house rules

Before you buy anything, set your house rules. Call it an **agent acceptance policy**. It's simply how your organisation decides to trust software acting for a person.

Keep the first version short. It should say:

- which jobs an agent may and may not do;
- what proof is needed as the consequences rise;
- which organisations you will trust to make which claims;
- what personal information you will collect and protect;
- whether an agent may pass work to another agent, and how its permission must narrow when it does;
- when permission runs out and how the person can cancel it;
- which limits your own service will enforce;
- when the person must approve the exact action;
- what record and receipt you will keep;
- how the person will be told what happened; and
- how complaints, disputes and recovery will work.

This doesn't require one supplier or a universal register. It does require decisions. Which evidence will you accept? From whom? How carefully will you check it? What happens to people who can't provide it? Those are service and policy choices, not details for your security team to wrangle with alone.

4.2 Ask for more proof when more is at stake

Looking up opening hours isn't the same as cancelling a tenancy. The checks should reflect the difference.

Level	Examples	A sensible starting point
Information only	Search public information; check opening hours; obtain guidance	Normal controls against excessive or abusive traffic. Knowing which agent is asking may help, but isn't essential.
Routine and reversible	Book an appointment; request a standard document; start a non-binding application	Identify the agent where possible; give it narrow, short-lived permission; tell the person what it did.
Consequential	Submit a claim or appeal; change an address; cancel a contract; disclose personal data	Prove which agent is acting; show approval for this kind of action; enforce limits in your service; issue a receipt; notify the person promptly; provide an accessible route to challenge it.
High-stakes or hard to reverse	Move significant money; surrender a right; alter someone's care, legal position or ownership	Check the person's identity where genuinely necessary; show them the exact action; seek a second confirmation; allow cooling-off or recovery where possible; keep stronger records and provide human review.

The policy also needs an answer for agents that can't produce your preferred proof. An unfamiliar agent isn't necessarily hostile. A familiar, correctly signed one isn't necessarily safe. Offer proportionate treatment and an accessible alternative route.

4.3 Keep the checks separate

For a consequential transaction, aim to answer eight questions:

- 1. Where did the request come from?** Check the sender's digital signature where the channel allows it.
- 2. Which agent is this?** Distinguish the continuing agent from the particular copy running now.
- 3. Who can answer for it?** Record an organisation or person you can contact and investigate. This doesn't, by itself, decide who is legally liable.
- 4. What did the person permit?** Look for permission that is short-lived, can be cancelled and is limited to a clear purpose. If an agent passes the task on to another agent, the permission should only get narrower.

5. **Does that permission cover this exact action?** Tie the approval to what your service is about to do, not merely to a vague instruction written earlier.
6. **Will your service enforce the limits?** Don't rely on the visiting agent to restrain itself. Keep a durable receipt of the result.
7. **Which other claims matter?** Treat claims about the software, its testing, its model, its tools and its governance separately. Avoid one catch-all "trusted agent" score.
8. **Can the person see, stop and challenge it?** Notice, access to records and a practical route to repair the outcome are part of the design.

No product on the market today does all eight reliably between different organisations. Treat this as a checklist to design against, not as a description of a mature market.

4.4 A signature is evidence, not magic

An electronic signature can be legally valid and can be used as evidence. What it proves depends on the circumstances, including whether the person meant to approve the action and whether the law requires any particular formality.²⁶

A signed instruction from an agent is therefore a better record, not an instruction the person "can't later deny". The person may say that their account was taken over, that they were pressured, that the screen was misleading or inaccessible, that they lacked capacity, that the signature covered a different action, or simply that the agent disobeyed them.

Safeguards should include:

- showing the permission clearly before the person approves it;
- providing a receipt they can understand;
- making continuing permissions easy to inspect and cancel;
- asking again before an unusual or irreversible action;
- setting fair rules for loss, liability and recovery; and
- never assuming that a valid digital signature proves informed consent.

4.5 The history of card payments might be instructive

Some readers will remember early "Verified by Visa" journeys: a payment suddenly disappeared into another screen and asked for another password. That was 3-D Secure in its less graceful youth. Later versions used more information about the transaction to

let many ordinary payments pass without interruption. European rules requiring stronger checks helped to push adoption.^{27, 28}

The useful lessons are about judging risk, aligning incentives, agreeing common rules and, sometimes, regulating. But 3-D Secure still checks individual transactions. It doesn't show that a person can safely give an agent a broad standing permission and disappear from view.

The narrower lesson is the better one: **let routine, limited actions happen quietly; bring the person back when an action is unusual, risky or difficult to undo.**

Part 5: What to do over the next 12 months

5.1 Ten actions for every large service organisation

1. **Find the doors agents can already use.** Map where software can read, submit, change, cancel, claim, appeal, complain, disclose data or move money. Include websites, system-to-system connections, email, contact centres and services run by others on your behalf.
2. **Write your agent house rules.** Start with two or three services where mistakes matter. Bring together service design, operations, security, identity, legal, data protection, accessibility and complaints teams. None owns this problem alone.
3. **See what your internet front door can recognise, and what it can't.** Ask the teams and suppliers that protect your websites what signed agent traffic they can verify today. Run a small trial with at least one known agent operator, then repeat the journey through an ordinary logged-in browser with no agent signal. Keep the existing checks for suspicious behaviour and fraud: a signature is an extra signal, not a replacement.
4. **Test a narrow permission on a real service.** Choose something useful but reversible. Test it first with an identifiable agent, then test how the service remains safe if the same action comes from an opaque agent inside the person's session. Make your service enforce limits on the action, not merely trust the caller's label. Tell the person what happened, let them cancel the permission and provide a receipt that can't quietly be altered.
5. **Check that different agents can take part.** Include agents from more than one large platform and at least one independent system run by the user or their organisation. Judge the evidence and the outcome, not possession of a preferred badge.
6. **Improve notice and redress now.** People should be able to see that an agent acted, what it did, which permission it used, how to stop it and how to challenge the result. None of this needs to wait for a new technical standard.

7. **Start measuring the channel and the workload.** You will never identify every agent, or every piece of AI-written text, with certainty. Distinguish traffic with a verified signature, traffic that merely declares itself, traffic that looks automated and traffic you can't classify. Report the unclassified share prominently: 'unknown' must not quietly become 'human'. Record professional representation, repeated or near-duplicate submissions, length, evidence quality, handling time, outcomes, fraud, blocking and complaints. Don't turn an unreliable AI detector into an official fact. Say how confident you are in the numbers.
8. **Model what happens if asking becomes cheap.** Start with the AI-assisted demand and complexity already visible in complaints, recruitment and information requests, then test more demanding scenarios involving autonomous agents. Pay particular attention to applications, appeals, claims, complaints and switching: places where effort currently suppresses demand. Put the findings in front of policy and product leaders, not just operations and technology teams.
9. **Offer a structured route where it makes sense.** A governed system-to-system connection (usually called an API) may be cheaper to run and easier to check than an agent reading pages built for people. Make the governed route safer and easier than human emulation; if KYA adds only friction, agents will remain opaque. Decide who gets access, at what speed and price, and preserve a good route for people who don't use agents.
10. **Set common expectations for standards and suppliers.** Work with peers. Ask suppliers for evidence that can move between platforms, logs you can export, permissions and keys you can cancel or replace, and testing across different agents. Ask browser and operating-system suppliers whether a verifiable attribution signal survives into external service requests. Avoid formats that tie your organisation to one agent platform.

5.2 Additional priorities for public bodies

Act together. Councils, NHS Trusts and government departments shouldn't invent hundreds of incompatible rulebooks for similar services. Ideally, they would convene themselves (or be convened) to agree common approaches, share what's being learned and deal with agent operators collectively. Confident, respected and technically capable convening institutions have never been more vital across myriad public sector organisations.

Design for the person who most needs help. Test the controls with people who have limited documents, connectivity, English, digital confidence or capacity. Keep assisted and representative routes as well as routes for autonomous software. The service must work for the person, not just for the agent that produces the neatest data.

Use government identity services for the job they actually do. GOV.UK One Login and the developing GOV.UK Wallet can help people prove who they are or something about themselves.³² The EU Digital Identity Wallet will also support identity checks, digital certificates and signatures.³³ None is currently a ready-made, general-purpose way to let an AI agent act for somebody. They may become useful components, particularly when a service needs the person to confirm a higher-risk action.

Make the policy choice visible. If easier applications or appeals reveal that a service has depended on people giving up, leaders have a choice: clarify eligibility, add capacity, reform the process or change the policy. Targeted charges, rate limits or evidence requirements may alter demand, as the Ombudsman's experience suggests. Their purpose, legal basis and effect on access should be explicit. Blocking agents, or simply making the service worse, is not an honest substitute.

Preserve fair decision-making. An agent can produce a neat, well-structured submission that is still wrong. More automated traffic must not quietly reduce the care given to each case, remove lawful discretion or make it harder for somebody without an agent to be heard.

5.3 Additional priorities for established businesses

Treat agents as a new customer channel, and as a new layer inside existing channels. Once the measurement is good enough, put verified-agent volumes, unclassified traffic, conversion, fraud, disputes and accessibility effects into ordinary product and board reporting.

Talk to the companies that handle your payments. Ask what they can support, in which markets, under what rules for losses and disputes, and whether it is live or merely on the roadmap. This may require changes to contracts, checkout, fraud controls, consent, data, receipts and dispute handling. It's unlikely to be just another switch in the admin screen.

Review terms and dispute handling. Existing terms may not say what happens when a customer uses an independently supplied agent, including what evidence the business will accept and how disputes will be handled. Contract terms can't remove statutory rights. Take specialist legal advice before relying on agent-specific wording.

Rework processes that depend on customers not bothering. Complaints, cancellations, refunds, data requests and claims may become easier to pursue. The defensible response is to handle valid requests efficiently and structure the process so fraud can be detected. Friction designed chiefly to make customers go away will become harder to hide, and harder to defend.

Part 6: Regulation and the questions nobody has settled

6.1 What current regulation does say

Article 50 of the EU AI Act matters in two distinct ways. Since 2 August 2026, Article 50(1) has required providers of AI systems intended to interact directly with ‘natural persons’ to design them so those people know they are dealing with AI, unless that is obvious. The European Commission’s guidelines give this an agent-specific reading: where an agent is capable of dealing with a person, it should disclose its artificial nature and on whose behalf it acts, including in multi-agent chains.⁵

This could matter when a consumer’s agent emails, chats or negotiates directly with a staff member. It doesn’t amount to a machine-readable passport at the website’s edge. The Commission says background and machine-to-machine exchanges, and systems that merely collect data rather than conduct a genuine two-way exchange, fall outside Article 50(1). An agent silently completing a web form or invoking an app may therefore remain invisible to the receiving organisation.⁵

Article 50(2) is different again. It requires providers of generative systems to mark synthetic text and other content in a machine-readable way, subject to exceptions and technical feasibility. That mark concerns the provenance of content, not the identity of the agent, its principal or its authority to act. It may help identify AI-generated material; it isn’t KYA.⁵

Don’t conflate this with the Digital Markets Act (DMA). Apple says it is delaying Siri AI on iPhone and iPad in the European Union because of DMA requirements concerning access for competing assistants. In July 2026, the European Commission adopted measures requiring Google to give competing AI services effective access to 11 Android features. These disputes concern platform competition, interoperability, privacy and security. They aren’t about whether an external council, bank or retailer can identify an agent. They matter indirectly because they may determine which assistants can act deeply inside a device, and therefore how much opaque agent traffic emerges.^{44, 45}

The Competition and Markets Authority (CMA) makes one narrower point. Its March 2026 guidance is aimed at a business using its own AI agent to deal with customers, and says the business remains responsible for that agent, much as it would for an employee or outsourced provider.³ That doesn't answer the central question here: what proof an organisation may require from a consumer's independently supplied agent, or who bears the loss if that agent exceeds its instructions. The CMA's wider research treats those questions as unresolved.⁴

6.2 What remains unresolved

1. Will regulation or platform rules require a verifiable agent signal to survive from a browser, operating system or app action to the receiving service, including when the agent uses the person's authenticated session?
2. Who bears the loss when a consumer's agent exceeds its instruction: the person, the agent operator, the platform, the business receiving the request, a payment provider or somebody else?
3. What must the person be shown for their approval to be informed and accessible?
4. When one agent passes work to another, how do we make sure the permission gets narrower rather than wider?
5. How can a service turn an instruction written in ordinary language into exact limits its computers can enforce?
6. How does a cancellation reach every organisation and agent in the chain?
7. When is it enough to prove that somebody is eligible without revealing who they are?
8. How should public bodies apply duties of fairness, discretion and review to submissions written by agents?
9. In a dispute, which evidence should count strongly, and which evidence should never take away the person's right to challenge the decision?

The first serious court cases and regulatory decisions may change the answer quickly.



Where I might be wrong

1. **Autonomous consumer agents may remain a niche.** They may be too unreliable, costly or legally awkward for many people to trust with important tasks. AI-assisted submissions may still grow through ordinary forms, campaigns and professional representatives. That would weaken the demand case for KYA, but not the wider operational case for preparing services.
2. **Broad instructions may be too hard to turn into safe limits.** If a system can't translate what a person meant into precise rules, proof tied to the action may work only for simple menus and payment-like transactions.
3. **Payments may be the exception.** Commerce has strong incentives, common infrastructure and established ways to handle disputes. Welfare, health, justice and employment are rather less like buying a pair of trainers than technology demonstrations sometimes suggest.
4. **A few large platforms may set the rules first.** Proprietary approaches to trust and permission could become established before systems that work across platforms are ready.
5. **A major failure may produce identity-heavy regulation.** Serious fraud or harm could push governments towards requiring stronger personal identification everywhere, despite the costs to privacy and inclusion.
6. **Long chains of agents may remain impossible to inspect.** When agents hand work to one another, change the AI models they use and select new tools as they go, a reliable record from beginning to end may be impractical.
7. **Other parts of the world may take a different path.** The approaches reviewed here are shaped largely by US technology companies and European regulation. China will probably take a different path from either, which might prove globally influential given their AI capabilities.

8. **OS- and browser-level agents might remain opaque.** This would make universal KYA impossible, but does not make this briefing moot. It would split the world into a governed agent channel and an unobservable layer inside ordinary customer traffic. Organisations would need to shift more protection to action-level limits and user confirmation, while regulators decide whether platform-level attribution should be mandatory.

Fine. I might be wrong about all of them. Those are all reasons to test, to measure and to keep your options open. Not one of them is a reason to wait.



So, where does this leave you?

Today's digital services know how to recognise an account signing in and an approved piece of software calling an interface. They often don't know who, or what, operated that account. That distinction becomes critical when AI can use the same browser, app and operating-system pathways as the person.

The early UK evidence is a warning, though a sharper and narrower one than "agentic flooding" suggests. AI-assisted material is already adding work through the channels you already have. Autonomous consumer agents aren't yet running across services at scale. KYA deals with the delegated-authority problem and nothing else, so you will still need the ordinary operational controls for demand, fair access and redress.

There's no single badge that fixes this, and a badge won't always be presented. Where an agent provides evidence, a service needs to do several jobs:

- check where the request came from;
- recognise the agent separately from the person;
- ask for proof of what the person allowed;
- tie that proof to important actions;
- enforce the limits in its own systems;
- keep a receipt;
- collect no more personal information than the job requires; and
- make it easy for the person to see, stop and challenge what happened.

So design KYA as an explicit safe channel. Make the governed route easier and more reliable than pretending to be the person, because if KYA is nothing but extra friction, agents will simply stay opaque. For the ones that do stay opaque, make consequential actions safe anyway: limits, a fresh confirmation, notification, receipts and redress that protect the account holder whoever was doing the clicking.

You don't need to pick the winning standard in 2026. You do need three pieces of work running: a governed route for identifiable agents carrying delegated authority; consequential actions made safe even when the actor is opaque; and a better grip on how you measure and handle AI-assisted demand through every channel you own. Press

your operating-system, browser and agent suppliers for verifiable attribution, but don't wait for it. KYA isn't moot if agents can hide. It just can't be the whole answer. Waiting for the market to hand you a finished one is not a strategy.

Author's note on sources and method

To research this topic, I used three highly-capable LLMs from different vendors using a bespoke deep research skill that runs a research task through 5 personas (The practitioner, the academic, the sceptic, the economist and the historian). The skill then gets the LLM to do synthesis, peer review, confidence scoring and draft a report with full citations, before getting them to challenge their own work. As a final step, I got the most powerful current LLM to act as the judge across the three other LLMs' deep research reports. This LLM stage of research took a couple of hours. I then spent several days making sure I understood all this research and any LLM-derived conclusions from first principles, before deciding what I actually think.

This briefing draws on standards documents, official guidance, supplier announcements, academic work and selected secondary analysis reviewed between 29 July and 9 August 2026. Primary and official sources have been prioritised. Supplier announcements are treated as evidence of activity and stated capability, not independent proof of adoption or effectiveness. Forecasts, preprints, scenarios and hypotheses are labelled thus. The Agentic Flooding database has been used to find cases, not as independent proof of their cause. Its stronger UK cases were checked against the cited underlying sources where those were available.

Tom Loosemore, Founding Partner, Public Digital
tom@public.digital



About the author

Tom Loosemore is a founding partner of Public Digital. In 2011 he co-founded the UK's Government Digital Service, where he led the creation of GOV.UK and wrote the UK's first Government Digital Strategy. Before that he devised the BBC's internet strategy.

Most of his time now goes on AI, and on what happens to public services when citizens have capable agents and institutions don't. He writes about it weekly and builds with it constantly, which is where most of this briefing comes from.

He blogs at www.loosemore.com. You can reach him at tom@public.digital.

About Public Digital

Public Digital helps governments, public bodies and large organisations adapt to the agentic era. We work with leaders on AI strategy, operating models, service design, data, cyber security, capability building and de-risking major programmes, and we have worked across the UK, Europe, North and Latin America and Africa with clients including the Cabinet Office, the Home Office, Defra, UK Parliament, the NHS, BT Group and the governments of Nova Scotia, Madagascar and Edo State. We are a member of the kyu Collective.

Public Digital, 262 High Holborn, London WC1V 7EE. public.digital. General enquiries: contact@public.digital.

References

1. Adobe, 21 August 2025, *Generative AI-powered shopping rises with traffic to US retail sites* (4,700% year-on-year growth in generative-AI referral traffic; conversion remained below non-AI traffic). <https://business.adobe.com/blog/generative-ai-powered-shopping-rises-with-traffic-to-retail-sites>
2. Gartner, 26 August 2025, *Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026* (including forecast of a shift from native applications to agentic front ends). <https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025>
3. UK Competition and Markets Authority, 9 March 2026, *Complying with consumer law when using AI agents*. <https://www.gov.uk/government/publications/complying-with-consumer-law-when-using-ai-agents/complying-with-consumer-law-when-using-ai-agents>
4. UK Competition and Markets Authority, 9 March 2026, *Agentic AI and consumers*. <https://www.gov.uk/government/publications/agentic-ai-and-consumers/agentic-ai-and-consumers>
5. European Commission, July 2026, *Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of the AI Act*. https://ai-act-service-desk.ec.europa.eu/sites/default/files/2026-07/guidelines_on_the_implementation_of_the_transparency_obligations_for_certain_ai_systems_under_article_50_of_the_ai_act_bzptwqhk0ikg1dtlddap41psfy_131215.pdf
6. Cloudflare, 15 May and 1 July 2025, *Web Bot Auth* and cryptographic verification in the Verified Bots programme. <https://blog.cloudflare.com/web-bot-auth/> and <https://blog.cloudflare.com/verified-bots-with-cryptography/>
7. IETF, Web Bot Auth Working Group. <https://datatracker.ietf.org/wg/webbotauth/>
8. Google Cloud, 16 September 2025, *Announcing Agent Payments Protocol (AP2)*; AP2 Agent Authorization Framework. <https://cloud.google.com/blog/products/ai-machine-learning/announcing-agents-to-payments-ap2-protocol> and https://ap2-protocol.org/ap2/agent_authorization/
9. Mastercard, 29 April 2025, *Mastercard unveils Agent Pay*. <https://www.mastercard.com/us/en/news-and-trends/press/2025/april/mastercard-unveils-agent-pay-pioneering-agentic-payments-technology-to-power-commerce-in-the-age-of-ai.html>
10. Visa, 18 December 2025, *Visa and partners complete secure AI transactions*. <https://corporate.visa.com/en/sites/visa-perspectives/newsroom/visa-partners-complete-secure-agentic-transactions.html>
11. IETF Datatracker, *AI Agent Authentication and Authorization*, draft-klrc-aiagent-auth-03, individual Internet-Draft. <https://datatracker.ietf.org/doc/draft-klrc-aiagent-auth/>
12. NIST NCCoE, 5 February 2026, *Accelerating the Adoption of Software and AI Agent Identity and Authorization*. <https://csrc.nist.gov/pubs/other/2026/02/05/accelerating-the-adoption-of-software-and-ai-agent/ipd>

13. NIST, 17 February 2026, *AI Agent Standards Initiative for Interoperable and Secure AI Agents*. <https://www.nist.gov/news-events/news/2026/02/announcing-ai-agent-standards-initiative-interoperable-and-secure>
14. OpenID Foundation, 7 October 2025, *Identity Management for Agentic AI*. <https://openid.net/new-whitepaper-tackles-ai-agent-identity-challenges/>
15. Microsoft Learn, *What is Microsoft Entra Agent ID?* <https://learn.microsoft.com/en-us/entra/agent-id/what-is-microsoft-entra-agent-id>
16. Google Cloud documentation, *Agent Identity overview*. <https://docs.cloud.google.com/iam/docs/agent-identity-overview>
17. Model Context Protocol, *Authorization* specification. <https://modelcontextprotocol.io/specification/2025-06-18/basic/authorization>
18. Michael and Roesner, 20 July 2026, *User Permissions for AI Agents, from Interfaces to Runtime Enforcement*, arXiv preprint. <https://arxiv.org/html/2607.13718v2>
19. South et al., January 2025, *Authenticated Delegation and Authorized AI Agents*, ICML 2025 position paper. <https://arxiv.org/html/2501.09674v1>
20. W3C, 15 May 2025, *Verifiable Credentials Data Model v2.0*. <https://www.w3.org/TR/vc-data-model-2.0/>
21. A2A Project, *Agent2Agent Protocol Specification*, including Agent Card signing. <https://a2a-protocol.org/latest/specification/>
22. Model Context Protocol, *The MCP Registry*. <https://modelcontextprotocol.io/registry/about>
23. Ethereum Improvement Proposals, *ERC-8004: Trustless Agents*. <https://eips.ethereum.org/EIPS/eip-8004>
24. Xiong et al., 2026, *Can Trustless Agents Be Trusted? An Empirical Study of the ERC-8004 Decentralized AI Agent Ecosystem*, arXiv preprint. <https://arxiv.org/abs/2606.26028>
25. Government Digital Service, 16 June 2022, *How we are improving inclusion for digital identity in government*. <https://gds.blog.gov.uk/2022/06/16/how-we-are-improving-inclusion-for-digital-identity-in-government/>
26. Law Commission, *Electronic execution of documents*. <https://lawcom.gov.uk/project/electronic-execution-of-documents/>
27. Visa, *The history and evolution of 3-D Secure*. <https://corporate.visa.com/en/solutions/visa-protect/insights/3d-secure.html>
28. European Commission, *Payment Services Directive: frequently asked questions* (including Strong Customer Authentication). https://ec.europa.eu/commission/presscorner/detail/en/MEMO_15_5793
29. Ministry of Justice, *Tribunal Statistics Quarterly: January to March 2022*. <https://www.gov.uk/government/statistics/tribunal-statistics-quarterly-january-to-march-2022/tribunal-statistics-quarterly-january-to-march-2022>
30. Ministry of Justice, *Tribunal Statistics Quarterly: January to March 2026*. <https://www.gov.uk/government/statistics/tribunals-statistics-quarterly-january-to-march-2026/tribunal-statistics-quarterly-january-to-march-2026>

31. Tom Loosemore, Computer Weekly, 14 April 2026, *Flood warning: How citizens' AI agents will swamp public services* (opinion and scenario). <https://www.computerweekly.com/opinion/Flood-warning-How-citizens-AI-agents-will-swamp-public-services>
32. UK Government, March 2026, *Making public services work for you with your digital identity* (GOV.UK One Login and GOV.UK Wallet). <https://www.gov.uk/government/consultations/making-public-services-work-for-you-with-your-digital-identity/making-public-services-work-for-you-with-your-digital-identity>
33. European Commission, *European Digital Identity*. https://commission.europa.eu/topics/digital-economy-and-society/european-digital-identity_en
34. Chris Schmitz, *Agentic Flooding Database*, UK cases (accessed 2 August 2026). <https://chrisschmitz.ai/flooding?country=UK>
35. Financial Ombudsman Service, *Embracing AI's transformational impact on consumer complaints*. <https://www.financial-ombudsman.org.uk/businesses/resolving-complaint/our-insight/embracing-ai-transformational-impact-consumer-complaints>
36. Financial Ombudsman Service, *Financial Ombudsman Service receives over 305,000 complaints in 2024/25 and Annual complaints data and insight 2025/26*. 2024/25 and 2025/26
37. Bar Standards Board, *Our Business Plan for 2026–27*. <https://www.barstandardsboard.org.uk/about-us/how-we-regulate/our-business-plan.html>
38. The Guardian, 9 November 2025, *AI-powered nimbyism could grind UK planning system to a halt, experts warn*. <https://www.theguardian.com/politics/2025/nov/09/ai-powered-nimbyism-could-grind-uk-planning-system-to-a-halt-experts-warn>
39. NHS Employers, 22 August 2025, *Guidance on the use of artificial intelligence in candidate applications*. <https://www.nhsemployers.org/articles/guidance-use-artificial-intelligence-candidate-applications>
40. Bedford Independent, *Council says AI tools are making Freedom of Information requests more complex*. <https://www.bedfordindependent.co.uk/council-says-ai-tools-are-making-freedom-of-information-requests-more-complex/>
41. OpenAI Help Center, updated 30 July 2026, *ChatGPT Work's Cloud browser allowlisting*. <https://help.openai.com/en/articles/11845367-chatgpt-works-cloud-browser-allowlisting>
42. Google for Developers, updated 16 July and 4 May 2026, *Google user-triggered fetchers; Authenticate requests with Web Bot Auth* (experimental). <https://developers.google.com/crawling/docs/crawlers-fetchers/google-user-triggered-fetchers> and <https://developers.google.com/crawling/docs/crawlers-fetchers/web-bot-auth>
43. Anthropic Help Center, updated July 2026, *Get started with Claude in Chrome*. <https://support.claude.com/en/articles/12012173-get-started-with-claude-in-chrome>
44. Apple Newsroom, 8 June 2026, *Due to DMA, Siri AI delayed in EU for iOS 27 and iPadOS 27*. <https://www.apple.com/newsroom/2026/06/due-to-dma-siri-ai-delayed-in-eu-for-ios-27-and-ipados-27/>
45. European Commission, 16 July 2026, *Commission provides guidance to Google for AI interoperability on Android*. https://digital-markets-act.ec.europa.eu/commission-provides-guidance-google-ai-interoperability-android-and-sharing-google-search-data-under-2026-07-16_en

46. Google, May–June 2026, *Android 17, Gemini Intelligence, Gemini Spark and Chrome auto browse*.
<https://android-developers.googleblog.com/2026/06/Android-17.html> and
<https://blog.google/products-and-platforms/platforms/android/gemini-intelligence/> and
<https://blog.google/innovation-and-ai/sundar-pichai-io-2026/> and <https://blog.google/products-and-platforms/products/chrome/gemini-3-auto-browse/>
47. The Economist, 6 August 2026, *From parking tickets to planning, the state is under siege by AI* (UK and international AI-assisted demand; quotes the author).
<https://www.economist.com/britain/2026/08/06/from-parking-tickets-to-planning-the-state-is-under-siege-by-ai>

Glossary

A2A (Agent2Agent Protocol). A proposed common way for agents to find and communicate with one another. Its Agent Cards are rather like directory entries: they describe who provides an agent, where to find it and what it claims to do. A digital signature can show who published the card, but not whether the claims are true.

Action-bound authorisation evidence. Technical shorthand for proof that connects a person's approval to the important action a service is about to take. AP2's payment instructions are examples.

Agent. In this briefing, software that can plan and take one or more actions for a person or organisation. Calling it an agent doesn't make it a legal person.

Agent attribution. A signal, ideally cryptographically verifiable, that tells a receiving service that software initiated a request and identifies the operator or agent surface. Attribution isn't the same as proof that the person authorised the action.

Agent evidence and acceptance policy. An organisation's house rules for software acting on somebody's behalf: which interactions it accepts, what proof it requires and what happens when something goes wrong.

Agentic flooding. This phrase is often used broadly for increased applications, appeals, complaints or claims produced with AI. This briefing distinguishes **AI-assisted flooding**, which is already visible in some UK services, from **autonomous-agent flooding**, in which software acts across services under delegated authority. The latter remains a plausible scenario, not a measured general trend.

API (application programming interface). A structured route that lets one computer system ask another for information or an action. It's the machine equivalent of a designed front door, rather than software trying to read and operate a page made for people.

AP2 (Agent Payments Protocol). A Google-led proposal for agent payments. It uses signed digital records to capture what somebody asked for and the details of the eventual purchase.

Delegated authority. Permission given by a person or organisation to software. Good permission is limited by the job, what the agent may do and how long it may act.

Digital signature. A mathematical seal attached to a message or document. It can show who held the signing key and whether the content changed afterwards. It doesn't, on its own, prove that the content is true or that a person gave informed consent.

HTTP Message Signatures. A technical standard for placing a digital signature on the messages that pass between services on the web.

KYA (Know Your Agent). The collection of checks and records a receiving organisation uses when software presents evidence of its identity and delegated authority. It covers where the request came from, which agent made it, what the person allowed and who can be held to account. It isn't one product, standard, register or universal method of detecting agents hidden inside ordinary user traffic.

MCP (Model Context Protocol). A proposed common way to connect agents to tools and sources of data. Its directory helps agents find those tools. It can use established access-control standards such as OAuth.

OAuth and OpenID Connect. Widely used standards for giving software limited access to an account and, separately, confirming information about the user. They underpin the familiar "Sign in with Google" and "Allow this app to access your account" journeys.

Opaque agent. An agent acting through the person's existing browser, app or operating-system session without presenting a reliable separate identity to the external service. The service may be unable to distinguish it from the person using the same account.

Principal. The technical name for the person or organisation on whose behalf an agent acts.

Prompt injection. Instructions hidden in a webpage, document, email or other content that try to hijack an AI agent and make it do something its user didn't intend.

Relying party. The organisation checking and deciding whether to accept a digital certificate, signed request or other proof. In this briefing it is normally the service being asked to act.

Revocation. Cancelling a digital pass or an agent's permission before it would otherwise expire. Cancellation only works if every receiving system checks the current status and obeys it.

SPIFFE / WIMSE. Technical work on giving each running piece of software its own short-lived, verifiable digital pass. These names are mainly relevant to identity and security teams.

Verifiable Credential. A signed digital certificate. It can carry a claim, for example that somebody is over 18, and show who issued it and whether it was altered. That doesn't prove that the claim is true or that the issuer is trustworthy.

Web Bot Auth. An emerging, voluntary way for bot and agent operators to sign web traffic so the receiving organisation can verify who sent it. It's currently used by some cloud-hosted agent traffic, not all agents or all requests. It can identify the sender. It can't prove that the request is safe or authorised by the user, and the absence of a signature doesn't prove that the visitor is human.