UNIVERSITY OF TARTU

Estonian Centre of Excellence in AI

Nortal

Kratt Nation

DIVIDING WORK BETWEEN PEOPLE, AI AGENTS AND RULES

DISCUSSION PAPER 2026



Kratt Nation

DIVIDING WORK BETWEEN PEOPLE, AI AGENTS AND RULES

DISCUSSION PAPER 2026

In Estonian folklore, a *kratt* gathers wealth for its maker, who must know how to keep it under control. Today's AI agents can prepare invoices, apply for benefits and handle work that is difficult to express as a fixed rule. For a country short of staff, that can add valuable capacity. Drawing on Estonia's experience, this paper examines how to divide work between people, AI agents and rules within democracy and the rule of law.

The person or organisation deploying an AI agent remains responsible for its use. That responsibility needs practical control: a payment limit must be enforced by the payment system, not merely written into the agent's instructions. Public and private service providers should test each proposed task, measure errors and expand independent action only when evidence supports it. Each service needs channel and continuity requirements suited to its purpose and applicable law. Essential services must remain practically accessible when agents stop.

Estonia already has many of the foundations this requires. Other countries will start from different legal and technical conditions. In each setting, the immediate task is to divide a service's work between rules, agents and people, establish the necessary controls and test the arrangement in practice.

This discussion paper was developed through voluntary collaboration between government, academia and the private sector. Its preparation was supported by the Estonian Centre of Excellence in Artificial Intelligence, EXAI (Ministry of Education and Research, TK213). This English edition adapts the Estonian manuscript for an international readership.

Working group: Andres Raieste (Global Head of Public Sector, Nortal), Dr Mihkel Solvak (Associate Professor of Technology Research, Tartu University), Keaty Sivelt (Chief Data & AI Officer, Ministry of Justice and Digital Affairs of Estonia), Toomas Vaks (Cyber Risk Manager, Swedbank), Madis Tapupere (independent expert), Priit Livak (Chief Government Technology Officer, Nortal) and Tõnu Grünberg (Deputy Secretary General for Digital Infrastructure and Cyber Security, Ministry of Justice and Digital Affairs of Estonia).

This paper may be cited as: Raieste, A., Solvak, M., Siivelt, K., Vaks, T., Tapupere, M., Liivak, P., & Grünberg, T. (2026). *Kratt Nation: Dividing work between people, AI agents and rules*.

CONTENTS

Executive summary 4

Introduction 6

1. The decisions agents change 10

2. Whose interests agents serve and what they may decide 14

2.1. The principal and the agent 14

2.2. Rules, agents and people 15

2.3. Automation and delegation to AI 15

2.4. Limits are enforced as rules 16

2.5. Action classes and autonomy 17

2.6. Decisions that remain with people 20

3. The person, the organisation and the agent 22

3.1. The personal agent 24

3.2. Three routes to a service 25

3.3. The external agent interface as the preferred route 26

4. What the organisation is responsible for 29

4.1. Internal agents 29

4.2. Parity between human and agent channels 30

4.3. Five principles for every agent-enabled service 33

5. What governments need to provide 35

5.1. Agents do not replace the need to improve administration efficiency 35

5.2. Adapt each organisation's systems to existing digital foundations 37

5.3. Let people see what agents did on their behalf and with their data 38

5.4. Shared authorisation when repeated needs justify it 39

6. Scenarios and service continuity in a Kratt Nation 42

References 44

EXECUTIVE SUMMARY

This paper focuses on public administration and regulated services, including banking, telecommunications and utilities. It uses 'organisation' for a public body or private entity responsible for the service, systems or actions being discussed. A service is the work an organisation does to meet a defined customer or staff need, including the steps performed behind the scenes. The paper examines which tasks and decisions may be delegated to agents and how responsibility is divided, starting with existing organisations and what they can do within current law. Agents must operate within democracy and the rule of law; their capabilities do not redefine those boundaries or confer authority to act.

Agents can handle work that is difficult to express as a fixed rule. They can read applications and supporting documents, identify missing information, prepare a draft assessment with references to its sources and explain an organisation's response in everyday language. A personal agent can also act while its user is away. These capabilities may be particularly useful where an ageing population or staff shortages put pressure on public services. Their wider value depends on clear responsibility for the agent's work and the quality of the whole service.

Software cannot take over a person's or organisation's legal duties. Those deploying agents retain their responsibilities, while liability for particular harm may also involve an operator, producer or other party. We propose four conditions for the use of AI agents:

1. An agent must not receive a person's authentication or signing credentials.

2. Its limits must be enforced by the organisation receiving the action.

3. Channel coverage and continuity requirements must reflect each service and applicable law, with essential services practically accessible when agents stop.

4. Decisions reserved to people by law or the organisation's policy must remain with people, with a practical route to challenge decisions that affect them.

An organisation should examine each step of a service and decide whether it should be performed by conventional software following predefined rules, by an AI agent or by a person. Stable tasks that can be specified precisely usually suit rule-based automation. An agent is more useful where inputs vary and require interpretation. For example, in a grant service, rules check objective eligibility conditions, an agent prepares an assessment against the scheme's objectives, and an official makes the discretionary decision.

An instruction asks the agent to observe a limit. Enforcement requires the receiving system to block actions that exceed it. Documents or websites may contain instructions that divert an agent, but they should not be able to change the payment system's authorisation rules. The scope of independent action should depend on the consequences of error and the practical possibility of repair. Requiring approval of every minor action risks turning review into a routine click. Organisations should first test the agent alongside staff, measure errors through appropriate samples and test

any proposed expansion before granting it. Serious errors or deteriorating results should narrow or suspend the permission.

An agent using a person's credentials through an ordinary website can be difficult for an organisation to distinguish from the person. A dedicated agent interface allows the person to authorise specific activities in the organisation's own system. The organisation checks each action against that authorisation and reserves the necessary steps for human confirmation. This gives the organisation a clearer basis for control and a way to restrict agent traffic while keeping services open to people.

Organisations must meet applicable requirements to inform users when they are interacting with AI. Their customer-facing agents should also identify the organisation they serve. Every service using agents should have a named service owner, and each agent should have someone responsible for enforcing its limits and able to stop it immediately. Staff should be able to continue essential work during an outage, and the time and cost of changing model provider should be known and tested. Time released by an agent adds capacity but does not automatically become a budget saving. Management should decide whether to reduce waiting times, improve service quality or use that capacity elsewhere.

The necessary foundations include reliable identification, checks on a person's authority, lawful data access and systems that distinguish agent sessions from human users. Where these exist, organisations can adapt individual services; elsewhere, missing capabilities need to be addressed within a bounded use case. Government can provide shared guidance so that each organisation does not have to design consent screens and authorisation logic alone. People should be able to see both actions performed on their behalf and the data used. A shared authorisation accepted by several organisations becomes worth pursuing when repeated needs justify its cost and complexity.

A personal agent may help someone navigate fragmented services, but it does not remove the fragmentation. Organisations should simplify services themselves and agree who coordinates a complete response to a defined need. Otherwise, the cost moves to the person or provider paying for the agent, while more easily generated applications and complaints may add work for public bodies.

Some choices remain political. Governments need to decide how to balance growth, wider social outcomes, additional capacity and financial savings. Legislatures may need to consider liability and conflicts between an assistant provider's commercial interests and its user's interests. Societies must decide which roles they wish to retain for people. Countries with less widely supported languages also face choices about language quality, cost and access to training material. Whether a service should be equally fast and good without an agent remains a separate policy choice.

Organisations can begin within their existing authority. For a selected service, they should establish what an agent may do, who checks the result and how the service continues or is restored without it. The proposed action classes and coverage measures are intended for testing in real services and have not yet been validated as a national model.

INTRODUCTION

This paper asks which tasks and decisions a person or organisation should entrust to an AI agent, under what conditions, and who remains responsible. It is written primarily for policymakers and public-service leaders, with detail for their legal and technical advisers. The organisation responsible for a service or its infrastructure should organise how agents use it. People should retain meaningful access to the service and a way to question its decisions.

In Estonian folklore, a kratt is a creature made from old household objects that gathers wealth for its maker. Similar motifs occur in other European traditions: Germany's money-bringing Geldmännlein and Sweden's bjära. English folklore offers a looser parallel in Puck, a mischievous spirit. Andrus Kivirähk gives the difficulty of managing a kratt a memorable voice in his novel Rehepapp ehk november (2000): "You don't know how to handle a kratt."[1]

In the age of AI agents, these old stories invite practical questions about whose interests a helper serves, who bears the cost and how its actions can be kept within bounds. This paper examines those questions through the division of work: which tasks belong to rules, which can be entrusted to AI agents, and which decisions remain with people. These arrangements must operate within democracy and the rule of law, with clear responsibility and limits enforced by the systems through which agents act.

In this paper, an agent is software that interprets a task and can use tools to take actions. We use the international term throughout the analysis and retain kratt as the cultural name.

An agent's instructions may guide its behaviour, but enforceable limits belong in the systems through which it acts. The person using a commercial assistant and the organisation providing a public service have different responsibilities. Chapter 3 explains that relationship from the service user's perspective. The remaining chapters address organisational choices and the support governments can provide.

The proposals are intended to work together within applicable law. The four conditions in the summary are requirements advocated by this paper; identical legal duties do not apply everywhere. Action classes, evidence-based changes to autonomy and service-coverage measures need testing in practice. Legal examples are identified as EU or Estonian examples and need assessment against local law before use elsewhere. Existing protections may already require human involvement.[2] We use 'should' for recommendations and identify the source and scope of legal requirements.

Agents may connect services across organisational boundaries, but those boundaries sometimes protect legitimate interests. Estonia's X-tee data exchange layer, for example, connects distributed systems rather than creating a single central database. Access to another organisation's data still requires an appropriate basis and agree-

1 Comparable figures include Sweden's skratte and bjära, Germany's Schrat and Geldmännlein, Poland's skrzat and, more loosely, England's Puck. Quotation: Kivirähk, Rehepapp ehk november (2000), translated from Estonian for this paper.

2 For example, GDPR Article 22 and Consumer Credit Directive Article 18(8). These safeguards have defined scopes and are not a general prohibition of all automated decisions.

ment.[3] Governments can reform administrative structures and simplify services while preserving clear responsibility and control over data use. Democratic authority must also be reflected in the technical design: who may decide, what may be delegated and who can intervene.[4] An agent's capabilities do not change a public authority's legal competence.

Identity systems, administrative law and service organisation differ between countries. Chapter 5 distinguishes the necessary functions from the existing and proposed Estonian arrangements used to illustrate them. The paper does not attempt a general account of (agentic) AI or a complete programme for funding, procurement and skills. It starts from capabilities that can be tested today. Organisations can decide how to allocate work without predicting the future limits of the technology.

Organising a service with agents, while the organisation remains accountable for it, involves three stages (Figure 1). First, each step is assigned to a rule, an agent or a person. Next, the agent's permissions and scope of independent action are set. Finally, review, responsibility and service continuity or recovery are organised.

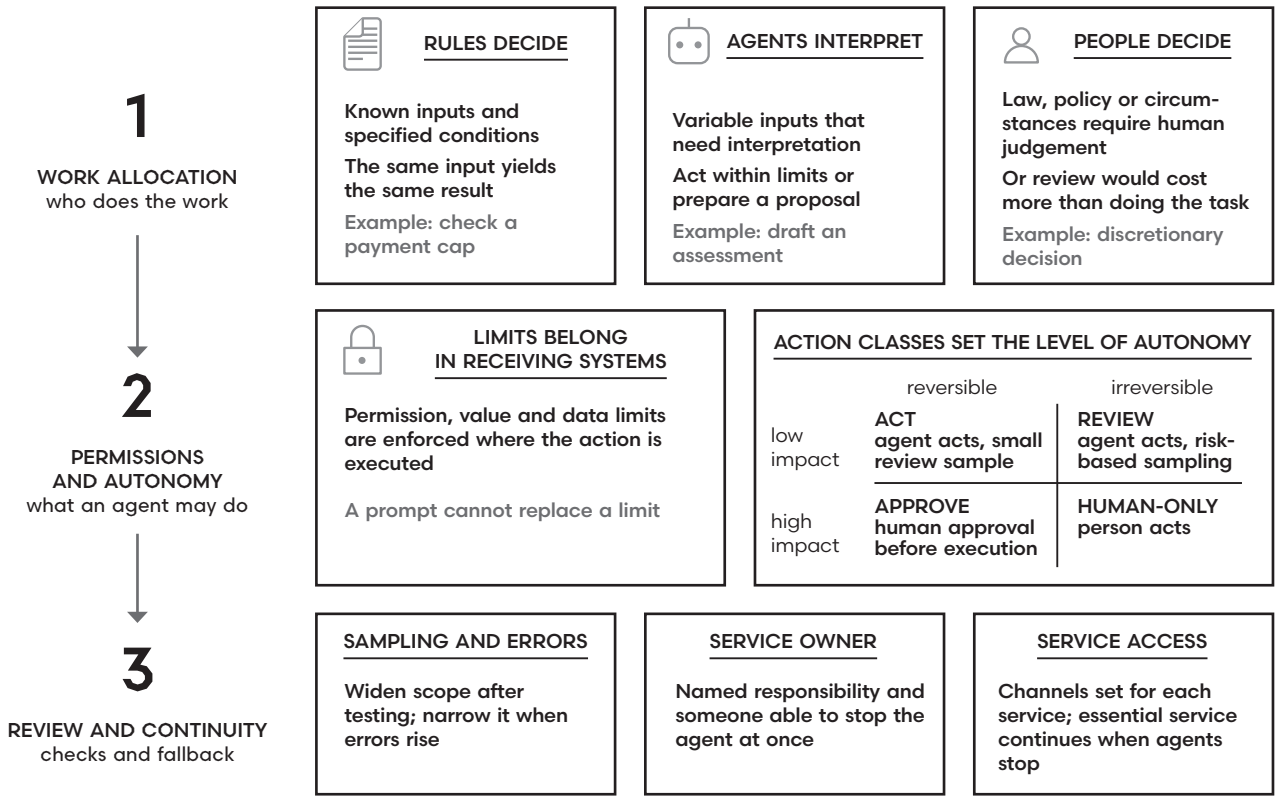


Figure 1. Three stages in organising a service with agents: allocate the work, set permissions and autonomy, then organise review and continuity.

3 Estonian Information System Authority (RIA), 'Introduction of X-tee'.

4 Raieste, Solvak, Velsberg and McBride, 'Government Efficiency in the Age of AI: Toward Resilient and Efficient Digital Democracies', 2025.

TERM	MEANING IN THIS PAPER
AI agent / kratt	Software that interprets a task and can use tools to take actions. Kratt is the Estonian cultural name retained in the title. Cooperating agents are treated as one arrangement for assigning controls; this does not erase the roles of their providers.
Principal	The person or organisation deploying an agent and delegating a task to it. Using an agent does not remove the principal's duties. The term does not imply that software has legal personality.
Organisation	A public body or private entity responsible for the service, systems or actions being discussed.
Agent operator	The organisation operating the agent as a service, for example a commercial assistant provider.
Model provider	The organisation supplying the model used by the agent. This role may differ from the operator and the provider of the complete AI system.
Area of responsibility	The part of an arrangement for which a person or organisation is responsible: the user's tasks and permissions, or the organisation's systems and agents. Responsibility requires practical control. This is not an exhaustive allocation of legal liability.
Organisation's service agent	The organisation's customer-facing AI agent, operating within its remit and limits. In this paper, 'the organisation's AI agents' refers to its service agents and internal agents.
Internal agent	An AI agent integrated into an organisation's systems to support or perform defined staff tasks.
External agent interface	An interface through which an organisation exposes defined actions to external agents, with checks against each action's authorisation.
Authorisation	A machine-readable technical permission granted in the organisation's own system by a person who has authenticated and holds the relevant rights. It defines on whose behalf an agent may act, what it may do and within which limits. It does not itself confer legal representation or determine liability.
Human interface and human route	The human interface is an organisation's website, application or staff-assisted channel. The human route shows which needs can be met without a personal agent. Channel coverage and continuity requirements are set per service; essential services must remain practically accessible.
Action class	A proposed control for an individual step: ACT, REVIEW, APPROVE or HUMAN-ONLY. Consequences and practical reversibility inform the class; steps reserved to people by law or the organisation's policy override the matrix.
Service parity	A measure of the needs covered by the human and agent routes, and of how many of those needs have an explicit decision on channel choice and the agent's role. Coverage is compared with targets for each service and role. This does not give agents the same authority as people.
Service owner	The person accountable for the service as a whole, able to arrange changes and remedies. In a shared service, the owner coordinates participants without taking over their legal competence.

Table 1. Terms used in this paper.

Chapter 1 considers the wider decisions affected by agents, followed in Chapter 2 by the division of work and controls outside the model. Chapters 3 and 4 examine how people and their agents deal with organisations, then the organisation's own agents, systems and responsibilities. Shared support and infrastructure are the subject of Chapter 5; Chapter 6 brings the proposals together in two hypothetical scenarios. Table 1 defines the main terms.



1. THE DECISIONS AGENTS CHANGE

Research on AI often focuses on labour and productivity. Studies report substantial gains in some knowledge-work tasks, with results depending on the task, worker and organisation.[5] Those effects matter to public bodies, which also face decisions about responsibility, culture and national dependence. Before deciding what an agent may do, it helps to distinguish four dimensions: economic, legal, cultural, and sovereignty and security. Table 2 separates the choices of individuals, organisations and government.

Economically, an agent may increase the amount of work one person can complete. That additional capacity can itself be a policy objective where staff are scarce. Jevons's account of resource efficiency illustrates why cheaper activity can also increase demand.[6] Cheaper code generation, for example, may lead to more software being built. Released staff time therefore does not automatically reduce expenditure. For each service, management should decide how the capacity will be used and measure completed cases and outcomes. Governments should decide whether the main aim is growth and wider social outcomes or cost savings and organisational efficiency. These aims need different policies and incentives; shorter queues are one possible service outcome.

Legally, deploying software does not transfer the deployer's duties to it. Responsibility for operating a service, accountability for explaining and correcting its decisions, and liability to compensate for harm are distinct questions. The EU's revised Product Liability Directive expressly covers software, including AI systems, but its new regime concerns products placed on the market or put into service after 8 December 2026 and requires national transposition.[7] Other applicable law also allocates obligations by role.[8] Depending on the facts and jurisdiction, an agent operator, producer, credential issuer or receiving organisation may bear liability. A person's authorisation alone does not settle that allocation.

Operational responsibility needs practical means of control. Every service using agents should have a service owner able to change its behaviour and coordinate a remedy when something goes wrong. This does not make the service owner the sole party liable for every loss. Legislatures should assess whether existing rules adequately cover agent-mediated actions before creating a separate liability regime.

Cultural choices concern roles people may wish to retain for humans in education, care, cultural life or justice, even where an agent could deliver a better result at lower cost. Technical feasibility does not settle what society wants to delegate. Estonia's

---

5 Review: Calvino, Reijerink and Samek, 'The Effects of Generative AI on Productivity, Innovation and Entrepreneurship', OECD, 2025. In experiments, professional writing tasks took 40% less time (Noy and Zhang, 2023), while a pooled estimate from three companies attributed about 26% more completed software-development tasks to use of the coding assistant (Cui et al., 2026). These task-level results do not establish equivalent gains across complete public-service workflows.

6 Jevons, The Coal Question, 1865, Chapter VII. The analogy identifies a possible demand response; it is not evidence of a fixed effect from AI deployment.

7 Product Liability Directive (EU) 2024/2853, Articles 2, 4 and 22, and the 7 May 2026 corrigendum. The new regime covers products placed on the market or put into service after 8 December 2026. National transposition and the facts of a claim matter.

8 For example, GDPR Article 82(2)–(5), on the respective liability of controllers and processors and recourse between responsible parties.

national song celebration, Laulupidu, brings large choirs together and has a particular place in cultural identity. Whether AI-generated songs belong there is a cultural choice. Allowing an agent to exercise someone's vote would raise constitutional questions. Such decisions belong to society and its democratic institutions.

Service leaders should identify where users expect to deal with a person and publish their commitments to human support. An undisclosed AI agent acting for the organisation can frustrate that expectation even when its answer is correct. In the EU, Article 50 of the AI Act requires providers of systems intended to interact directly with people to ensure they are informed that they are interacting with AI, unless this is obvious in the circumstances.[9] This paper also recommends stating whose interests the agent serves and what it cannot do.

Sovereignty and security concern the dependence a country is willing to accept. A country may rely on model suppliers subject to another jurisdiction's export controls. Several providers can share this source of disruption, so using more than one supplier does not necessarily remove the dependence. The policy question is which dependencies can be managed, and which would make a critical service too vulnerable.

For Estonia, language is one part of that dependence. Research on tokenisation has found that equivalent content can require more text units, or tokens, in Estonian than in English.[10] For services priced by token, this can raise costs, although the difference depends on the model and pricing. If the gap persists, users or organisations bear the extra cost or face an incentive to switch languages. Other less widely supported languages face similar questions. Organisations should measure language quality and cost; governments should decide which public language resources to make available for training, subject to their rights and restrictions.

An outage or supply restriction can remove access to a model, so critical services need a workable response. When information is sent to an externally hosted model, the organisation must decide what data may leave its systems and under what conditions. Information classification and applicable law may rule out particular uses even where the model appears capable of the task.

---

9 EU AI Act, Article 50(1). The provision addresses providers of systems intended for direct interaction with natural persons and includes an exception where the interaction is obvious in context.

10 Petrov et al., 'Language Model Tokenizers Introduce Unfairness Between Languages', 2023, Appendix C, p. 23. This establishes differences in tokenisation, not a universal cost ratio or comparative model quality for every language.

DIMENSION	WHAT CHANGES	INDIVIDUAL CHOICE	THE ORGANISATION'S CHOICE	POLICY CHOICE
Economic	More work per person; released time is not automatically a saving	Tasks to delegate	Use of capacity; outcome measures	Growth and social outcomes versus cost savings
Legal	Delegation leaves people's and organisations' duties in place	Scope of permission	Service ownership, limits and remedies	Whether existing law adequately covers agent actions
Cultural	Some roles remain human by deliberate choice	Preferred form of assistance	Where human support is expected	Roles society reserves to people
Sovereignty and security	Dependence on language support, supply and data flows	Which assistant to trust	Data restrictions, language quality and continuity	Acceptable foreign supply dependence; shared foundations and language resources

Table 2. Four dimensions of choice and the decisions belonging to individuals, organisations and government.



Whether AI-generated songs belong at Estonia's national song celebration is a cultural choice.

PHOTO: ANDREA FORLANI / ESTONIAN WORLD

2. WHOSE INTERESTS AGENTS SERVE AND WHAT THEY MAY DECIDE

Across all four dimensions, organisations first need to decide what to delegate and how to check the work. This is difficult when the process behind an agent's output is only partly visible.

2.1. THE PRINCIPAL AND THE AGENT

An AI agent works for someone, and that someone is not necessarily the person giving it a task. Economics describes a related difficulty as the principal–agent problem: how can a principal secure good performance from a delegate whose interests may differ and whose work cannot be fully observed, without spending more on incentives and monitoring than the work is worth? A principal may see the result, such as the customers brought in by a sales representative, without knowing how much came from the representative's effort and how much from the product. Behind an AI agent there may also be an operator with a business model of its own, in effect a second principal whose interests the agent may also serve.

The problem became a major subject of research in the 1970s.[11] Jensen and Meckling described in 1976 how the interests of parties to a delegation can diverge, and how detecting and limiting that divergence has a cost.[12] Two difficulties are particularly relevant here.

Information asymmetry. The user sees an output without necessarily seeing the work that produced it. Avoiding that work was the reason for delegating it. The output may depend on external information, tools and other agents, and the user may not know whether it was produced through an authorised process.

Diverging objectives. Software need not have interests of its own for the problem to arise. The operator or organisation behind it does have interests. An agent may misunderstand a user's goal, optimise for an operator's commercial objective, or gradually depart from what the user intended. It may take risks the user would have declined, leaving the user or others to bear the consequences.

Principal–agent theory often uses incentives linked to results because direct monitoring is expensive or incomplete. Later work has examined better incentives[13] and questioned when information asymmetry and conflicting interests are the right starting assumptions.[14] Research on AI agents extends these questions to automated delegation.[15] Sub-delegation to other agents or the presence of several principals can make the work harder to observe.[16]

---

11 Spence and Zeckhauser, 'Insurance, Information, and Individual Action', 1971.

12 Jensen and Meckling, 'Theory of the Firm', 1976.

13 Holmström, 'Moral Hazard and Observability', 1979; for an overview, Gailmard, 'Accountability and Principal–Agent Theory', 2014.

14 Fehr, Gächter and Kirchsteiger, 'Reciprocity as a Contract Enforcement Device', 1997; Waterman and Meier, 'Principal-Agent Models: An Expansion?', 1998; Miller and Whitford, 'Trust and Incentives in Principal-Agent Negotiations', 2002.

15 Kim, 'Deep learning and principal–agent problems of algorithmic governance', 2020; Dowding and Taylor, 'Algorithmic Decision-Making, Agency Costs, and Institution-Based Trust', 2024; Feng and Chandra, 'Principal–agent issues when governments embrace AI agents', 2026.

16 Rajala and Jalonen, 'Beyond Simplification in Public Sector Accountability', 2025; Feng and Chandra, 'Principal–agent issues when governments embrace AI agents', 2026.

A person may also defer so consistently to an agent's recommendations that the agent effectively determines the choice.[17]

Personal autonomy, understood as the ability to set goals and act on them, can increase when people deliberately delegate administration and gain time for other things. It can diminish when an agent stays within its technical permissions but quietly steers their choices.[18]

Aligning AI systems with human values is a wider question than this paper addresses.[19] A user may choose an agent based on any model. A receiving organisation therefore needs controls that remain effective regardless of that model's internal behaviour.[20]

Training incentives and contractual commitments may improve behaviour, but they cannot replace enforceable limits. Political theory offers a useful parallel: procedures can constrain delegated public power when incentives are an imperfect control.[21] An organisation needs controls it can implement and sustain.[22] Their cost belongs in the assessment of whether delegation is worthwhile; legal safeguards still apply.

2.2. RULES, AGENTS AND PEOPLE

Whether a step can be described by a stable rule or requires interpretation of variable information determines how work is divided between rules, agents and people. The seriousness and reversibility of a possible error then limit the agent's freedom within its allotted work. Technical capability does not confer authority.

A rule handles known inputs according to specified conditions. An agent handles work in which inputs vary and interpretation is needed. A person decides where the law, the organisation's requirements or the circumstances require human judgement. This allocation is made for each step and all three can contribute to one service.

2.3. AUTOMATION AND DELEGATION TO AI

In conventional automation, work is specified as rules and then performed repeatedly. Building and maintaining the rules has a cost, but processing another case is usually inexpensive and predictable. Rules still require a defined response to exceptions; they cannot resolve a case that their designers have not accounted for.

An agent interprets the task again on each occasion: it reads the document, works out what to do and uses the available tools. This can lower the initial cost of supporting variable tasks. Each case still uses computing resources and may need checking,

---

17 Giacomini and Grandi, 'I chose to let the machine choose', 2026.

18 Laitinen and Sahlgren, 'AI Systems and Respect for Human Autonomy', 2021; Fischli et al., 'Agents, Alignment, and the Many Faces of Autonomy', 2026.

19 Gabriel, 'Artificial Intelligence, Values, and Alignment', 2020.

20 Chan et al., 'Infrastructure for AI Agents', 2025, Sections 1 and 4.2 (arXiv v3, pp. 2–4 and 13–15).

21 McCubbins, Noll and Weingast, 'Administrative Procedures as Instruments of Political Control', 1987.

22 Schillemans and Busuioc, 'Predicting Public Sector Accountability', 2015, pp. 206–208, on whether accountability forums exercise oversight in practice. The cost assessment in the text is this paper's recommendation.

while identical inputs can produce different outputs. The agent can address cases for which no explicit rule has been written.

If an organisation asks an agent ten thousand times a year whether the amount on a structured electronic invoice exceeds a limit, it pays repeatedly for interpretation that could have been specified once. For an invoice received as an image, optical character recognition can extract the amount and the limit check remains a rule. A specialised model is usually cheaper and more consistent for a high-volume task such as classification or text recognition. Receiving a machine-readable invoice may remove the extraction work altogether. A fixed rule, however, cannot judge whether a vaguely described line item corresponds to work the organisation actually ordered.

Michael Polanyi argued in 1966 that people know more than they can explicitly describe. David Autor later examined how this limit had shaped resistance to automation.[23] AI moves that boundary by performing some tasks for which people could not specify the rules in advance. Where a stable rule can be written and reused at scale, conventional automation should be the first option considered.

Organisations also have tasks performed only a few times a year, for which building dedicated automation has never been worthwhile. An existing agent and its controls may make these tasks economical to support. With so few cases, however, an error rate is difficult to estimate. The agent can prepare the work while a person reviews every case.

An agent can also help build conventional automation by drafting code, deriving candidate rules from examples and preparing tests. The organisation must validate those rules before deployment. The resulting process can then run without an agent.

2.4. LIMITS ARE ENFORCED AS RULES

If an agent may make payments up to €50, that limit is a rule and must be enforced deterministically (Table 3). A sentence in a prompt is open to interpretation; a contractual promise also requires interpretation and enforcement. Neither substitutes for a system check. The same applies to permitted data, the person represented, the duration of authorisation and steps requiring human involvement. The system must prevent any actor from exceeding the permissions of the role or channel being used. The agent may interpret the task, but cannot rewrite its authority.

The limit belongs where the action is executed: the payment system at a bank or the case-management system at a public authority.[24] A receiving organisation cannot rely on controls inside an external commercial agent that it neither operates nor inspects. The user sets the scope of the agent's authorisation and the organisation enforces it, together with the organisation's rules and legal restrictions that the user cannot waive.

---

23 Polanyi, The Tacit Dimension, 1966; Autor, 'Polanyi's Paradox and the Shape of Employment Growth', 2014.

24 Bogdanov et al., study of AI and machine-learning technology risks and mitigation, 2024, Section 6.2.2, p. 78: a model's request to a connected service must not carry greater rights than the user's. An agent may need narrower rights still.

Agents also read documents and websites containing text that can resemble instructions. Prompt injection can redirect their behaviour, so the organisation should not depend on the model always rejecting such instructions.[25] A prompt saying "do not pay more than €50" influences behaviour. A payment-system check that rejects an amount above €50 enforces a boundary without interpreting the document's instructions. That boundary must hold even when the agent misunderstands its task.

	ENFORCED SYSTEM LIMIT	INSTRUCTION OPEN TO INTERPRETATION
Location	The system executing the action	A prompt, contract or another model's assessment
Basis	Authoritative permissions and validated action data	Model output or text read by the agent
Can a prompt change it?	No, if the control and its inputs are properly protected	Instructions may be misinterpreted or overridden
Effect	An action outside the enforced boundary is blocked	An unwanted action may become less likely
Suitable uses	Permission scopes, data access and spending limits, with audit logs	Tone, assistance and preliminary screening

Table 3. Enforced limits and instructions open to interpretation.

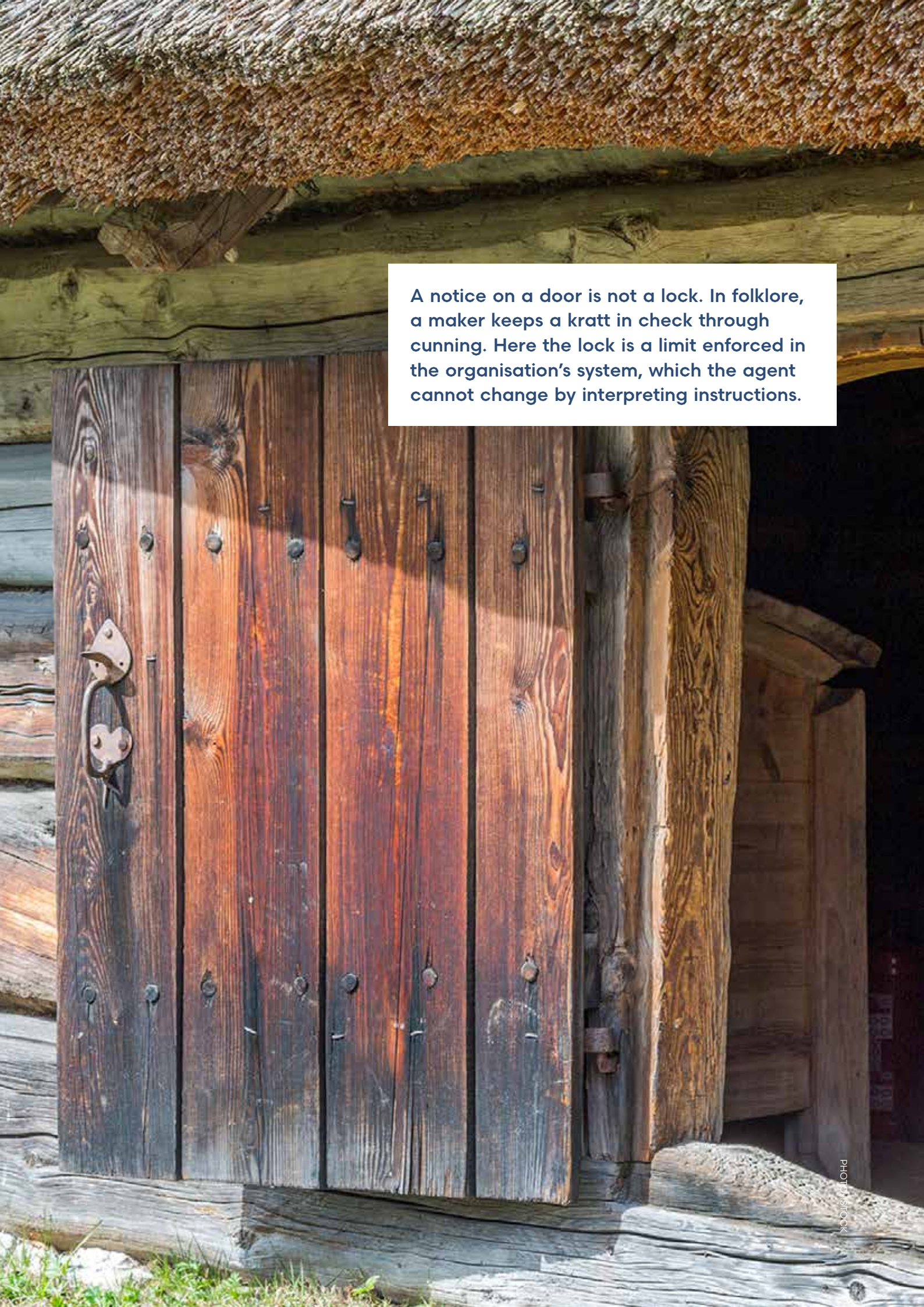
An enforced limit does not establish that an action inside it was correct or served the user's interests. An agent can pay the wrong invoice within a valid authorisation, or make many individually permitted payments whose combined effect was never intended. Aggregate limits on total value or action count help with the latter risk. Review against the underlying evidence is still needed to assess whether permitted actions were appropriate. Logs establish what happened and the recorded basis for it.

A notice on a door is not a lock. In folklore, a maker keeps a kratt in check through cunning. Here the lock is a limit enforced in the organisation's system, which the agent cannot change by interpreting instructions.

2.5. ACTION CLASSES AND AUTONOMY

This paper proposes four action classes for deciding how independently an agent may act. They reflect a step's effects on people's rights, property or access to services. Preparation, approval and execution are separate steps and can receive different classes. Classification asks how serious the consequences could be and whether the action can realistically be reversed or its consequences remedied, including the cost of correction. Table 4 gives a starting matrix: ACT, REVIEW, APPROVE and HUMAN-ONLY. A decision reserved to a person by law or the organisation's policy is always HUMAN-ONLY. The organisation may impose stricter controls where the context requires them.

25 OWASP, 'LLM01:2025 Prompt Injection'; Bogdanov et al., 2024, Section 6.2.2, p. 78.



A notice on a door is not a lock. In folklore, a maker keeps a kratt in check through cunning. Here the lock is a limit enforced in the organisation's system, which the agent cannot change by interpreting instructions.

PHOTO: ISTOCK

The organisation sets thresholds for each action type in advance. Its system assigns the class using structured facts such as amount and recipient; an agent's description cannot lower it. Classification controls an individual step and does not replace a risk assessment. Many small irreversible actions may together cause substantial harm. In the grant example, registering a submission is REVIEW because it creates an official record; a routine request for a missing attachment may be ACT only if it cannot change a deadline, create a new procedural requirement or automatically suspend processing. A decision not to examine an incomplete application requires at least prior human approval, and remains HUMAN-ONLY wherever applicable law reserves that decision to an official.

CLASS	IMPACT	REVERSIBLE	CONTROL EXAMPLE	PUBLIC AUTHORITY EXAMPLE	BANK EXAMPLE
ACT	Low	Yes	Agent acts; actions logged; small human review sample	Low-consequence explanation; attachment request without changed deadlines, new procedural duties or automatic suspension	Prepare an invoice or payment draft
REVIEW	Low	No	Agent acts; human sampling sized to risk	Register a submission in the case record	Execute a low-consequence payment within a valid authorisation and limit
APPROVE	High	Yes	Human approval before execution	Non-examination for an uncorrected defect, only where human-approved execution is lawful and an effective remedy exists	Add a trusted recipient; change a limit (always needs approval)
HUMAN-ONLY	High	No, or reservation overrides	Person performs the reserved step	Make a discretionary administrative decision	Make a lending decision reserved to a person

Table 4. Proposed action classes. Steps reserved to a person by law or the organisation's policy override this matrix.

Checking every REVIEW action would recreate routine approval; checking none would leave the organisation ignorant of performance. A random sample, sized for the consequences and the evidence needed, estimates error while keeping review manageable. It should cover important case groups, since an average can conceal failures in a particular scheme or population. A single severe error may justify intervention regardless of the average. Low-consequence free-text communications in ACT also need sampling. Communications that can affect rights or deadlines require stronger controls from the outset.

An error rate below the agreed threshold could (taking into account also other aspects) justify considering a wider operating range, such as a higher amount before approval is required. When the operating range is widened, the threshold changes within the same classification framework. Performance on small cases alone does not establish safety on larger ones. The new range should first be tested with human approval or in shadow operation, then introduced with increased sampling. Rising

errors narrow the range again. A material change in model, tools, data or task also triggers renewed evaluation. Human-reserved decisions and changes to the agent's own authorisation are never released merely because its error rate falls.

Ten wording errors do not have the same weight as one wrongful rejection. Errors should be weighted by consequences, and the organisation should also track how often people substantially revise the agent's proposals. The organisation decides how much autonomy to permit, based on results measured by action class for the relevant task and case groups; it can revoke that permission.

2.6. DECISIONS THAT REMAIN WITH PEOPLE

A person decides where law or the organisation requires it. This paper recommends that an official make and sign discretionary administrative decisions. A bank may likewise reserve lending decisions to people. Rule-based automatic decisions remain possible where applicable law permits them. In the EU, national measures implementing the revised Consumer Credit Directive are to apply from 20 November 2026; they include a right to human intervention where a creditworthiness assessment involves automated processing of personal data.[26] GDPR safeguards also apply where relevant. An agent may prepare a decision, but the decision-maker needs the evidence, authority and time to reject its recommendation. European data-protection guidance describes meaningful human involvement as review by someone able to change the decision.[27]

The EU AI Act classifies systems by intended use. Answering a balance enquiry is different from assessing a person's creditworthiness or eligibility for essential public assistance. The latter uses are listed in Annex III, subject to the Act's classification rules and exceptions; not every grant assessment is high-risk. Under the amended timetable, the relevant Annex III requirements generally apply from 2 December 2027, with transition provisions to check for each system.[28] The action classes proposed here operate within a use case and do not replace the Act's risk categories.

A person may be better placed to do the work directly if checking an agent's result costs more without improving the outcome. Repeated approval of hundreds of proposals can turn oversight into a formality. Bainbridge described how reliable automation can leave people poorly prepared to detect and address rare failures.[29] The AI Act's human-oversight provisions address over-reliance on automated output and the ability to intervene or stop a high-risk system.[30] Deterministic checks should handle technical violations, leaving reviewers the questions that require judgement. In public administration, rejecting an unauthorised machine action must remain distinct from refusing to receive a person's application.

---

26 Consumer Credit Directive (EU) 2023/2225, Articles 18(8) and 48. The right concerns a creditworthiness assessment involving automated processing of personal data. Member States' implementing measures apply from 20 November 2026; national implementation must be checked.

27 Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling', 2018, pp. 21 and 27, endorsed by the European Data Protection Board; GDPR Article 22.

28 EU AI Act, Article 6(2)–(3), Annex III points 5(a)–(b), and Article 113 as amended by Regulation (EU) 2026/1744. Check Article 111 transition provisions and whether the system falls within the stated use and exceptions.

29 Bainbridge, 'Ironies of automation', 1983, pp. 775–777.

30 EU AI Act, Article 14(4)(b) and (e), and Article 26(2), with the relevant application and transition dates.

At a public authority. A company's agent prepares and submits a grant application through an authorised interface. The system checks authorisation and structured information, records receipt and flags defects or questions about deadlines or eligibility for the required procedure. In Estonia, a software gate cannot replace the duty to receive applications and allow defects to be corrected.[31] The authority's internal agent drafts an assessment of how the proposal meets the scheme's objectives, with references to the application and rules. It does not make the discretionary administrative decision, which can have substantial consequences and may be difficult to reverse. An official considers the draft and the original sources and makes the decision.

At a bank. A customer's agent is asked to pay incoming invoices. Before each payment, the system checks authorisation, limits and applicable payment requirements. A small payment within the permitted range may go ahead with sampled checks of execution and compliance: its consequences are limited, but it may not be recoverable. The bank's checks do not establish whether every invoice serves the customer's purpose; the agent flags an uncertain invoice for the customer. Raising a limit has greater consequences. Although it can be reversed, the customer must approve it in the bank's own application. Otherwise, the agent could raise its limit and immediately make a large payment.

31 Estonian Administrative Procedure Act, Sections 14–15, current text. An administrative authority generally receives an application despite defects and allows correction, subject to statutory exceptions. Channel authentication, legal receipt and a decision on the merits must be distinguished.

3. THE PERSON, THE ORGANISATION AND THE AGENT

An agent-enabled service has two areas of operational responsibility. A person is responsible for the tasks and permissions they give their agent; an organisation is responsible for its systems and the actions those systems accept. Each needs practical means of control. Within its lawful authority, the organisation determines who may use the service, what their agents may do, and how work is divided between rules, agents and people. These responsibilities do not settle every question of legal liability: operators, providers and other parties retain their own duties.

The two areas remain separate. A personal agent submits requests through an organisation's interface within the user's authorisation; it does not gain general access to internal systems. Behind the interface, a rule, one of the organisation's AI agents or conventional software may handle the request. This chapter examines the organisation as the counterpart to a personal agent. Chapter 4 considers agents inside organisations. The allocation of work, enforced limits and action classes apply to both.



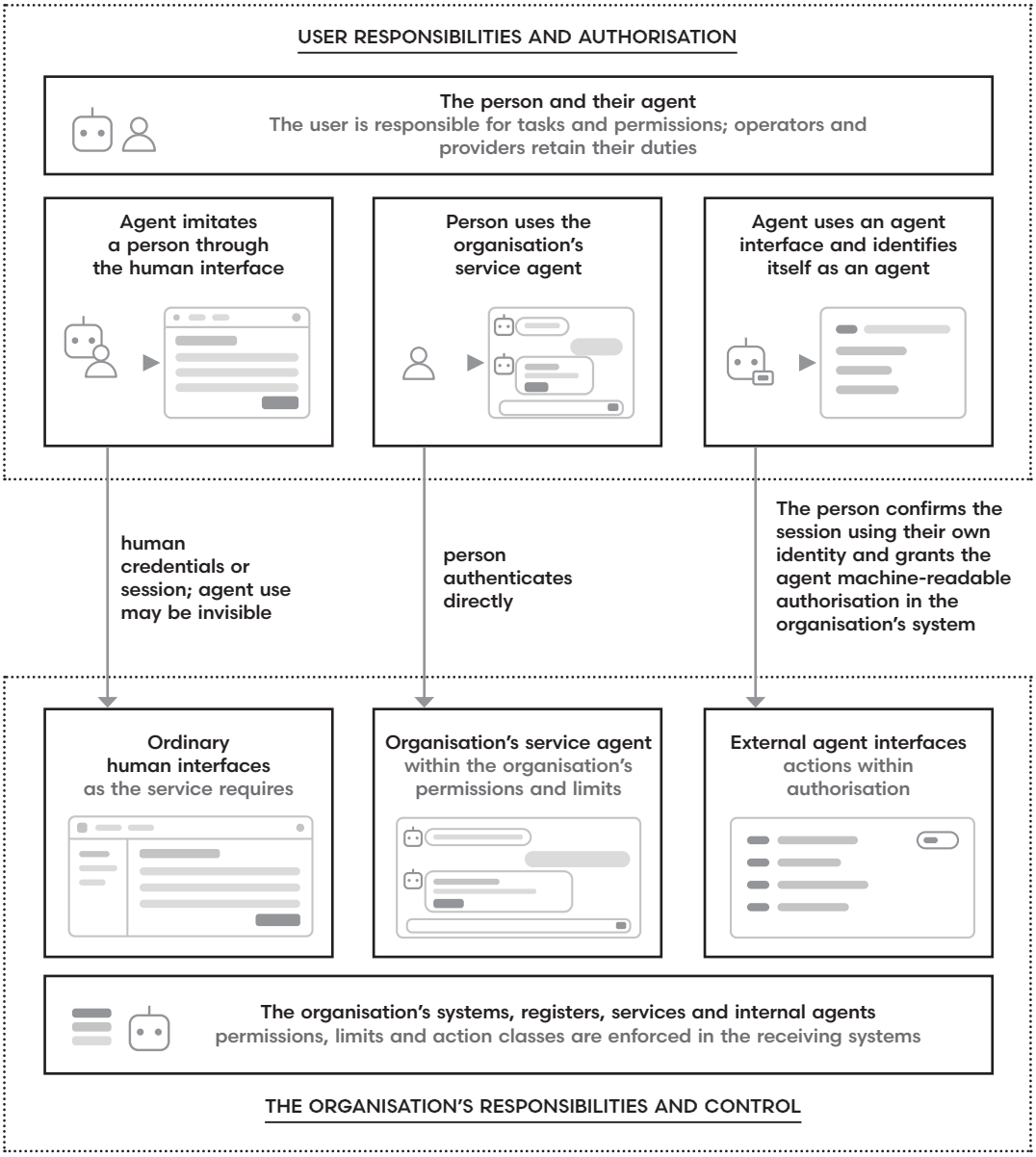
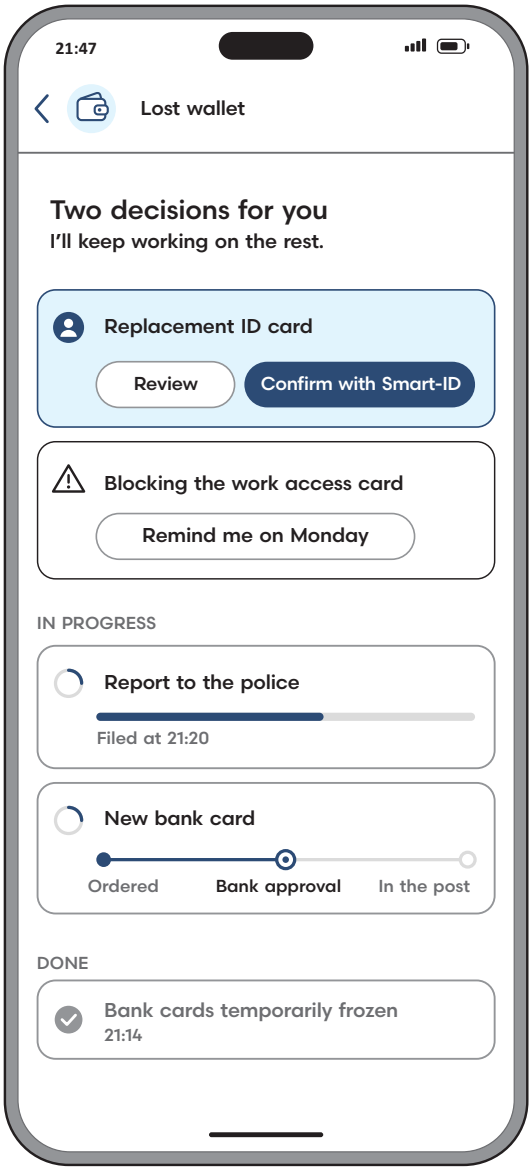


Figure 2. Responsibilities of the user and the organisation, with three routes to a service. The organisation enforces permissions and limits in its systems. The diagram does not allocate all legal liability between users, operators, providers and organisations.

3.1. THE PERSONAL AGENT

A person may already use a commercial assistant or an agent built into their phone and prefer it because it knows their circumstances. They choose which tasks to authorise. In a public service, a personal agent can provide independent help, identify missing information or help prepare an appeal.

In an illustrative scenario, Mari discovers one evening that she has lost her wallet and asks her agent for help. It asks what was inside. Through her bank's agent interface, it requests a temporary freeze on her cards. Mari's existing authorisation covers this, and the bank freezes them. Applying for replacement identity documents falls outside that authorisation, so the agent prepares the application and Mari confirms it herself. At her employer, the agent gets stuck: there is no agent interface and the instructions require contacting a person directly. Finally, it tells Mari what has been done, what remains pending and where she can get help tomorrow. Each organisation has acted within its own authority; Mari has had less to keep track of herself.



It can save time navigating fragmented services, but organisations should still coordinate their own procedures. Lacking a personal agent must not prevent access to essential public services or the exercise of rights. Channels for other features may differ by service where the organisation makes and justifies that choice (see 4.2).

A customer asking a commercial agent to find the best offer may not see how a seller's commission influences the recommendation. Authorisation, enforced limits and action classes do not resolve this conflict when the agent acts within its permissions. Whether operators should owe users a duty to prioritise their interests, and whether particular services require licensing, are questions for the legislature and relevant sectoral rules.

Figure 3: An illustrative example of a general commercial AI agent assisting a person in resolving the case of a lost wallet - this involves interacting with both public and private sector entities and performing actions, some of which require the person's confirmation. This requires the organizations involved having interfaces for agents.

3.2. THREE ROUTES TO A SERVICE

A person or their agent can reach an organisation in three ways, differing in what the organisation can observe and control (Figure 2 and Table 5).

An agent uses the human interface. It operates the person's browser or screen through their logged-in session or credentials. The organisation may be unable to distinguish the agent's actions from the person's or assign it narrower rights. Sharing authentication secrets creates security risks and may breach applicable credential terms. It does not automatically make the user liable for every resulting loss: the facts and relevant law still determine liability. Under EU payment law, for example, recorded authentication alone does not necessarily establish that a disputed payment was authorised.[32] A useful, delegated route gives people a practical reason to avoid credential sharing.

A person uses the organisation's service agent. A bank or public authority offers an agent with permissions and limits that it controls. The agent is more useful when it can complete permitted steps as well as answer questions. It serves the organisation's purpose: a bank's agent offers the bank's products, while a public authority's agent follows its service remit and procedures. Users should know whose interests and authority it represents. EU AI Act Article 50 requires providers of systems intended for direct interaction to ensure that people are informed they are interacting with AI, unless this is obvious in context.[33] This paper additionally recommends a clear statement of whose agent it is and what it cannot do.

An organisation rarely has a reason to recreate a universal commercial assistant. A more practical model is a service-specific agent, such as the concept behind Estonia's Bürokratt public-service assistant programme. It understands a defined service and can perform permitted steps. It may support a coordinated service across organisations where the service owner and each participant's competence have been agreed.

An agent uses a delegated interface. An organisation can expose defined actions through an interface such as a Model Context Protocol (MCP) server. The user authenticates and authorises access on the organisation's own website or application. The protocol supports access; the organisation's rules determine the authority to act.[34] The interface makes permitted operations and their conditions readable by agents and returns receipts that people and agents can understand. During an incident, the organisation can restrict this separately authorised access while keeping the designated human route available.[35] MCP is one implementation option.

---

32 PSD2, Article 72, together with Articles 73–74. Recorded use of a payment instrument is not necessarily sufficient to prove authorisation, fraud, intent or gross negligence. Applicable law and the circumstances determine liability.

33 EU AI Act, Article 50(1). Identifying whose agent it is and explaining its limits are additional recommendations in this paper.

34 Model Context Protocol, 'Authorization'. Authentication and protocol access do not themselves define all application-level permissions.

35 Chan et al., 'Infrastructure for AI Agents', 2025, Section 4.1 (arXiv v3, pp. 11–13). A separate channel does not ensure that every agent uses it or prove that only software operates its credentials.

ROUTE	OPERATIONAL RESPONSIBILITY	ORGANISATION SEES	ORGANISATION CONTROLS	RESPONSE
Agent uses the human interface	User is responsible for tasks and permissions; providers retain their duties; liability depends on facts and law	A human session; agent use may be invisible	Existing user permissions, without reliable agent separation	Offer a useful route with separate, limited permissions
Person uses the organisation's service agent	Organisation runs the service and agent within its remit	Its own agent's actions	Agent access, tools and execution limits	Support useful actions with clear ownership and safeguards
Agent uses a delegated interface	User is responsible for tasks and permissions; organisation checks and executes; operators retain their duties	Delegated session, permissions and recorded actions	Each action against authorisation and the organisation's rules	Begin with reading; add separately authorised actions when needed and tested

Table 5. Three routes to a service: visibility, control and the organisation's response.

3.3. THE EXTERNAL AGENT INTERFACE AS THE PREFERRED ROUTE

Personal agents should reach organisations through a dedicated agent interface and must not receive the person's authentication secrets or signing credentials. A person's authenticated confirmation and an agent's delegated action are different events and should be recorded as such. In the EU, a qualified electronic signature has the legal effect of a handwritten signature.[36] That does not make every authenticated action proof of informed intent. An agent may misunderstand instructions or be redirected by material it reads; using the person's credentials obscures the distinction.

With separate delegated access, the organisation can check each action against its authorisation, refer consequential steps to the person and suspend the agent while preserving the person's designated service channels. A log identifies the channel used, although it cannot prove who physically operated the software. An accountant can, for example, use their own identity and delegated rights to access a client's tax affairs, leaving the client's credentials with the client. The accountant is a person capable of legal representation; software is not. An agent's authorisation here means a technical permission verifiably and inseparably bound to an accountable person or organisation, not a new legal personality or a universal power of attorney.

The receiving organisation need not inspect the agent's internal reasoning to enforce a boundary. It needs to establish who authorised access and check that the authorisation is still valid, the person has the relevant rights and the requested action fits the permission. It also needs to establish whether it can execute the action now and how the result can be remedied or challenged.[37] Strong identity verification answers only part of this. Checking that the agent understood the user's goal remains the

36 eIDAS Regulation, Article 25(2).

37 South et al., 'Authenticated Delegation and Authorized AI Agents', 2025, Sections 2.1 and 3.3–3.4 (pp. 2–3 and 6–8). The operational questions in the text are this paper's synthesis.

responsibility of the user and operator within their respective roles; the receiving interface cannot establish it on its own.

Agent interfaces should support useful actions as well as reading. Organisations should begin with read access and add each write permission only when it is needed and tested.[38] An action checked against a separate authorisation gives the organisation more control than the same action performed through an indistinguishable human session. If useful operations are missing, users may resort to routes that the organisation cannot distinguish from human activity.

A month-long task, such as paying incoming invoices, may use one authorisation across several sessions. Later sessions may use a different product or model version. Agent and session identifiers help connect actions and investigate incidents, but neither identification nor registration establishes authority to act.[39] Logs should identify the agent and session and record the operator and actions performed, with appropriate access controls and retention periods.

At a bank, authorisation specifies permitted recipients, amounts and information access. At a public authority, it may identify a case or application type, distinguish preparation from submission and define accessible data. It should also state whether subdelegation is allowed. The organisation needs structured, unambiguous limits because a free-text instruction interpreted only by the agent cannot serve as an enforceable permission.

An action can proceed without another confirmation where it fits the authorisation, applicable law permits it and its class does not require a person. Other steps go to the person in the organisation's own website or application, with enough information to understand and refuse them. Before opening the interface, the organisation should define how it handles duplicate requests, expired permissions and revocation during an action already in progress. Revocation prevents further authorised use; it does not by itself undo completed actions.

Granting permissions creates another opportunity for fraud: a person may be deceived into authorising an attacker's agent. A genuine confirmation screen provided by the organisation does not prevent a deceived person from consenting. Narrow defaults, notification through a channel the agent does not control, a delay before expanded powers take effect and rapid revocation can reduce the risk. The interface guidance in Section 5.2 and the organisation's security procedures should cover these cases.

Agents may ask other agents for information or delegate subtasks. The receiving organisation still checks each action against the authority of the person or organisation on whose behalf it is requested. Subdelegation cannot enlarge that authority. A message from another agent is input to the task, not permission to replace the user's instructions or the organisation's rules.

---

38 Model Context Protocol, 'Security Best Practices', 'Scope Minimization'.

39 Chan et al., 'Infrastructure for AI Agents', 2025, Sections 3.1, 3.3 and 8 (arXiv v3, pp. 6–8, 10–11 and 21). Identification alone does not determine legal liability; the need for registries remains a research question.

EU payments provide a partial precedent. PSD2 and its technical rules enable regulated third-party account-information and payment-initiation services with customer consent, alongside strong customer authentication requirements and applicable exemptions.[40] Several banks offer read-only interfaces for customer agents, including LHV in Estonia and First Internet Bank in the United States; some offerings remain in beta or limited pilots.[41] A bank can also enable agent payments within a bounded authorisation where applicable authentication, authorisation and consumer-protection rules permit it. Before execution, its system checks the permission and payment conditions and obtains any legally required customer confirmation in the bank's own system. The customer's authentication credentials remain with the customer. In 2026, card networks also reported live European agent purchases under defined permissions, including a Santander–Mastercard pilot and Visa transactions with participating merchants.[42]



40 PSD2, Articles 66, 67 and 97; liability provisions include Articles 73, 74 and 90. Delegated Regulation (EU) 2018/389, Articles 30–33; relevant authentication exemptions include Articles 10a, 13 and 14. Applicability depends on the payment and service involved.

41 LHV Pank, launch announcement, 15 July 2026, and lhv.ai; First Internet Bank, MCP connection FAQ. LHV's public beta is read-only; access conditions differ by bank. Checked 20 October 2026.

42 Mastercard and Banco Santander, press release, 2 March 2026; Visa, agentic-commerce announcement, 2 July 2026. Provider-reported deployments and pilots.

4. WHAT THE ORGANISATION IS RESPONSIBLE FOR

The organisation receiving an action is responsible for ensuring that its systems execute only permitted and controlled actions. Within its lawful authority, it defines how permissions are checked and action classes assigned, specifies who gives any required approval and sets the applicable budgets. It also determines the data and tools available to its internal agents and enforces their limits outside the model. These controls support the organisation's existing service responsibilities; they do not transfer those responsibilities to software.

4.1. INTERNAL AGENTS

An internal agent works through an organisation's systems and may perform some of the same steps as an employee. It can have access to the organisation's information unavailable to a personal agent, under permissions set for its task. The European Banking Authority reported in September 2025 that 92% of EU banks were using AI, and that more than half of surveyed banks used general-purpose AI models or AI agents in customer-facing processes.[43] These figures describe adoption, not the proportion of work performed autonomously. Internal agents still need action classes, sampling and staff tools that meet the service's continuity requirements.

An agent normally performs low-consequence actions itself, within enforced limits. It prepares actions with serious but remediable consequences for human approval; separation-of-duties requirements may call for a second person. It does not perform actions with serious, irreversible consequences or steps reserved to people. For each action type, the organisation records how it can be reversed or its consequences remedied, who does this and who bears the cost of correction, subject to applicable liability rules.[44] Organisations should reduce unnecessary approvals while making every required approval a substantive decision: the reviewer sees the source evidence and can refuse the action.

Before an agent receives permission to act independently, it can work in parallel on the same cases as staff. Staff continue the live service while the agent produces a separate result that neither reaches the customer nor changes the case. An experienced reviewer resolves disagreements without assuming that the human result is always correct. The trial establishes evidence about the task before live autonomy, but duplicates work, so the service owner should limit it by duration or case count. Real data remain subject to the same information and data-protection restrictions during the trial.

An agent may later remain beside an official as a checker, flagging inconsistencies between a draft decision, applicable rules, previous decisions and source information. Its findings can help identify mistakes but are themselves interpretations. They do not enforce a boundary or replace the official's decision.

---

43 European Banking Authority, 'Rising application of AI in EU banking and payments sector', 25 September 2025, pp. 1–2. Adoption figures do not measure autonomous action volumes or effectiveness.

44 Chan et al., 'Infrastructure for AI Agents', 2025, Section 5.2 (arXiv v3, p. 19). Remedy arrangements can shift correction costs and weaken incentives to prevent errors.

Each of the organisation's AI agents should have a named person responsible for enforcing its limits and able to stop it immediately. Stopping it must not require a justification or prior approval. The person needs an accessible technical control and a procedure for handing work over. The amended AI Act also requires providers and deployers to support the development of AI literacy among relevant staff.[45]

At a public authority, an internal agent might draft assessments with references to both an application and the scheme's rules. An official makes and signs the decision. If officials approve drafts faster than they could read them, average review time and the share of substantially changed drafts fall. Both can also fall because the agent improves or cases become simpler. The head of the authority should therefore review these indicators regularly alongside sampled reviews, case mix and complaints.

Where a system evaluates natural persons' eligibility for essential public assistance benefits, its use falls within the AI Act's Annex III framework, subject to the Act's classification provisions.[46] Under the amended timetable, relevant high-risk deployer obligations generally apply from 2 December 2027, subject to transition provisions. They include assigning competent people to oversight, monitoring operation, keeping logs and informing affected people where required.[47] Comparing agent outputs with source evidence during sampling helps staff develop this competence, alongside training in permissions, escalation and continuity. An organisation can also be a provider if it develops or commissions a system and puts it into service under its own name. For high-risk systems, rebranding, substantial modification or a change of intended purpose can also bring provider obligations under Article 25.[48] These can include risk management, technical documentation and conformity assessment.[49] The organisation should establish its role and applicable duties before deployment, including when it uses a purchased model.

An operator may not see harm that its agent causes at another organisation.[50] For incidents crossing organisational boundaries, service owners should agree whom to notify, what information may lawfully be shared and who is responsible for seeing the remedy through. These arrangements should form part of existing incident management.

4.2. PARITY BETWEEN HUMAN AND AGENT CHANNELS

An organisation needs to know which customer and staff needs agents can meet and how much of the service this covers. This paper proposes service parity as a measure

---

45 EU AI Act, Article 4 as amended by Regulation (EU) 2026/1744: measures to support the development of appropriate AI literacy among relevant staff.

46 EU AI Act, Article 6(2)–(3) and Annex III point 5(a). Defined procedural or preparatory uses may fall outside the high-risk classification where the statutory conditions are met; profiling natural persons remains high-risk. Calling an output a 'draft' does not settle classification.

47 EU AI Act, Article 26, read with amended Article 113 and Article 111 transition provisions. Specific duties and exceptions depend on the system and use.

48 EU AI Act, Article 3(3)–(4) and Article 25(1). Provider status depends on the system, its intended purpose and the conditions of development, deployment or modification.

49 EU AI Act, Article 16; provider status following changes is addressed in Article 25.

50 Chan et al., 'Infrastructure for AI Agents', 2025, Section 5.1 (arXiv v3, pp. 17–19).

of whether the same need can be met through different channels and which activities have deliberately been reserved to people. Parity does not require identical speed or every function in every channel. Table 6 illustrates the comparison in a fictional public authority.

SERVICE NEED	HUMAN INTERFACE (CUSTOMER OR STAFF)	EXTERNAL AGENT INTER-FACE	INTERNAL AGENT	HUMAN DECISION
Read scheme conditions	Yes: self-service	Yes: read	Yes: assist	Not needed
Prepare and submit an application	Yes: self-service	Yes, within applicant's permission	No	Not needed
Authority requests a missing attachment	Yes: official in case system	No	Yes: routine request within limits	Not needed for this routine step
Applicant supplies the requested attachment	Yes: self-service	Yes, within applicant's permission	No	Not needed
Assess the substance	Yes: official in case system	No	Draft with sources	Official reviews
Make the discretionary decision and issue the administrative act	Yes: official in case system	No	No: reserved step	Official decides

Table 6. Service parity in an illustrative public authority. "No" means a deliberate prohibition. "Not needed" means no additional human decision for that step; authorisation and any legally required confirmation still apply. The authority's request for an attachment and the applicant's response are separate steps.

Human access shows which service needs can be met without a personal agent (Figure 4). The organisation separately assesses how staff continue providing the service when its own agents stop. A customer-facing form alone does not establish that capability.

The service owner sets channel coverage and continuity or recovery targets for each service. Essential and convenience services may have different targets, reported separately; lower consequences may justify more agent autonomy in a convenience service. A service or function may be available only through an agent or only through a human interface when this is a deliberate, reasoned choice within applicable law. The organisation assesses whose access is restricted and who bears additional costs. Essential services and the exercise of rights, such as challenging a decision or accessing one's own data, must remain practically accessible. Temporary loss of an agent-only function may be acceptable where its consequences have been assessed and a recovery deadline set.

Agent coverage shows which service needs agents can fulfil, including whether an applicant's agent can submit an application or only prepare it. When setting limits, organisations should account for the possibility that agents will use human-facing interfaces to imitate users if an action is unavailable through the agent interface. Agent interfaces should begin with read-only access and expand gradually. Preparing

a human decision counts as agent support, but should be reported separately from autonomous completion. Coverage of external interfaces and internal agents should also be shown separately.

The last column in Table 6 identifies decisions that remain with people and the responsible role, such as the official making a discretionary decision. For each service need, the organisation should decide which channels serve it and whether a human decision is required. Both "Not needed" and a reasoned prohibition count as decisions; an unexamined need is a gap. Each new function requires this assessment, whichever channel first offers it. Decision coverage records whether these choices have been made, not whether the selected channels work in practice.

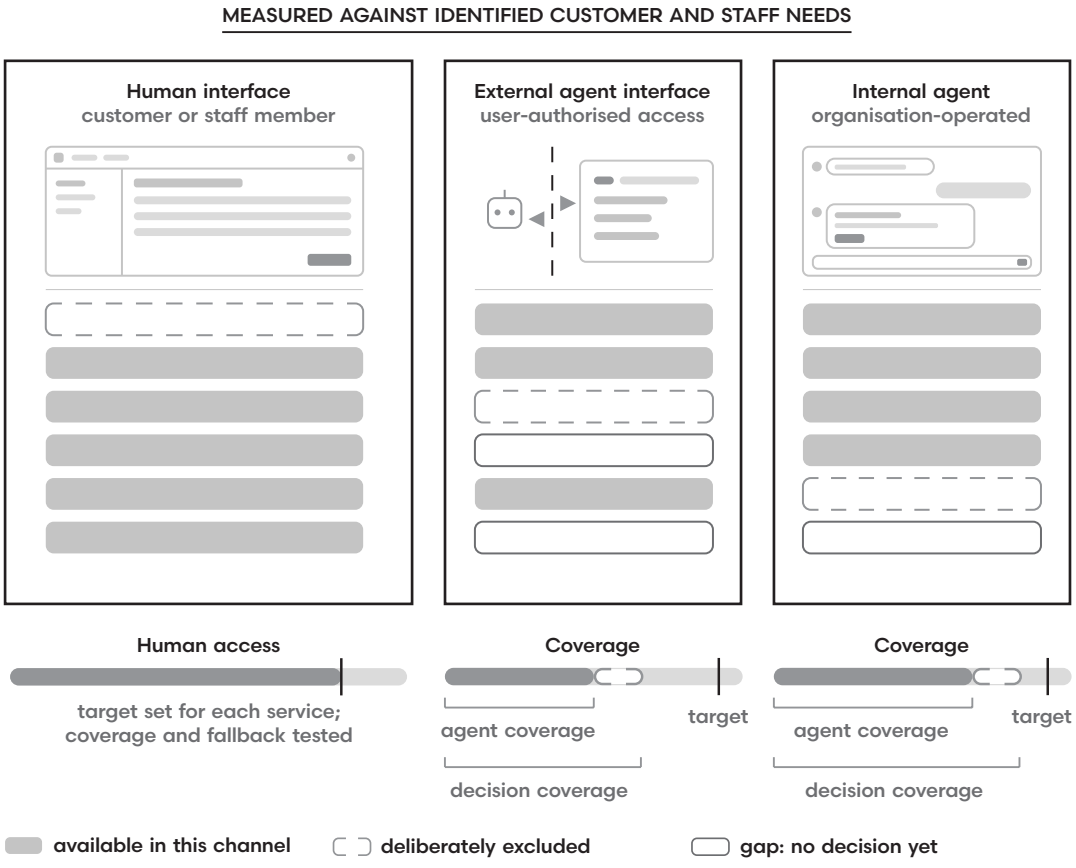


Figure 4. Channel coverage is measured against customer and staff needs and compared with service-specific targets. Access without a personal agent and staff continuity without the organisation's own agents are separate questions.

Parity helps reveal gaps in access, but service improvement must be judged by whether the person's need is resolved. A coordinated service may remove several applications or repeated requests for evidence without adding agent actions. The service owner should measure time to resolution, the person's active administrative effort and total work across the participating organisations, including review and correction. A step removed through simplification is not a coverage gap.

Goodhart's law warns that a measure can become less informative when used as a target.[51] Organisations may add agent functions that offer little value, or label difficult needs as prohibited to improve decision coverage. A prohibition does not increase agent coverage, which is why the two measures are separate. Service needs should be the primary unit of measurement. Every prohibition should have a recorded reason, and management should review prohibitions and targets annually against actual access, waiting times and outages. Organisations may also weight needs by volume to show how much actual traffic agents handle. A simple count remains useful because high volumes on a few easy tasks cannot conceal gaps elsewhere.

4.3. FIVE PRINCIPLES FOR EVERY AGENT-ENABLED SERVICE

Each principle below needs a responsible owner within the organisation, evidence that it works and a response when it does not.

1. Preserve service in the languages people are entitled to use. Language can affect model quality and token costs, but the difference varies by model and task.[52] The service owner should test performance in the relevant languages and budget for it. In Estonia this includes preserving service in Estonian. Elsewhere the same principle follows the jurisdiction's language obligations and users' needs. Poor performance should narrow the agent's role or trigger human assistance, rather than exclude users from the service.

2. Enforce information restrictions where data leave the organisation. Information classification, legal basis and purpose determine what may be sent to an external model. Structured records may remain useful after direct identifiers have been replaced, but pseudonymisation does not make personal data anonymous or remove data-protection duties. Free text can contain identifying details that need additional review. Removing a name does not authorise a disclosure otherwise prohibited. The information owner and security function should define the permitted flows and block the rest.

3. Meet continuity requirements when agents stop. The service owner defines which work must continue without agents, how long other functions may be unavailable and how they will be restored. The organisation agrees how much work staff can complete without agents and the longest acceptable wait for a critical case, then tests the fallback before launch and regularly thereafter. A provider outage or price change must not stop an essential service. Bainbridge's account of automation explains the difficulty of retaining rarely used skills;[53] new staff may never learn preparation that agents always perform. Where continuity depends on staff taking over, they should regularly complete some cases without agent-prepared material. Maintaining this capability consumes staff time and reduces the agent's net benefit; the organisation must budget for it.

---

51 Goodhart, 'Problems of Monetary Management', 1975; the widely used formulation is discussed by Strathern, 'Improving ratings', 1997, p. 308.

52 Petrov et al., 'Language Model Tokenizers Introduce Unfairness Between Languages', 2023. Test actual service costs and language quality with the intended model.

53 Bainbridge, 'Ironies of automation', 1983, pp. 775–776.

4. Make provider replacement practicable. Every procurement involving a model provider should provide for replacing it, with the time and cost known and the change tested in practice. Duplicating every feature may be too expensive; identify essential capabilities and test their transfer. EU financial institutions have specific ICT continuity and exit obligations under DORA; other organisations need requirements suited to their legal setting.[54] With a new provider or model version, the organisation starts with narrower permissions and measures the error rate again. If an acceptable exit cannot be demonstrated, it should limit the dependency.

5. Justify the total cost by the service benefit. Measure cost and time per completed case, including review, correction, integration and continuity. Compare these results with those from rules-based automation and human work, taking account of better evidence, fewer defects and shorter waits. If checking costs more than doing the task and does not improve the outcome, the task is poorly chosen. As review takes a larger share of the work, narrow the task, convert stable steps into rules or stop using the agent. Include work transferred to applicants, staff and other organisations.

54 DORA, Articles 11, 28 and 30. These are financial-sector obligations, used here as an example rather than a rule for every public body.

5. WHAT GOVERNMENTS NEED TO PROVIDE

Governments can help immediately by simplifying services, clarifying how existing infrastructure can support delegated action and giving people visibility over what is done on their behalf. Shared authorisation across organisations is a further step to consider when repeated needs justify it. Countries should begin with the functions a service requires, taking account of their own legal and technical starting point.

5.1. AGENTS DO NOT REPLACE THE NEED TO IMPROVE ADMINISTRATION EFFICIENCY

An agent can help people navigate fragmented services by remembering deadlines and entering the same information in several places. This may hide fragmentation from organisations while moving its cost to whoever pays for the agent. Access to essential public services and the exercise of rights with reasonable effort should not depend on a personal agent's availability or capability. Organisations should reduce duplicate work and coordinate their procedures; AI must not become a reason to preserve poor administration.

Sometimes a service can dispense with the application altogether. Where the law defines entitlement clearly and the necessary information is already available, a rules-based process can initiate a proactive service. In Estonia, registration of a child's birth can trigger a family-benefit offer from the Social Insurance Board; the parent confirms it in self-service, subject to the relevant circumstances.[55] This removes an application that an agent might otherwise prepare. Before assigning application reading to an agent, the service owner should ask whether the application is needed at all.

In Estonia, support for everyday assistive devices and medical devices has been divided between the social insurance and health insurance systems. Orthopaedic shoes and individual insoles can fall under different arrangements, while workplace aids involve the employment service. For a person, these may be parts of one need. In June 2026, the Ministry of Social Affairs announced plans to bring assistive and medical devices into a unified health-insurance arrangement during 2027–2029.[56] This is an announced reform, not an already completed service. An agent filling in separate applications would make the current division easier to navigate without resolving it.

Cheaper drafting may also increase demand on public bodies. Some additional applications, information requests and appeals will represent legitimate claims that people previously lacked the time or confidence to make; others may be repetitive or poorly founded. A 2026 study collected 84 potential cases across 11 jurisdictions in which officials or other sources attributed increased demand to AI use. Its method does not establish how widespread the effect is or whether AI caused the increase.[57] Organisations should measure their own workload before assuming the cause. Simpler services and an agent interface that accepts structured requests under verified authorisation can

---

55 Estonian Social Insurance Board, 'Millised perehüvitised kaasnevad lapse sünniga?', 27 January 2026, 'Pakkumus – mis see on?'. The parent confirms the offer in self-service; eligibility and family circumstances still matter.

56 Estonian Ministry of Social Affairs, 'Abivahendid ja meditsiiniseadmed liidetakse ühtseks süsteemiks', 17 June 2026; Estonian Unemployment Insurance Fund, 'Tööalane abivahend'. The 2027–2029 consolidation is an announced plan.

57 Schmitz, Hammond and Chan, 'Characterizing Agentic Flooding of Government Services', arXiv:2608.16603v2, 2026, abstract and Sections 1, 3.2 and 4. The authors identify potential cases; their method does not support causal attribution or a quantitative estimate of the aggregate effect.

reduce avoidable work. Intake checks must follow the applicable procedure. Technical checks must not become an unlawful barrier to submitting a claim, correcting a defect or exercising a right. Fees and other barriers can also deter legitimate claims. Access to rights must not depend on whether a person has an agent or how capable it is.

When a person's need crosses organisational boundaries, the responsible ministry and participating bodies, including local government where relevant, should agree who coordinates the whole service and where the person can turn (Figure 5). The European Interoperability Framework, the UK Service Standard[58] and Estonia's life-event services support this approach.[59] The arrangement needs agreed powers, funding and escalation routes. The service owner remains responsible for coordination and each organisation retains its decision-making powers, including when an agent connects the steps. Information should be reused where lawful, and people and participating organisations should see who is responsible for the next step and where a case is waiting. The service owner should measure resolution through one contact, the person's remaining administrative effort and total work across the participating organisations. Integration has costs[60] and some overlap may serve a purpose, so restructuring should follow an assessment of the actual problem.[61]

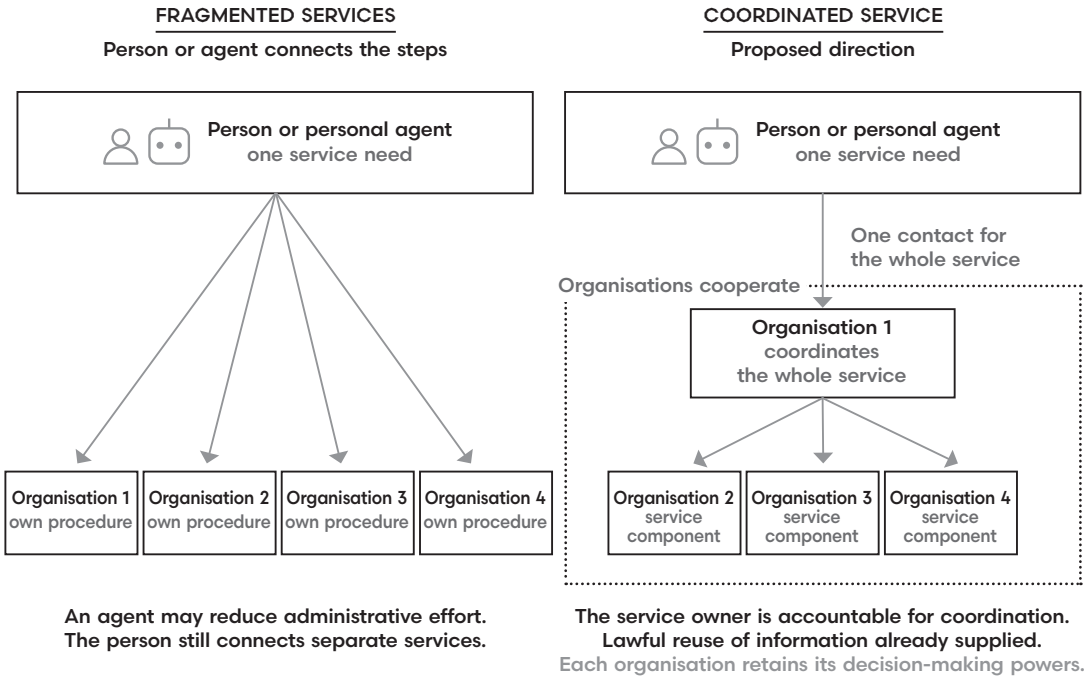


Figure 5. Fragmented and coordinated arrangements for a defined public-service need. The service owner is accountable for coordinating the whole service; each organisation retains its decision-making competence.

---

58 European Commission, New European Interoperability Framework, 2017, Recommendations 11 and 25 and Section 4.3.1; GOV.UK, Service Standard, point 2, 'Solve a whole problem for users'.

59 RIA, 'Sündmusteenused' (life-event services); European Interoperability Framework, Recommendation 29.

60 OECD, Integrating Social Services for Vulnerable Groups, 2015, pp. 51–52 and 197–208.

61 US Government Accountability Office, Fragmentation, Overlap, and Duplication: An Evaluation and Management Guide, GAO-15-49SP, 2015.

5.2. ADAPT EACH ORGANISATION'S SYSTEMS TO EXISTING DIGITAL FOUNDATIONS

Estonia already has electronic identity, digital signatures, registers containing representation rights and a distributed data exchange layer. EU law gives qualified electronic signatures the legal effect of handwritten signatures.[62] Other countries may provide the necessary functions differently. For a selected service, the organisation needs reliable authentication, a way to check the user's authority, lawful access to information and revocable permissions enforced at execution. It does not need to reproduce Estonia's entire infrastructure. Its access-management system must distinguish the person on whose behalf an action is taken from the agent session performing it, so that the agent receives separate, narrower permissions.

LHV's bank required from the state mainly only national digital identity solution to enable the agentic interface. Customers authenticate on the bank's own page; the bank handles consent, permission selection, logging and revocation. Its published beta documentation describes an access token lasting one hour and a refresh token lasting 30 days, with revocation through internet banking.[63] Organisations would otherwise implement these security-sensitive functions repeatedly, although LHV's particular arrangements are not a general rule for public services.

Organisations first need two practical guides, prepared with service owners, security specialists and legal advisers. The external-access guide should explain separate agent credentials, permission scopes, session binding, execution checks, revocation and failure cases. The internal-agent guide should cover work allocation, review, stopping and continuity. Both should use suitable open standards and include tests with representative commercial agents. Shared software components are justified where organisations would otherwise duplicate security-critical work and a common implementation would reduce cost or risk.

Government can demonstrate the guidance through a bounded service on its own portal, with Estonia's eesti.ee as one candidate. The proposed external agent interface would let a person's chosen agent read portal information and perform permitted actions. The portal's own service agent could prepare an application and submit it after the person has confirmed it. When the portal routes a request to another organisation, the receiving service must identify the requesting organisation, verify its authority and the legal basis for the data use, and apply its own legal and procedural requirements. A personal agent's authorisation cannot supply a missing legal basis for processing between organisations.

The pilot should start with reading public information and a person's own records, then enable one action once authorisation checks have been tested. Its service owner should publish the pilot's scope and compare time to resolution, user effort, work required of the organisation and errors with the simpler existing route. If permission

---

62 RIA pages on electronic identity services, Pääsuke representation-rights management and X-tee; eIDAS Regulation, Article 25(2). These establish existing foundations, not universal readiness for agent delegation.

63 LHV Pank, announcement, 15 July 2026; lhv.ai, 'Security and trust' and 'Common questions', checked 20 October 2026. Published token lifetimes and read-only scope describe this beta implementation.

limits or revocation fail, suspend the affected action. Do not expand it if it fails to shorten resolution time or reduce work. Table 7 separates existing foundations from proposed support.

FUNCTION OR SUPPORT	USE AND STATUS IN THE ESTONIAN EXAMPLE
Reliable electronic identity	Existing foundation. The person authenticates and grants permission in the organisation's system; the agent receives separate access credentials.
Digital signature	Existing foundation where law or procedure requires a signature. Human signing credentials remain with the person.
Registers and representation rights	Existing sources for checking whether a person may act for an organisation or another person. The service still checks the particular authority needed.
Lawful data exchange	X-tee connects distributed systems. Reuse depends on a lawful basis and access agreements.
Visibility of data access and actions	Data Tracker already covers data-access records from connected systems. Agent-action records and linked permission controls are a proposed extension.
Legal framework	Existing rules govern identity, signatures, representation, data use and decisions. Each proposed agent use needs a specific applicability assessment.
Guidance and test suite	Proposed shared support for delegated access, enforcement, revocation, continuity and interoperability.
Portal reference service	Proposed bounded pilot demonstrating the guidance, with measured outcomes and stopping conditions.

Table 7. Functions organisations need, with Estonia as an example of a starting point.

5.3. LET PEOPLE SEE WHAT AGENTS DID ON THEIR BEHALF AND WITH THEIR DATA

Delegation makes administration less visible to the person. People need a usable way to inspect actions performed on their behalf and the use of their data, identify the responsible organisation and revoke permissions. Without this visibility and control, their responsibility for the tasks and permissions they give their agent is largely formal.

Estonia's Data Tracker offers a starting point: connected databases can show which organisation accessed a person's information, when and for what purpose.[64] Extending it to agent actions would be new work. For each action, an external agent interface should record on whose behalf it was taken, the authorisation used and the operator through which it was performed. The record should identify which data were accessed or supplied for the action and for what purpose. Organisations should keep equivalent records for their internal agents. A portal could show these records and where to revoke permission. Records should remain at the organisations and be retrieved on the person's request, avoiding a new central history of the services they use. Access, purpose, retention and protection of the combined view still require explicit controls.

64 RIA, 'Andmejälgija' (Data Tracker), checked 17 October 2026. The existing service displays data-access information from connected systems. Agent-action visibility and linked revocation in this paper are proposals.

A log may name a procedure or legal instrument that tells the person little. A portal agent could use the person's records and logs to explain what happened and why, or help them understand an official letter or decision. An agent can explain a log independently, with sampling, where mistakes would have minor, readily correctable consequences. This is a suitable first task for the reference service. An explanation of a decision should link to the decision and its challenge instructions. Where an explanation could affect rights, deadlines or obligations, verified standard text or human review is needed. The explanation does not replace the authoritative documents, and the person must be able to obtain a correction. The service owner should test whether explanations improve understanding before expanding their use.

Frequent questions reveal which letters and procedural steps need improvement. The service owner should use these questions to improve the underlying service so that people can understand it without relying on an agent.

5.4. SHARED AUTHORISATION WHEN REPEATED NEEDS JUSTIFY IT

An agent can be useful precisely because it acts while its user is away. Within one organisation, a permission can already specify an agreed scope and duration, subject to applicable law and security requirements. Shared authorisation becomes relevant when several organisations need to interpret and accept the same permission. Organisations can agree early on what shared permissions mean while developing useful local services without waiting for a universal system.

The developing Agent Payments Protocol (AP2) describes advance authorisation linked to a cryptographic key, with transaction checks against the agreed limits.[65] A key demonstrates control of a credential; it does not by itself establish the agent's legal identity, trustworthiness or authority for every service. The European Digital Identity Wallet framework places the wallet under the user's sole control.[66] This supports keeping human credentials separate from delegated access, but does not itself prescribe one architecture for all agent actions. This paper proposes a separate, scoped credential for delegated agent actions. It must be accepted only by interfaces designated for agent access, not as a credential for human-facing sessions. Revocation needs both a way to withdraw permission and a way for every receiving system to learn and enforce that change within a defined time.[67]

Nationally recognised agent identifiers, shared permissions and certification may become useful as activity across organisations grows. They should be evaluated against a concrete problem before being treated as prerequisites for deployment. Even when an agent is recognised, each action still needs to be checked. Participating organisations must agree the meaning of a permission: for example, whether a

---

65 Agent Payments Protocol (AP2), specification v0.2 and 'Agent Authorization Framework', checked 20 October 2026; Google, AP2 contribution to the FIDO Alliance, 28 April 2026. Standardisation is continuing. This payment protocol is an architectural example, not generally applicable law or proof of interoperability across public services.

66 eIDAS as amended by the European Digital Identity Framework Regulation, Article 5a(4)(a). The wallet is under the user's sole control; the proposed separation of agent credentials is this paper's design recommendation.

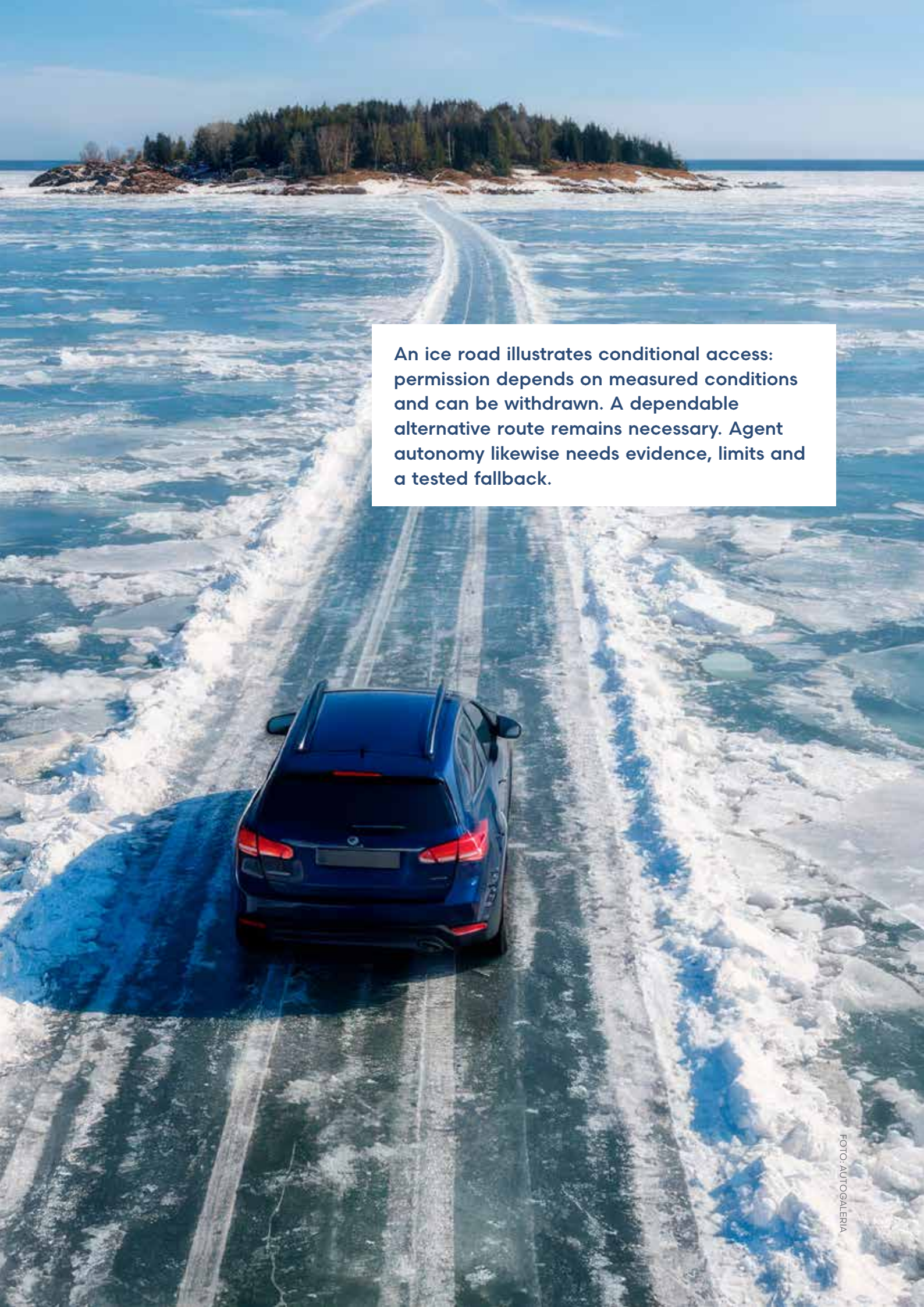
67 South et al., 'Authenticated Delegation and Authorized AI Agents', 2025, Section 3.3 (p. 7). Effective revocation additionally depends on timely enforcement by receiving services.

monthly €500 cap includes fees, how concurrent actions consume it and what happens when a permission is revoked. A shared solution is worth testing where many services require similar permissions, users repeatedly grant the same scope and one coordinated service cannot remove the repetition.

Guidance can reduce future integration work by defining consistent names for reading, preparation, submission and payment permissions, together with expiry, revocation and logging fields. Sector-specific arrangements may be easier to test than a universal mandate. A pilot should compare separate permissions granted in each organisation with the shared arrangement, measuring user effort, implementation and operating cost, enforcement failures and revocation delay. Organisations should retain the simpler approach if the shared system adds cost or ambiguity without sufficient benefit.

POSSIBLE COMPONENT	PROBLEM IT COULD ADDRESS	WHEN TO INVESTIGATE
Common permission format	Several organisations interpret the same scope consistently	Repeated uses exist; scope, aggregate limits and remedies can be defined without ambiguity
Shared authorisation service	People manage and revoke permissions in one place	Separate permissions impose significant effort; revocation reaches every recipient; no central activity history is collected; privacy and availability safeguards are tested
Permission credential bound to a key	Delegated actions across organisations while the user is absent	Several issuers and recipients need it; the credential works only at agent interfaces; scope, expiry and revocation are enforced; a key does not confer legal authority

Table 8. Possible components of shared authorisation and the evidence needed before developing them.



An ice road illustrates conditional access: permission depends on measured conditions and can be withdrawn. A dependable alternative route remains necessary. Agent autonomy likewise needs evidence, limits and a tested fallback.

FOTO: AUTOGALERIA

6. SCENARIOS AND SERVICE CONTINUITY IN A KRATT NATION

In a Kratt Nation, rules, authorisations and interfaces are readable by agents, whose limits are enforced outside the model. Service owners set channel coverage and continuity or recovery requirements for each service. Essential services remain practically accessible without a personal agent and when the organisation's own agents stop. Two fictional scenarios show a grant authority using the process in Table 9.

STEP	WHO PERFORMS IT	CLASS	CONTROL
Check agent access and permission	Rule in receiving system	Not applicable	Reject unauthorised machine actions; preserve a lawful route for the person to submit
Record receipt; flag defects or eligibility issues	Case system; staff where judgement is required	Not applicable	Do not treat a technical gate as the legal decision to refuse receipt; allow correction where required
Register a submission where interpretation is needed	Agent	REVIEW	Sample official records; use a rule if registration is fully specified
Request a missing attachment	Agent within narrow scope	ACT	Sample text; no new procedural requirement, changed deadline or automatic suspension
Leave an incomplete application unexamined	Official, with agent preparation	APPROVE or HUMAN-ONLY under local law	Official checks evidence, legal basis and remedy and can refuse; human-reserved decisions remain human
Draft an assessment against the scheme's objectives	Internal agent	ACT	Draft cites application and scheme rules; no case effect without review; head regularly reviews reading time, changes and samples
Make the discretionary decision	Official	HUMAN-ONLY	Official examines sources and signs; monitor review quality
Explain the decision	Agent or official according to consequence	ACT for a low-consequence summary	Link decision and challenge instructions; sample free-text summaries; verify rights and deadlines through rules or human review
Manage the whole service	Service owner	Not applicable	Keep the web-form route; meet channel and continuity targets; adjust autonomy thresholds using measured error rates; test fallback and provider change; assess total cost, time and harm

Table 9. An illustrative grant process. Technical checks remain distinct from legal receipt and decisions on an application.

The service owner assesses the whole process using four measures: total cost per completed case including review and correction; time to resolution; errors weighted by consequences; and the volume of work staff can complete without agents. The authority also checks whether applicants retain usable access and effective ways to challenge decisions.

Blue scenario: agents become an ordinary part of the service

A grant authority introduced agents three years ago. During the first three months, an agent worked alongside staff in shadow mode. Experienced reviewers compared its assessments and proposed document requests with the evidence and staff's work. Only after testing the task did the authority allow it to send routine requests for missing attachments within the limits in Table 9. The agent now assembles evidence so an official can decide once the file is ready. Cases that used to take a month are resolved in a week. Staff numbers have stayed the same, but the queue has disappeared.

In the second spring, the authority introduced a new model version with narrower initial permissions and increased sampling. For one small grant scheme, the agent began requesting documents already attached. Some applicants paused while trying to obtain the documents, although the requests had no power to suspend the procedure. Individual requests looked reasonable and initially went unnoticed. During their monthly random review, officials found two such cases. The service owner immediately required staff approval for these requests, investigated the affected case group and had stalled applications restored to processing. After the provider corrected the error and a new task-specific check showed acceptable performance, the authority restored the limited permission.

Red scenario: continuity and sovereignty

A few years later, several major model providers announce one morning that export restrictions have forced them to suspend their services in Europe.[68] The authority loses access to the models on which its agents depend. Applicants can still use the ordinary web form and officials continue processing cases, although the queue grows. The agents had prepared draft assessments three times faster; newer staff take much longer during the first few days. The fallback is familiar: the authority practises it twice a year and staff complete some cases without agent material each month. One provider outside the affected supply chain remains available in Europe. Provider replacement has also been tested, and two weeks later the chief information officer moves the authority's agents to that provider's model. The agents begin with narrower permissions until the authority has measured their performance again. Critical cases are prioritised throughout the disruption.

Putting a kratt to work means knowing how to handle it. In a democratic state, each agent must have an identifiable person or organisation responsible for its use, with the power to enforce its limits within the rule of law. Essential services must continue to be provided the day the agents are switched off.

68 A narrower disruption occurred in June 2026: Anthropic suspended Fable 5 and Mythos 5 for all users on 12 June following a US export-control directive; its 1 July update reported access restored. Its other models remained available. Anthropic, statement of 12 June and 'Redeploying Fable 5', 30 June, updated 1 July 2026.

REFERENCES

EU and Estonian legal references are used as jurisdiction-specific examples.

Legislation

Digital Operational Resilience Act (DORA). **Regulation (EU) 2022/2554**.

eIDAS Regulation. **Regulation (EU) No 910/2014**, as amended.

European Digital Identity Framework. **Regulation (EU) 2024/1183**, amending eIDAS.

General Data Protection Regulation (GDPR). **Regulation (EU) 2016/679**.

Payment Services Directive (PSD2). **Directive (EU) 2015/2366**. Strong customer authentication: **Delegated Regulation (EU) 2018/389**, amended by Delegated Regulation (EU) 2022/2360.

Consumer Credit Directive. **Directive (EU) 2023/2225**.

EU AI Act. **Regulation (EU) 2024/1689**, amended by **Regulation (EU) 2026/1744**.

Product Liability Directive. **Directive (EU) 2024/2853**, with the **2026 corrigendum**.

Estonia. **Administrative Procedure Act**, Sections 14–15, current text accessed 17 October 2026.

Research and other sources

Agent Payments Protocol (AP2). **Specification v0.2** and **Agent Authorization Framework**, ap2-protocol.org, accessed 20 October 2026.

Anthropic. **"Statement on the US government directive to suspend access to Fable 5 and Mythos 5"**, 12 June 2026; **"Redeploying Fable 5"**, 30 June 2026, updated 1 July 2026, anthropic.com.

Article 29 Working Party. **"Guidelines on Automated individual decision-making and Profiling"**, WP251rev.01, 2018; endorsed by the European Data Protection Board.

Autor. **"Polanyi's Paradox and the Shape of Employment Growth"**, NBER Working Paper 20485, 2014, nber.org.

Bainbridge. "Ironies of automation", Automatica 19(6), 1983, pp. 775–779, **doi:10.1016/0005-1098(83)90046-8**.

Bogdanov, Etti, Kamm, Ostrak, Stomakhin, Toomsalu, Valdma and Veldre. "Tehisintellekti ja masinõppe tehnoloogia riskide ja nende leevendamise võimaluste uuring" [study of AI and machine-learning risks and mitigation], Cybernetica for RIA, D-16-377, 27 February 2024, **ria.ee**.

Brednich, Rolf Wilhelm, ed. Enzyklopädie des Märchens, vol. 9. Berlin and New York: de Gruyter, 1999. "Mandragora", cols 115–116.

Calvino, Flavio, Jelmer Reijerink and Lea Samek. **"The Effects of Generative AI on Productivity, Innovation and Entrepreneurship"**, OECD Artificial Intelligence Papers, No. 39, 20 June 2025, doi:10.1787/b21df222-en.

Chan, Wei, Huang, Rajkumar, Perrier, Lazar, Hadfield and Anderljung. "Infrastructure for AI Agents", Transactions on Machine Learning Research, 2025, **arXiv:2501.10114v3** (19 June 2025).

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng and Tobias Salz. **"The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers"**, Management Science, advance online publication, 27 February 2026, doi:10.1287/mnsc.2025.00535.

Dowding and Taylor. "Algorithmic Decision-Making, Agency Costs, and Institution-Based Trust", Philosophy & Technology 37, article 68, 2024, **doi:10.1007/s13347-024-00757-5**.

Duden. "Schrat", meaning and etymology, duden.de, accessed 18 October 2026.

Einfeldt, Rand. "Puck and Folklore", Early Modern England Encyclopedia, University of Victoria, 2026, lemdo.uvic.ca.

Eloundou, Manning, Mishkin and Rock. "GPTs are GPTs: Labor market impact potential of LLMs", Science 384(6702), 2024, pp. 1306–1308, **doi:10.1126/science.adj0998**.

Estonian Information System Authority (RIA). **"Andmejälgija"** [Data Tracker], accessed 17 October 2026.

Estonian Information System Authority (RIA). **"Andmevahetuskiht X-tee"** [X-tee data exchange layer].

Estonian Information System Authority (RIA). **"Elektroonilise identiteedi teenused"** [electronic identity services].

Estonian Information System Authority (RIA). **"Introduction of X-tee"**, ria.ee.

Estonian Information System Authority (RIA). **"Keskne volituste haldamise infosüsteem Pääsuke"** [central management of representation rights].

Estonian Information System Authority (RIA). **"Sündmusteenused"** [life-event services].

Estonian Ministry of Social Affairs. **"Abivahendid ja meditsiiniseadmed liidetakse ühtseks süsteemiks"** [planned consolidation of assistive and medical devices], 17 June 2026.

Estonian Social Insurance Board. **"Millised perehüvitised kaasnevad lapse sünniga?"** [family benefits following the birth of a child], 27 January 2026, accessed 20 October 2026.

Estonian Unemployment Insurance Fund. **"Tööalane abivahend"** [workplace assistive device].

European Banking Authority. **"Rising application of AI in EU banking and payments sector"**, 25 September 2025, eba.europa.eu.

European Commission. **"New European Interoperability Framework"**, 2017, ec.europa.eu.

Fehr, Gächter and Kirchsteiger. "Reciprocity as a Contract Enforcement Device: Experimental Evidence", Econometrica 65(4), 1997, pp. 833–860**, doi:10.2307/2171941**.

Feng and Chandra. "New development: Principal–agent issues when governments embrace AI agents", Public Money & Management, 2026, **doi:10.1080/09540962.2026.2642814**.

First Internet Bank. **"How does First Internet Bank's Model Context Protocol (MCP) connection work with Claude and ChatGPT?"**, FAQ, firstib.com, accessed 20 October 2026.

Fischli, Franklin, Manzini and Gabriel. "Agents, Alignment, and the Many Faces of Autonomy", Minds and Machines 36, article 34, 2026, **doi:10.1007/s11023-026-09786-9**.

Gabriel. "Artificial Intelligence, Values, and Alignment", Minds and Machines 30(3), 2020, pp. 411–437, **doi:10.1007/s11023-020-09539-2**.

Gailmard. ''Accountability and Principal–Agent Theory'', in Bovens, Goodin and Schillemans (eds.), The Oxford Handbook of Public Accountability, Oxford University Press, 2014, **doi:10.1093/oxfordhb/9780199641253.013.0016**.

GAO. **''Fragmentation, Overlap, and Duplication: An Evaluation and Management Guide''**, GAO-15-49SP, 2015, gao.gov.

Giacomini and Grandi. ''New development: I chose to let the machine choose—the rise of 'AIcratism''', Public Money & Management 46(3), 2026, pp. 387–391, **doi:10.1080/09540962.2025.2570740**.

Goodhart. ''Problems of Monetary Management: The U.K. Experience'', Papers in Monetary Economics, Vol. I, Sydney: Reserve Bank of Australia, 1975, pp. 1–20.

Google. **''We're donating Agent Payments Protocol to the FIDO Alliance to support the future of secure, agentic payments''**, 28 April 2026. blog.google.

GOV.UK. **''Solve a whole problem for users''**, Service Standard, point 2, updated 29 January 2026, gov.uk.

Holmström. ''Moral Hazard and Observability'', The Bell Journal of Economics 10(1), 1979, pp. 74–91, **doi:10.2307/3003320**.

Institute for Language and Folklore (ISOF). Realkatalogen, Uppsala, 2021, p. 163, ''Bjära, mjölkhare'', isof.se.

Institute of the Estonian Language. Etymological Dictionary, **entry ''kratt''**, eki.ee.

Institute of the Polish Language, Polish Academy of Sciences. Wielki słownik języka polskiego PAN, entry ''skrzat'', meaning and etymology, wsjp.pl, accessed 18 October 2026.

Jensen and Meckling. ''Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure'', Journal of Financial Economics 3(4), 1976, pp. 305–360, **doi:10.1016/0304-405X(76)90026-X**.

Jevons. The Coal Question, London: Macmillan, 1865 (2nd edition 1866), Chapter VII, ''Of the Economy of Fuel''.

Kim. ''Deep learning and principal–agent problems of algorithmic governance: The new materialism perspective'', Technology in Society 63, article 101378, 2020, **doi:10.1016/j.techsoc.2020.101378**.

Kivirähk, Andrus. Rehepapp ehk november. Tallinn: Varrak, 2000.

Laitinen and Sahlgren. ''AI Systems and Respect for Human Autonomy'', Frontiers in Artificial Intelligence 4, article 705164, 2021, **doi:10.3389/frai.2021.705164**.

LHV Pank. **lhv.ai**, agent-interface product documentation, ''Security and trust'' and ''Common questions'', accessed 20 October 2026.

LHV Pank. **''Pangandus liigub AI-assistentidesse: LHV loob klientidele uue võimaluse''** [banking through AI assistants], 15 July 2026.

Mastercard and Banco Santander. **''Santander and Mastercard complete Europe's first live end-to-end payment executed by an AI agent''**, press release, 2 March 2026, mastercard.com.

McCubbins, Noll and Weingast. ''Administrative Procedures as Instruments of Political Control'', Journal of Law, Economics, and Organization 3(2), 1987, pp. 243–277**, doi:10.1093/oxfordjournals.jleo.a036930**.

Miller and Whitford. ''Trust and Incentives in Principal-Agent Negotiations: The 'Insurance/Incentive Trade-Off''', Journal of Theoretical Politics 14(2), 2002, pp. 231–267, **doi:10.1177/095169280201400204**.

Model Context Protocol. Specification, **''Authorization''** and **''Security Best Practices''**, version 2026-07-28.

Noy and Zhang. ''Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence'', Science 381(6654), 2023, pp. 187–192, **doi:10.1126/science.adh2586**.

OECD. **''Integrating Social Services for Vulnerable Groups: Bridging Sectors for Better Service Delivery''**, 2015, doi:10.1787/9789264233775-en.

OWASP. **''LLM01:2025 Prompt Injection''**, OWASP Top 10 for LLM Applications 2025, genai.owasp.org.

Petrov, La Malfa, Torr and Bibi. ''Language Model Tokenizers Introduce Unfairness Between Languages'', NeurIPS 2023, **arXiv:2305.15425**.

Polanyi. ''The Tacit Dimension'', Garden City, NY: Doubleday, 1966.

Raieste, Solvak, Velsberg and McBride. ''Government Efficiency in the Age of AI: Toward Resilient and Efficient Digital Democracies'', 2025, **nortal.com**.

Rajala and Jalonen. ''Beyond Simplification in Public Sector Accountability: Contradictions Between Principal-Agent and Complexity Theories'', Public Administration Review 85(6), 2025, pp. 1757–1770, **doi:10.1111/puar.13941**.

Schillemans and Busuioc. ''Predicting Public Sector Accountability: From Agency Drift to Forum Drift'', Journal of Public Administration Research and Theory 25(1), 2015, pp. 191–215, **doi:10.1093/jopart/muu024**.

Schmitz, Hammond and Chan. ''Characterizing Agentic Flooding of Government Services'', **arXiv:2608.16603v2**, 19 August 2026.

South, Marro, Hardjono, Mahari, Whitney, Greenwood, Chan and Pentland. ''Authenticated Delegation and Authorized AI Agents'', 2025, **arXiv:2501.09674**.

Spence and Zeckhauser. ''Insurance, Information, and Individual Action'', American Economic Review 61(2), 1971, pp. 380–387.

Statens historiska museer. ''Mjölkhare'', collection terminology, samlingar.shm.se, accessed 18 October 2026.

Strathern. '''Improving ratings': audit in the British University system'', European Review 5(3), 1997, pp. 305–321, **doi:10.1017/S1062798700002660**.

Svenska Akademien. Svenska Akademiens ordbok, entry ''skratte/skratt'', noun 2, 1973, svenska.se, accessed 18 October 2026.

Visa. **''Visa and Banks Across Europe Reach the Next Phase of Agentic Commerce, Unlocking Secure Access to Merchants''**, 2 July 2026, visa.co.uk.

Waterman and Meier. ''Principal-Agent Models: An Expansion?'', Journal of Public Administration Research and Theory 8(2), 1998, pp. 173–202**, doi:10.1093/oxfordjournals.jpart.a024377**.