Cúram

Where to start with AI in health and human services

Nick Raponi, Senior Vice President
Global Implementation Services, Cúram

A structured approach for agencies serving citizens at scale

Executive Summary

Health and human services agencies are under sustained pressure: demand for benefits, care, and protective services continues to rise, workforces remain constrained, and citizens increasingly expect timely, transparent, and personalized digital engagement comparable to the commercial sector. At the same time, agencies operate under strict statutory, privacy, and governance requirements due to the sensitive personal information they hold.

Recent advances in artificial intelligence present a real opportunity to improve service delivery, reduce operational burden, and give caseworkers back time for work that requires human judgment. The challenge for the sector is not whether AI is capable, but whether agencies can match the right use cases to the right deployment model and govern them within their existing frameworks to deliver real value.

This paper provides a practical, vendor-agnostic starting point for agencies seeking to move from interest to action. It does not propose new governance models or wholesale transformation programs. Instead, it sets out a structured approach that helps leaders:

- Understand the governance implications of different AI deployment models
- Identify and prioritize credible AI use cases
- Sequence adoption to deliver early, low-risk value
- Build the confidence and capability required to scale responsibly

At the core of the approach is a **Human Services AI Use Case Pyramid**, which organizes opportunities by service impact and governance complexity, and a set of **eight practical steps** that take an agency from use-case identification through to sustained operation.

The paper emphasizes starting with lower-risk, high-learning opportunities — most often in workforce support and engineering productivity rather than citizen-facing services — and using those early wins to build capability before moving into higher-value and higher-sensitivity use cases.

The key point is simple: in HHS, the main barrier to AI adoption is the ability to choose the right use cases, apply the right deployment model, and govern each one in proportion to its real-world consequences. And the stakes are long-term: agency systems typically run for 10–20 years or more after go-live, so the choices made now will shape delivery and maintenance costs for decades.

For agencies facing growing demand, constrained resources, and rising expectations, this approach offers a credible path to adopting AI deliberately and at pace—delivering tangible benefits while maintaining trust, accountability, and control.

The case for action

Health and human services agencies around the world are operating under considerable pressure. Demand for benefits, care, and protective services continues to grow. Workforces are stretched, experienced caseworkers are harder to retain, and citizen expectations, shaped by digital experiences in the commercial sector, now require interactions with government that are easy to use and access, timely, transparent, and personalized.

AI can help agencies reduce the burden of complex processes, improve responsiveness for citizens, and overcome complexities caused by legacy systems, all of which can free caseworkers to spend more time on vital tasks that require human judgment.

AI has the potential to materially improve service delivery in HHS. It can support more consistent eligibility decisions, reduce avoidable appeals, and strengthen the fairness and defensibility of outcomes - areas that directly influence public trust.

This is a fundamental shift, from static, rules-based digital systems to AI-enabled operating models that continuously learn, adapt, and assist human decision-making.

The outcome is that digital transformation is no longer the end state. AI-enabled operations become the new baseline, reshaping how service delivery, casework, and core systems maintenance are performed.

Rapid advances in AI capability present significant opportunities to agencies that can apply new technology safely in a human services context. In many organizations it is already likely that staff are informally using public AI tools to draft, summarize, and explain. In HHS, the main challenge is often not the technology itself, but how to align the use case, the deployment model, and the level of governance required.

This paper is intended to help human services leaders move from interest to action. It does not propose new governance because agencies already have governance structures. Instead, it provides a practical sequence that can be followed within an agency's existing governance, helping to identify the right starting use cases, choose an appropriate deployment approach for each, and structure adoption of new processes so that early wins build the confidence and capability needed for broader change.

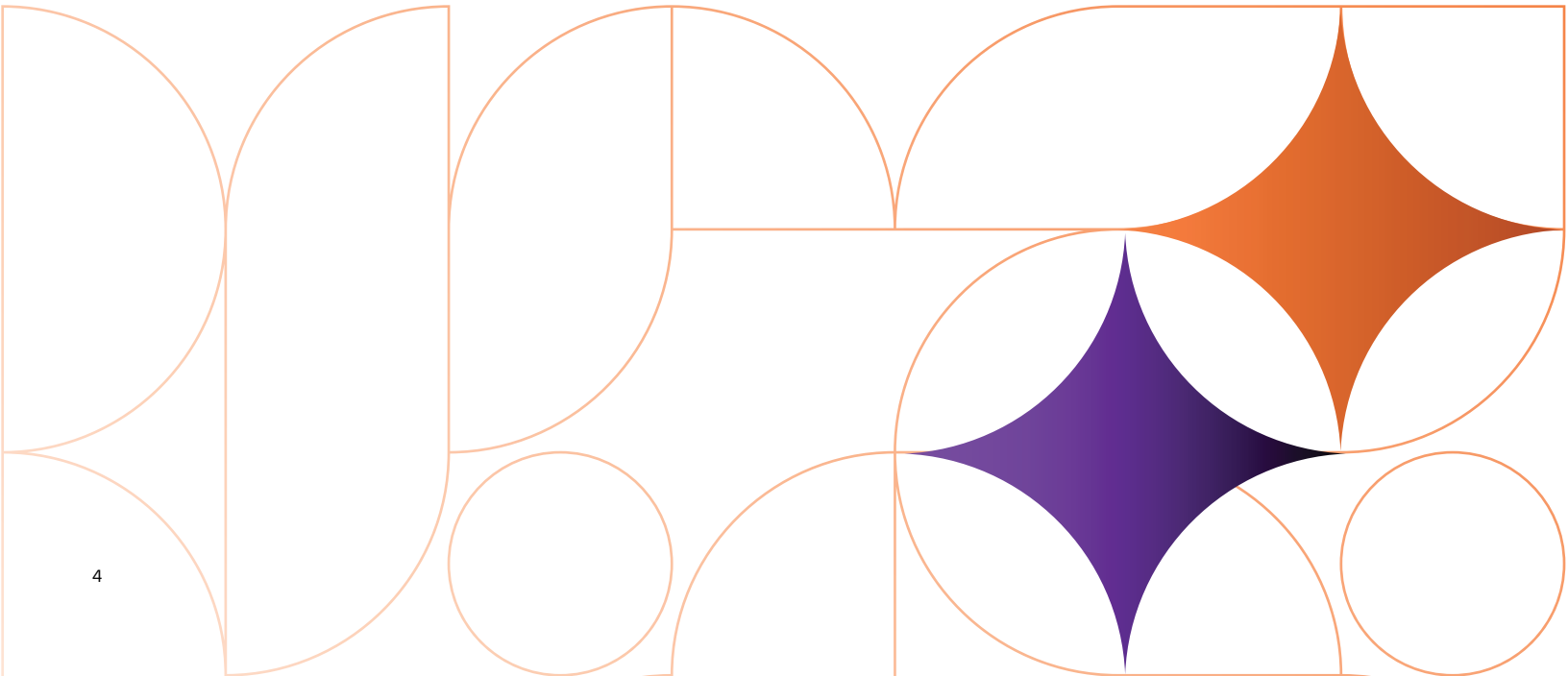
Understanding the AI landscape, deployment models, and governance implications

Before identifying use cases, leaders need a clear and shared understanding of the AI deployment options available. In practice, the choice of deployment model is inseparable from the governance approach an agency can adopt for a given use case.

Key terms:

- **Zero data retention:** A contractual and technical configuration in which the AI provider does not store prompts, responses, or derived data beyond the immediate transaction. Typically, critical for use cases that touch personal information (PI) on third-party infrastructure
- **Tenancy isolation:** Whether your data shares logical or physical infrastructure with other customers. Enterprise tenants typically offer stronger isolation than public consumer services
- **Sovereign / in-country hosting:** Ensures data does not cross jurisdictional boundaries, often a regulatory requirement for human services data
- **Retrieval-Augmented Generation (RAG):** Supplying the AI model with relevant agency content at query time, without retraining the model. Generally lower-risk than fine-tuning because agency data is not embedded in model weights
- **Fine-tuning:** Adapting a model's weights using agency data. Powerful, but introduces governance complexity around data lineage, model portability, and audit trail
- **Open vs. closed models:** Open-weight models can be self-hosted for maximum control. Closed models are typically more capable but accessed as a service
- **Software Development Life Cycle (SDLC):** The development process used by technical teams to implement business solutions
- **Maintenance and Operations (M&O):** The ongoing activities to keep a system running and up to date with legislative and business requirements post go-live. Typically runs for 10 to 20+ years post go-live, meaning savings here are critical to an agency's long-term cost management

The table below lists the common deployment options with a summary of their respective governance concerns. Agencies typically operate a portfolio of deployment options, selecting the approach that is appropriate for the sensitivity of the data and process in each use case.



Deployment model	Data sensitivity tolerance	Protection of data	Example use cases	Pros	Cons
Public LLM (consumer tier)	**Non-sensitive, public information only**	- Lowest control - Data may be retained or used for training	- General research, public policy explainers (no PI)	- Lowest cost (often free/fremium) - Minimal technical setup - Instant access to state-of-the-art models - Fastest deployment time	- Complete lack of data privacy - Data may be used for model training - No audit trails or service level agreements (SLAs) - Not suitable for any sensitive data
Enterprise tenant with zero data retention	**Moderate; suitable for some internal data**	- Contractually controlled - No training on agency data	- Casework copilots on non-PI content - Staff productivity	- Better privacy via contractual guarantees - Prevents data reuse for training - Leverage advanced models - Provides tenant isolation - Creates a dedicated space for agency work	- Higher cost than consumer tiers - Requires enterprise licensing - Still involves trust in a third-party vendor (albeit contractually regulated) - Not suitable for highly sensitive data
Private cloud (single tenant)	**High; PI permissible with controls**	- Strong isolation - Agency-controlled keys and logging	- PI-bearing casework AI - Document ingestion	- Strong isolation and dedicated resources - Full agency control over keys and access - Comprehensive logging and auditing - Scalable - Supports PI-bearing data - Strong SLAs	- Higher operational complexity than SaaS - Higher cost (infrastructure and management) - Potential vendor lock-in to the cloud provider - Requires specialized internal skills
On-premise / sovereign hosting	**Highest; including regulated and classified data**	- Maximum control - Full data residency	- High-risk decision support - Sensitive populations	- Maximum control (physical and network security) - Full data residency and sovereignty - Lowest risk to data privacy (no external trust required) - Handles the most sensitive and classified data	- Highest capital expenditure (hardware and facilities) - Highest operating expenditure (staff, maintenance, power) - Slower to innovate - Difficult to scale quickly - High management overhead
Hybrid + RAG pattern	**Variable by component**	- Model hosted externally - Sensitive data retrieved locally and not retained	- Many casework and customer-facing scenarios - Flexible balance of capability and control	- Flexibility - Uses the best-of-breed model while keeping data private - Cost-effective relative to pure private solutions - Leveraging RAG for precision - Allows a gradual transition	- Highest system complexity - Creates multiple dependency points and failure nodes - Potential latency issues - Complex data orchestration and model management

Deployment choice and governance approach are inseparable. Most agencies will operate a portfolio across these models.

The Hybrid + RAG architecture leverages an external LLM for advanced reasoning while querying and retrieving relevant secure data locally from internal agency sources. This approach balances access to state-of-the-art capability with strict data privacy control.

In practice, the same use case can fall into very different risk categories depending on where and how the model runs, and with suitable governance and approvals a use case could be deployed in several deployment models. That is why the governance conversation should begin with the deployment choice, not with a vendor demo.

Leaders should also give explicit attention to vendor independence and portability, so the agency does not become locked into a single provider or exposed to unnecessary price volatility over time.

Step 1: Define, understand, and prioritize use cases

This is the starting point. Before any procurement, pilot, or platform decision, an agency needs a clear, shared view of where AI can credibly help and in what order to pursue those opportunities. We recommend organizing candidate use cases using the pyramid below.

The human services AI use case pyramid

The pyramid has three layers. The top two are split vertically to separate use cases that involve PI from those that do not, which is the distinction likely to drive many of the governance considerations.

Top layer - Customer-facing AI
Value: enables more direct and improved communication and services to citizens.

- *Non-PI examples:* general benefits and eligibility information chatbots; multilingual policy and entitlement explainers; service navigation assistants that help citizens find the right programs and services
- *PI examples:* authenticated self-service entitlement queries; personalized case status updates; proactive notifications about renewals or required actions

Middle layer - Casework-facing AI
Value: reduce time spent on data handling so that caseworkers can focus more on direct interaction with clients, better identify and assist the most vulnerable, and (where the agency chooses) automate routine cases so human attention is directed where it has the greatest impact.

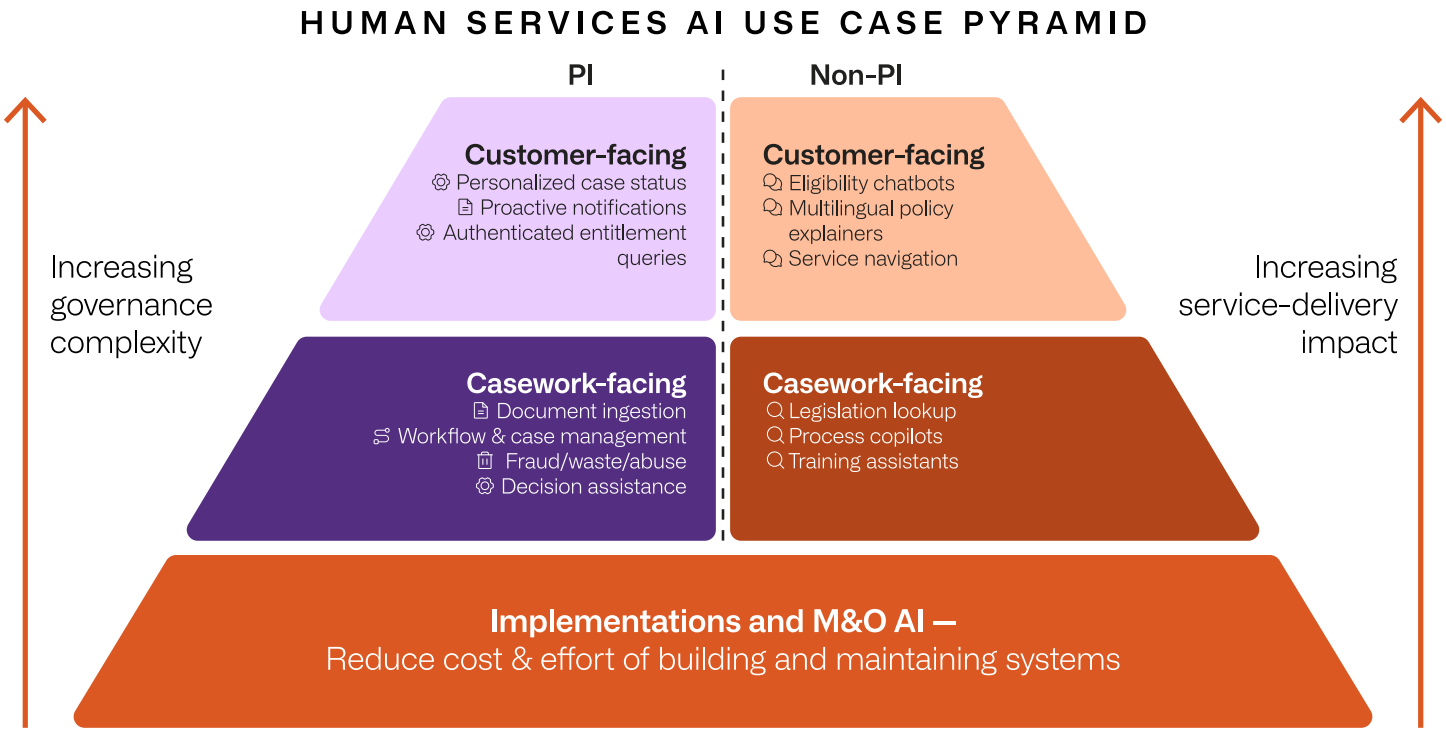
- *Non-PI examples:* legislation and policy lookup agents; process and procedure copilots; training and onboarding assistants; precedent and prior-decision search across de-identified material
- *PI examples:* document verification and ingestion; intelligent workflow and case management; fraud, waste, and abuse identification; decision-assistance and guidance for caseworkers

Bottom layer - AI for delivery and engineering
Value: significantly reduce the cost and effort of building and maintaining the systems that support the agency.

- *Examples:* code generation and legacy modernization; automated test generation; AI enhanced SDLC; technical documentation; log analysis, and incident triage; data quality and reconciliation

A general principle is to **start lower in the pyramid where governance complexity is lower and learning is faster**, then use the capability and confidence built there to move upwards into higher-value, higher-sensitivity layers. This approach enables agencies to start their AI journey quickly and safely, build the skills and experience needed to move into higher-value applications, and deliver early benefits such as lower costs and faster systems delivery.

In practice, the best early AI use cases in HHS are often not citizen-facing. They are more likely found in workforce support, knowledge access, and engineering productivity, where agencies can prove value sooner and with lower governance complexity.



Start where governance complexity is lowest, move upwards as capability and confidence grow.

What we are seeing across the sector

Across the sector, interest in AI is high, but progress is uneven. Many agencies are exploring the technology, running small experiments, or allowing (perhaps unknowingly) informal use by staff, yet remain hesitant to move into governed deployment because the governance path is unclear.

Early value is often clustering in workforce support, engineering productivity, and internal knowledge access, while higher-risk citizen-facing use cases can attract attention before the agency is ready to deliver them well.

This matters because it changes the leadership question. The issue is not simply whether AI is available or affordable. It is whether the agency can manage AI as a portfolio of use cases with different risk profiles, different deployment options, and different approval needs.

Spotlight use cases

To better understand each layer, here is one illustrative spotlight per layer, starting at the bottom:

Implementations and M&O spotlight:

Modernization Copilot: An SDLC assistant that helps delivery teams understand, design, and develop solutions directly from legislation and analysis documentation. It can also further automate the creation, management, deployment, and execution of test frameworks that support both the initial delivery and subsequent M&O of long-running agency systems over 10-20 years or more.

Data sensitivity: Low-medium (code and technical artefacts, not PI. Note however that some agencies consider codebases and configurations to be sensitive assets, and these codebases may contain third party Intellectual Property that also needs sensitive handling).

Governance considerations: IP, license attribution, secure code handling.
Value: Materially shorter modernization cycles, lower contractor spend, retained institutional knowledge. Significantly cheaper modernization and ongoing M&O of systems.

Casework spotlight:

Document Ingestion and Verification Assistant: Incoming citizen-supplied documents (identity, income evidence, medical letters and more) are classified, extracted, and pre-validated before reaching a caseworker, who confirms and adjudicates.
Data sensitivity: High (PI).

Governance considerations: private or sovereign hosting; human-in-the-loop on every consequential decision; auditability; bias testing across demographic groups.

Value: reduced time-to-decision, fewer routine touches per case, fewer avoidable errors, and more caseworker time for complex or vulnerable cases.

Customer-facing spotlight:

Multilingual Service Navigation Assistant: A conversational assistant on the agency's public website helps citizens identify which programs they may be eligible for and what to prepare, in their preferred language, without authentication.

Data sensitivity: Low to moderate (no to some PI captured; session data).

Governance considerations: Accuracy of policy information, content currency, accessibility, contestability, clear hand-off to human channels.

Value: reduced call-center load on routine enquiries; improved access for underserved populations.

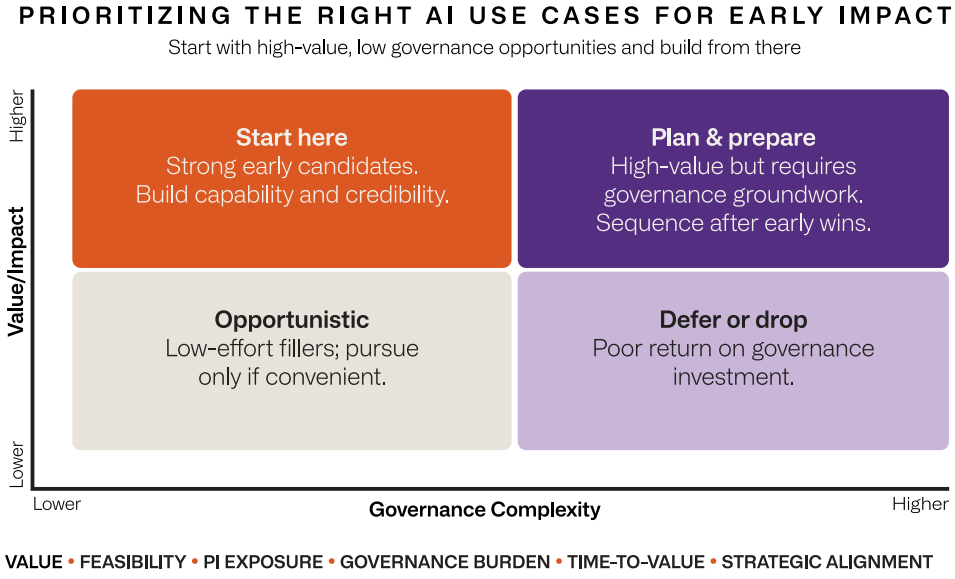
Many agencies start the AI conversation in the wrong place. They begin with vendor demonstrations or citizen-facing chatbot ideas, when the better starting point in HHS is usually a governance-aware portfolio view of use cases. That is what helps agencies separate what is attractive from what is workable.

Lightweight prioritization lens

Every candidate use case should be given a priority. We recommend a simple, transparent evaluation lens with six criteria to determine priority:

- **Value / impact:** To citizens, to caseworkers, or to operating cost
- **Feasibility:** Data availability, technical maturity, integration burden
- **Data sensitivity / PI exposure:** What data is touched, and how
- **Governance burden:** Approvals, assurance, monitoring effort required
- **Time-to-value:** How quickly meaningful benefit can be demonstrated
- **Strategic alignment:** Fit with agency outcomes and digital strategy

This can be visualized as a 2x2 of **Value/ Impact** against **Governance Complexity:**



The most likely early candidates typically sit in the high value, lower governance complexity quadrant - most often in the bottom layer and the non-PI half of the middle layer of the pyramid.

Responsible AI

Responsible AI must not be an afterthought - it should be built into every use case.

In human services, this is not only about good practice. It is about whether decisions and recommendations affecting eligibility, payments, appeals, safeguarding, and access to support are fair, defensible, and open to challenge, especially where people may be vulnerable or in crisis.

- **Human-in-the-loop** for every consequential decision affecting a citizen
- **Bias and fairness testing** across the demographics the agency serves, before deployment and on an ongoing basis
- **Explainability** sufficient for caseworkers to understand and for citizens to challenge
- **Auditability** of inputs, outputs, and model versions
- **Contestability** to offer clear, accessible routes for citizens to question an AI-influenced outcome
- **Alignment to recognized frameworks** relevant to the jurisdiction, such as the NIST AI Risk Management Framework, ISO/IEC 42001, and/or the risk tiering of the EU AI Act, applied through the agency's existing governance

These considerations should be treated as **acceptance criteria** for every use case advancing through the journey, not just as a final-stage review. If an agency cannot explain how an outcome was reached, defend it through internal review or external scrutiny, or provide a credible route for appeal or correction, the use case is not ready.

Steps 2-8:From candidate to capability

Step 1 produces a prioritized portfolio. Steps 2–8 turn that portfolio into real capability. Do not over-engineer the process. Run the steps in order and document only what you need.

Step 2 - Understand the agency's governance structure and prerequisites

Purpose: Establish how existing privacy, security, procurement, ethics, and assurance functions will engage with AI work.

Key questions: Who must be consulted and/or informed? Who should approve? What artefacts (privacy impact assessments, security reviews, ethics reviews) are mandatory? Where is AI-specific guidance still being formed?

Typical outputs: A governance map, a stakeholder RACI, a list of standing prerequisites.

Step 3 - Understand the governance requirements of the top potential use cases

Purpose: For each shortlisted use case, identify the specific governance obligations appropriate for the data, decisions, redaction, logging, retention and any citizen and caseworker impact.

Key questions: What PI is involved? What decisions does the AI influence, and how reversibly? Which deployment models are permissible?

Typical outputs: A per-use-case governance profile, refined feasibility view, deployment model recommendation.

Step 4 - Identify success factors

Purpose: Define what "good" looks like before building anything.

Key questions: What measurable outcomes will demonstrate value (cycle time, accuracy, caseworker time reallocated, citizen satisfaction, cost avoided)? What guardrails define unacceptable behavior? What will trigger a stop or pivot?

Typical outputs: Success metrics, guardrail definitions, decision criteria for proceeding past PoC.

Step 5 - Seek approval to investigate; identify early costs and funding pathways

Purpose: Secure a mandate to proceed and a realistic view of investment.

Key questions: What is the indicative cost of a proof of value, and of subsequent scaling? What funding routes exist (operating budget, transformation funds, cross-agency programs, grants)? Who must endorse?

Typical outputs: An approved investigation mandate, an indicative cost envelope, an initial funding plan.

Step 6 - Undertake an MVP Proof of Concept

Purpose: Test the use case end-to-end at meaningful but contained scale.

Key questions: Is the value hypothesis confirmed? Are the governance controls workable in practice? What did caseworkers and (where relevant) citizens actually experience?

Typical outputs: A working MVP, evidence against success metrics, lessons on operating model and change impact.

Step 7 - Assess and transition to implementation

Purpose: Turn a successful MVP into sustained service.

Key questions: Has the value anticipated been recognized? What has been learned and must change for implementation? What is required for production-grade security, monitoring, support, and accessibility? How is the change managed for staff and citizens? Who owns the capability ongoing?

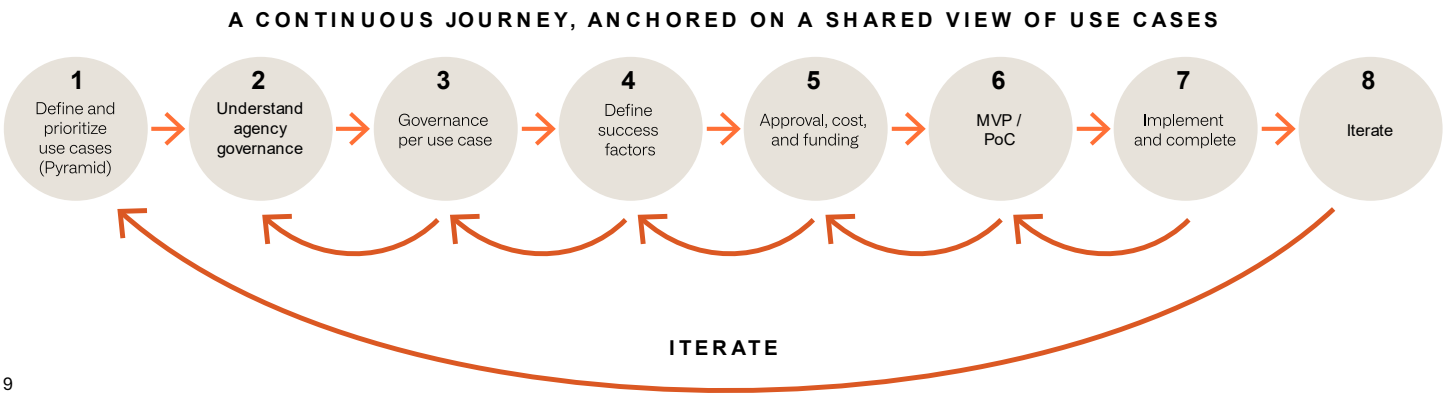
Typical outputs: A production deployment, an operating model, a defined product owner and support model.

Step 8 - Iterate

Purpose: Recognize that AI capabilities and models change rapidly and treat each capability as a living product.

Key questions: Is value being delivered? Have new model capabilities or deployment options brought new opportunities? What is the next use case that this capability unlocks?

Typical outputs: A continuous improvement backlog, periodic reassessment against the pyramid, planned review and update cycles.



Where to start

We believe that one of the most common reasons human services agencies stall on AI is not model capability itself, but the lack of a shared, structured way to align use-case choice, deployment model, and governance requirements.

That is why AI in HHS should be managed as a governed portfolio, not treated as a single technology decision. The pyramid and the eight steps are designed to provide this.

Agencies that take the following three steps will put themselves in a position to move from cautious interest to deliberate delivery, and to do so in a way that respects the citizens they serve.

Three actions to take immediately:

- **Convene the right stakeholders:** Pair business, operational and program leaders with CxO functions, privacy, security, and legal. The conversation must be a joint one from the start
- **Run a use case discovery workshop using the pyramid:** Populate each segment with candidates drawn from current pain points, then apply the prioritization lens to identify two or three credible early candidates
- **Baseline existing governance against AI-specific considerations:** Confirm what your current privacy, security, ethics, and assurance regimes already cover, and where AI-specific guidance needs to be developed or adopted

Conclusion

AI in HHS must be approached as a governed portfolio of use cases, not as a single platform decision or isolated effort.

Agencies that start with the right use cases, apply the right deployment models, govern with discipline, and hold firm on consequences, fairness, and accountability will be better placed to capture value without compromising trust. They will not only improve execution but also help reshape how human services are delivered.

The outcome is the emergence of more adaptive, responsive, and trusted human services systems that are better equipped to meet the needs of citizens at scale.

About Cúram

Cúram by Merative has over 25 years of experience helping national, regional, and local governments, and organizations across health and social ecosystems, to transform the delivery of social services, empower caseworkers, and help individuals and families access the programs they need to achieve better outcomes. Cúram solutions and services expertise are trusted in 12 countries and jurisdictions, and support over 970 government programs. Available in 7 languages, the Cúram platform connects benefits administrators, social services agencies, and case managers, to serve and protect 187 million citizens annually. Learn more at merative.com/curam

Produced in the United States of America
June 2026

SPM-000000063 Rev 1.0

Cúram