

HiiL POLICY BRIEF

# Beyond the Hype: Making AI Work for Justice Seekers

**Why AI must be governed as a justice service[1], not a technology product**

October 2026

## Executive Summary


Artificial intelligence promises to expand access to legal information, break down language barriers, and reduce administrative backlogs. However, governments and donors risk repeating the mistakes of past digital interventions by treating AI as a standalone technology solution. An AI tool can generate legally accurate answers and still fail a vulnerable person if it offers no actionable next steps, relies on unstable broadband, or disappears when a software license expires.

This brief argues for a fundamental paradigm shift: AI for justice must be evaluated, funded, and governed as a complete justice service, not a technology product. It proposes the Two-Layer Test, a new standard requiring that every AI initiative prove its tangible value to justice seekers (Layer 1) and its long-term viability for the institutions that maintain it (Layer 2). The test is designed to complement existing AI laws, risk assessments, or technical standards. It is operationalised through an AI Justice Scorecard, which governments, justice providers and funders can use when designing, financing or deploying AI-enabled justice services.

An AI system should not be considered successful just because it works. It should be considered successful when it helps people address justice problems, and when the institution responsible for the service can sustain, govern and, if necessary, safely replace it.


HiiL

1 In this brief, *justice services* encompass the full spectrum of services people use to prevent and resolve legal disputes. This ranges from broad informal services (rights education, legal triage, translation, and community paralegal support) to narrow formal processes (court proceedings, binding determinations, and legal aid eligibility assessments).

# 1. The Reality Check: AI is Already Here, but Unanchored

The global justice gap is staggering, with an estimated five billion people facing unmet justice needs globally.[2] AI is rapidly becoming part of the justice landscape. Justice ministries, legal aid organisations, donors and other justice providers are exploring its potential to expand access to legal information, support frontline professionals and make justice services more efficient. A recent survey of legal aid organisations found that 86% are already using at least one AI-powered tool. Yet adoption has moved faster than institutional readiness. The same organisations identified cost, data privacy and limited staff preparedness as significant barriers, rating their overall readiness at 5.8 out of 10.[3]

Greater adoption, however, does not necessarily translate into better access to justice. Introducing AI into a justice service does not, by itself, make that service more accessible or effective. This is the risk of tech solutionism: the belief that deploying an application automatically resolves systemic barriers.[4]

Much of the current debate rightly focuses on technical performance and responsible AI safeguards, including accuracy, privacy, security and bias.[5] These are essential, but from a people-centred justice perspective, they are not sufficient. A system can meet technical and regulatory requirements and still fail the person who needs justice. It may provide legally accurate information that a user cannot understand or act upon, require connectivity or digital skills that many users do not have, or offer no route to human assistance when a problem becomes too complex or consequential for an automated system.[6] The relevant question, therefore, is not simply whether an AI system works, but whether it helps people address their justice problems.

The examples in this brief draw substantially on HiiL's work with governments, justice providers, community leaders, and institutional partners. They illustrate practical challenges observed in that work; they are not intended to represent the full range of AI development or practice in the justice sector.

## What do we mean by a justice service?

In this brief, justice services means the range of services that help people understand, prevent, navigate or resolve problems involving their rights, responsibilities or access to remedies. These can include legal information, advice and referrals; support from paralegals, lawyers and customary or community-based intermediaries; help preparing forms and documents; and AI used by justice institutions to support administrative or decision-making processes.

This is a broad service-based definition, not a claim that every use of AI in the justice sector has the same legal status or risk. An information tool and a system that influences a court decision are not equivalent. The purpose of the Two-Layer Test is to assess the service people experience while recognising that the safeguards must be proportionate to the system's role and the consequences of error.

2 World Justice Project (2019). *Measuring the Justice Gap: A People-Centered Assessment of Unmet Justice Needs Around the World. Washingto*n, D.C.: World Justice Project.
3 HiiL (2025). AI and Access to Justice: A Snapshot of Current Trends and Future Potential. https://hiil.org/wp-content/uploads/2025/07/ai_and_access_to_justice__a_snapshot_of_current_trends_and_future_potential.pdf
4 Siddharth Peter de Souza, "The Spread of Legal Tech Solutionism and the Need for Legal Design," *European Journal of Risk Regulation* 13, no. 3 (2022): 373–388. https://doi.org/10.1017/err.2022.4
5 OECD, *Governing with Artificial Intelligence* (2025), on the need for human-centred AI governance in justice administration, continuous impact assessment, stakeholder engagement and redress. https://doi.org/10.1787/795de142-en
6 OECD, *OECD Framework and Good Practice Principles for People-Centred Justice* (2021), which identifies access barriers including cost, complexity, language, remoteness and discrimination and stresses that services should be appropriate and responsive to people's circumstances. https://doi.org/10.1787/cdc3bde7-en

## SPOTLIGHT: WHAT DOES PEOPLE-CENTRED AI LOOK LIKE IN PRACTICE?

Imagine a mother facing an illegal eviction who reaches out to a digital legal aid portal.

### A tech-centred approach

The AI accurately analyses her query and within seconds, generates a technically flawless, ten-page summary of the relevant housing law. The information is legally accurate, the system complies with privacy and security requirements, and the technology has performed exactly as designed. But the mother still does not know what she needs to do to stop the eviction. Left with a lengthy legal explanation while dealing with an immediate crisis, she becomes overwhelmed and abandons the process.

### A people-centred approach

The AI begins by asking three simple questions, in a language she understands, to identify the nature and urgency of her problem. Recognising that the eviction is imminent, it explains the immediate steps available to her and generates a one-page notice that she can use to challenge the eviction. Because her situation requires further assistance, the system routes her case to a local legal aid provider or paralegal, who contacts her within the hour and helps her take the next step.

**The difference:** Both systems may be technically sound. But while the first measures success by the quality of the AI's output, the second starts with the justice seeker's problem and what she needs to move towards a resolution. Technical accuracy, privacy and regulatory compliance are essential, but so too are accessibility, human support and the outcomes experienced by the justice seeker.[7]



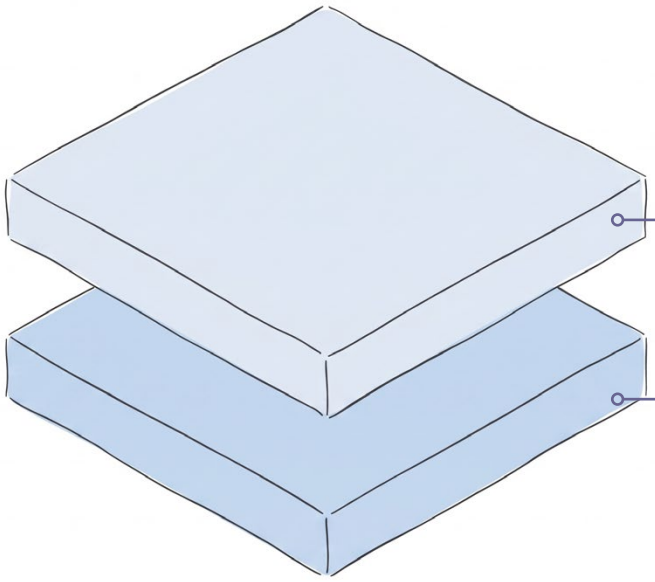
7 OECD, *OECD Framework and Good Practice Principles for People-Centred Justice* (2021), Annex A: people-centred justice services should be empowering, collaborative and integrated, focused on fair outcomes, and continually improved through evaluation and evidence-based learning. https://doi.org/10.1787/cdc3bde7-en

### LESSONS FROM THE FIELD: AI AS A SUPPORT TOOL, NOT A REPLACEMENT – NIGER[8]

In Niger, HiiL developed a digital legal-support platform and an AI-enabled voice channel to support traditional chiefs handling land-related disputes. The chiefs act as trusted intermediaries, using the service to access relevant legal and procedural information, inform their mediation, and explain available pathways to community members. The AI does not replace their judgement or mediation role; it strengthens their ability to provide informed, locally accessible support.

**The Takeaway:** AI can be particularly valuable when it strengthens trusted justice intermediaries and connects technology to an existing service ecosystem, rather than operating as a standalone substitute for human.

This is the starting point for a people-centred approach to AI in justice. Before asking how sophisticated a system is, or how widely it can be deployed, policymakers, funders and justice providers need to ask two fundamental questions: Does it create meaningful value for justice seekers, and can the institutions responsible for it sustain that value over time? These questions form the basis of the Two-Layer Test developed in this brief.



**LAYER 1: JUSTICE VALUE**

Does it actually help people solve their justice problems?

**LAYER 2: SERVICE VIABILITY**

Can the institution sustain it?

8 *Note on Field Evidence:* AI adoption in the justice sector is expanding rapidly across commercial, state, and non-profit domains worldwide. While this brief draws upon insights from broader global AI developments, its specific operational case studies highlight first-hand project implementations and field pilots led by The Hague Institute for innovation of Law (HiiL) and its partners. These case studies provide grounded, practical lessons from frontline justice environments.

# 2. Layer 1: Justice Value

## Does it actually help people solve their justice problems?

AI must earn its place in a justice service. The starting point should not be "Which AI tool should we use?" but "What justice problem are we trying to solve, and is AI the best available way to address it?". Layer 1, therefore, asks three basic questions: (1) Does AI offer a meaningful advantage over simpler alternatives for addressing the justice problem? (2) Where its use could affect a person's choices, rights or access to services, can the person challenge consequential outputs, and (3) are safeguards in place that reflect the level of risk involved?

### Test 1: The Non-AI Baseline

Before investing in AI, policymakers and justice providers should establish what could reasonably be achieved without it. In some cases, a clearer website, an updated service directory, a structured decision tree or a stronger referral arrangement may solve the problem more simply and at lower cost. AI should be used where it provides demonstrable additional value, for example, by making complex information easier to navigate, adapting guidance to an individual situation, overcoming language barriers or enabling frontline providers to serve more people effectively. The relevant comparison is not between having AI and doing nothing, but between an AI intervention and the best feasible alternative in terms of the value it delivers to justice seekers and the resources required to deliver it.

### Test 2: The Right of Contestability

Justice value also depends on preserving the agency and ability of justice seekers to challenge decisions that affect them. Where an AI system classifies a person's problem, assesses eligibility for a service, recommends a legal pathway or otherwise influences what happens next, it should not become a "black box" whose outputs cannot be questioned or reviewed. Justice seekers need to know when AI has played a role, understand the limits of that role, and have a clear and accessible way to challenge an output or request human review.[9] The precise legal right to challenge a decision will depend on the applicable law and the nature of the service; the practical standard proposed here is that providers should make review and redress accessible wherever AI materially influences consequential outcomes. This right of contestability becomes particularly important as AI moves from providing general information to influencing decisions that may affect a person's rights or access to services.

### Test 3: Match safeguards to Role and Risk

Not every use of AI in justice carries the same level or type of risk. The safeguards required should reflect what the AI is being asked to do and the consequences for the justice seeker if it gets that task wrong. A tool translating general legal information, for example, does not perform the same function or create the same risks as one determining eligibility for legal aid or drafting a document that will be filed in court.

9 N. Karusala, S. Upadhyay, R. Veeraraghavan and K. Gajos, "Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services," in *CHI Conference on Human Factors in Computing Systems* (2024). Available at: https://crcs.seas.harvard.edu/publications/understanding-contestability-margins-implications-design-algorithmic-decision

Policymakers should, therefore, avoid applying the same level of oversight to every AI application. Instead, safeguards should reflect the role the system performs, the potential harm arising from error, and the opportunities available for human review. Recent empirical research on legal AI has shown that even specialised retrieval-augmented tools continue to generate materially inaccurate or unsupported legal responses, underscoring the need for verification and supervision in consequential use cases.[10]

For practical purposes, AI applications in justice can be understood through four broad roles: **Guide, Bridge, Coach** and **Doer**. These roles help identify what can go wrong in different uses of AI and the safeguards needed in response.

The taxonomy is not intended to create a separate regulatory classification. Rather, it is a practical way for justice institutions to identify the function an AI system performs, the consequences of failure and the safeguards that should accompany it.

**Guide: Explains rights or procedures**

Typical uses include legal information, question-and-answer services and translation. The principal risk is that information may be inaccurate, incomplete or confusing. Safeguards should, therefore, focus on authoritative and validated sources, clear communication of the system's limitations, and mechanisms for identifying and correcting errors.

**Bridge: Identifies needs and connects people to services**

Typical uses include triage, issue spotting, eligibility screening and referrals. Errors can result in a justice seeker being misclassified, directed to the wrong service or missing assistance altogether. Safeguards should include clear referral rules, monitoring for misrouting and accessible routes to human review.

**Coach: Helps justice seekers prepare for the next step**

Typical uses include hearing preparation and step-by-step procedural guidance. Here, a key risk is that users may rely on automated guidance in circumstances requiring professional judgement or more individualised support. Safeguards should clearly establish the limits of automated guidance and provide escalation to human assistance where necessary.

**Doer: Produces material that may directly affect a justice process or outcome**

Typical uses include completing forms and drafting statements, applications or other legal documents, executing dispute resolution or settlement and penalty generation. Because errors may directly affect a person's rights or legal position, stronger safeguards are required, including verification of outputs before use and meaningful human review.

10 Varun Magesh et al., "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," *Journal of Empirical Legal Studies* (2025). The preregistered evaluation found substantial hallucination rates even in specialised legal research tools and highlighted the continuing need for professional supervision and verification. https://doi.org/10.1111/jels.12413

# 3. Layer 2: Service Viability

## Can the institution sustain it?

A justice service may create real value for users and still fail if the institution cannot sustain it. A service cannot be people-centred if it is unreliable, unaffordable, insecure, impossible to maintain, or dependent on actors over whom the accountable institution has no control.

Layer 2, therefore, asks whether the conditions are in place to sustain the justice value identified in Layer 1 over time. This requires looking beyond the AI model itself to the wider **sociotechnical service architecture**: the combination of technology, infrastructure, data, people, institutions and financing on which the service depends.[11]

### Pillar 1: Infrastructure and Access

The first consideration is whether the service can function in the environment in which people will actually use it. **Connectivity and Channel Fallbacks** focuses on the practical conditions that determine access: Connectivity, electricity, access to devices, digital literacy and language all shape whether an AI-enabled service that works in testing will remain accessible in practice. This means designing for real-world constraints rather than ideal conditions.

Where connectivity fails or digital access is limited, services should have accessible offline or low-tech alternatives, such as SMS, interactive voice response (IVR) or access through a physical paralegal office.

**Language, Speech and Code-Switching** are equally important to whether the technology works for the languages and ways in which people actually communicate. Performance should, therefore, be tested across relevant local languages and dialects, accents, noisy environments and code-switching, with local linguistic and legal experts involved in validating terminology and outputs. Emerging research consistently finds performance disparities between high-resource and low-resource languages, with evidence that language representation in training data is associated with model performance.[12]

Institutions also need **sufficient local technical capacity** to manage content, evaluate quality, oversee vendor relationships, troubleshoot problems and monitor system performance over time, rather than relying indefinitely on external expertise.

### Pillar 2: Data and Technical Sovereignty

Accessibility alone, however, is not enough. Institutions must also retain appropriate **data and technical sovereignty**. For AI-enabled justice services, **Knowledge Base and Data Integrity** is particularly important: statutes, case law, customary rules and other source materials must be authoritative, current and lawfully usable. There should also be a clear process, and a named institutional owner, responsible for keeping this information up to date. Systems also need to work with the wider justice ecosystem rather than creating new digital silos. Their architecture should allow information and services to connect across relevant systems and, where necessary, enable individual components to be updated or replaced.

**Interoperability** is, therefore, not simply a technical consideration: isolated "tech silos" risk reproducing existing service fragmentation at a digital scale. At the same time, sensitive justice-user data must be securely and lawfully managed, with appropriate controls for data access,

11 OECD, *Governing with Artificial Intelligence* (2025), which identifies governance, institutional capability, data, procurement and implementation challenges as core determinants of whether government AI can operate effectively. https://doi.org/10.1787/795de142-en

12 T. Alhanai et al., "Bridging the Gap: Enhancing LLM Performance for Low-Resource African Languages with New Benchmarks, Fine-Tuning, and Cultural Adjustments," *Proceedings of the AAAI Conference on Artificial Intelligence* 39 (2025): 27802–27812; and Z. Li et al., "Language Ranker: A Metric for Quantifying LLM Performance Across High and Low-Resource Languages," *Proceedings of the AAAI Conference on Artificial Intelligence* 39 (2025): 28186–28194. https://doi.org/10.1609/AAAI.V39I27.34996 ; https://doi.org/10.1609/AAAI.V39I27.35038

encryption, back-ups and incident response. Institutions should also understand where data is stored, who can access or reuse it, and whether it may be transferred to or used by external providers, including for model training.

**Hosting, Security & Data Sovereignty** determine how justice-user data is stored, protected and controlled. An institution should not become so dependent on a single vendor, proprietary model or other provider that the justice service cannot continue without it. It needs both the legal and technical ability to change critical providers without having to rebuild the service from scratch.Research on public-sector software-as-a-service has documented how organisations can underestimate lock-in and fail to establish effective exit arrangements for their own digital assets.[13] Clear accountability is also needed for decisions about the system, including who owns it, who can approve changes, manage operational risks, respond to complaints and, where necessary, suspend or shut down the system.



### LESSONS FROM THE FIELD: VALIDATE LEGAL CONTENT – TUNISIA

In Tunisia, HiiL partnered with the National Business Register (RNE), alongside technical and content partners, to develop a multilingual chatbot that helps MSMEs navigate business creation, ongoing compliance, and the prevention and resolution of disputes. The pilot showed that developing and validating the legal content can be more demanding than building the chatbot itself. Each pathway involves different laws, procedures, and institutional responsibilities, while the institution hosting the tool may not have the mandate or expertise to validate every topic it covers.



**The Takeaway:** A closed knowledge base does not, by itself, make an AI service reliable. Validation responsibilities should be assigned by topic and pathway to the relevant institutions and experts, with sufficient time and resources dedicated to thorough end-to-end testing. As Principle 4 of the UK Government's AI Playbook emphasises, where chatbot responses cannot be reviewed in real time, meaningful human control must operate at other stages of development and deployment. This requires testing before launch, ongoing assurance, regular checks, and mechanisms for users to flag issues and trigger human review.

13 M. Hedström et al., "Perceived and Actual Lock-in Effects Amongst Swedish Public Sector Organisations When Using a SaaS Solution," in *Electronic Government* (2021). The study examined 46 public-sector organisations and found significant shortcomings in licensing and exit planning for a widely used SaaS service. https://doi.org/10.1007/978-3-030-84789-0_5



**Pillar 3: Institutional and Financial Governance**

Finally, service viability also depends on strong **Institutional & Financial Governance** to ensure that the service can be effectively owned, managed and financed over time.

**Dependency on a single provider** must be considered from the outset. Critical functions should not depend on a single vendor, proprietary model or telecom agreement, and institutions must retain the legal and technical ability to switch providers without having to rebuild the application from scratch.[14] **Institutional Governance & Accountability** requires clear ownership of the service, with a named system owner and oversight body that have the authority to approve changes, manage operational risks, handle complaints and, where necessary, order the system to be suspended or shut down.

**Planning and budgeting for the full-cycle** requires looking beyond the initial costs of developing or procuring the technology to the cost of operating, maintaining, transitioning and eventually decommissioning the service. Financial planning should therefore account for ongoing model or API usage, legal-content updates, monitoring, maintenance and expert audits, as well as the resources needed to ensure continuity for justice seekers and protect institutional knowledge and data if the service is replaced or discontinued. This includes planning for a Decommissioning Budget, discussed further in Section 6.

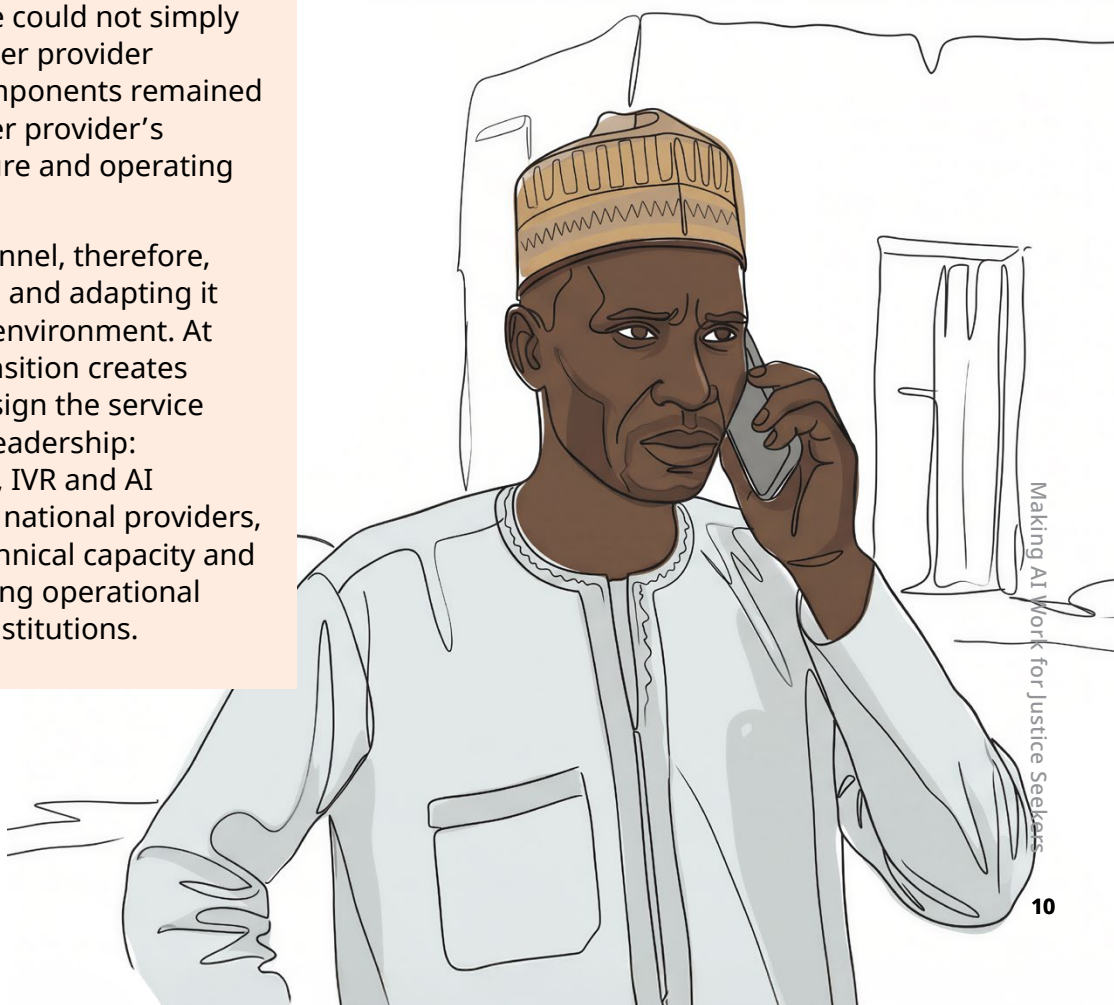
14 OECD, *Enabling Digital Innovation in Government: Defining GovTech* (2024), which identifies vendor lock-in and large ICT contracts as barriers to innovation and emphasises the need for governments to maintain responsive and secure digital services. https://www.oecd.org/en/publications/enabling-digital-innovation-in-government_a51eb9b2-en.html

## LESSONS FROM THE FIELD: WHEN A PROVIDER WITHDRAWS – NIGER

The Niger pilot, discussed earlier, showed how AI can support trusted justice intermediaries. However, it also highlights the risk of technology lock-in. The pilot combined a digital legal-support platform with an AI-enabled voice channel designed to reduce literacy and internet barriers. The voice channel showed potential value, but its operation depended on an external provider managing several interdependent components, including telecom arrangements, IVR infrastructure, AI components, technical configuration and support. When the provider discontinued the service, this dependency caused an immediate disruption and required HiiL and its local partners to restructure the service. Although HiiL obtained key project assets, including content, prompts, scripts, usage data and technical information, the service could not simply be transferred to another provider because some core components remained embedded in the former provider's proprietary infrastructure and operating model.

Restoring the voice channel, therefore, requires reconstructing and adapting it within a new technical environment. At the same time, this transition creates an opportunity to redesign the service around stronger local leadership: separating the telecom, IVR and AI components, engaging national providers, strengthening local technical capacity and progressively transferring operational responsibility to local institutions.

**The Takeaway:** Ownership of project assets does not necessarily guarantee service portability. Procurement and partnership agreements should assess not only whether data, content and configurations can be exported, but whether the service can realistically be transferred or reconstructed by another provider. Technical documentation, open or interoperable components, local implementation capacity, transition support, cost exit plans and advance testing of provider replacement are essential. A service model may deliver value to users and still remain structurally fragile if changing providers leads to significant service disruption.



| Pillar | Structural dimension | What should be assessed? |
|---|---|---|
| **1. Infrastructure and Access** | **Connectivity and Channel Fallbacks** | Can the service function under real-world bandwidth, electricity, device and digital-literacy constraints? Are accessible offline or low-tech alternatives available when the primary channel fails? |
| | **Language, Speech & Code-Switching** | Does the system perform reliably across the languages, dialects, accents and code-switching patterns of intended users, including in real-world conditions? Is terminology validated by local linguistic and legal expertise? |
| | **In-Country Technical Capacity** | Do national staff and administrators have the skills to manage content, evaluate quality, troubleshoot problems, oversee vendors and monitor performance without permanent dependence on external consultants? |
| **2. Data and Technical Sovereignty** | **Knowledge Base and Data Integrity** | Are legal and other source materials authoritative, current and lawfully usable? Is there a clear process and institutional owner responsible for keeping them up to date? |
| | **Architecture and Interoperability** | Can the system exchange information and work with relevant justice services and systems? Is the architecture sufficiently modular and documented to avoid creating another isolated technology silo? |
| | **Hosting, Security and Data Sovereignty** | Are appropriate controls in place for data storage, access, encryption, back-ups and incident response? Does the handling of justice-user data comply with applicable national requirements? |
| **3. Institutional and Financial Governance** | **Provider Dependency and Vendor Lock-in** | Can the institution change vendors, models or other critical providers without losing the service or having to rebuild it from scratch? |
| | **Institutional Governance and Accountability** | Is there a clearly identified system owner and oversight structure with authority to approve changes, manage risks, address complaints and intervene when necessary? |
| | **Full Lifecycle Financing** | Does financing cover not only initial development but the ongoing costs of operating, updating, monitoring, auditing and eventually transitioning or replacing the service? |

# 4. Alignment with International Justice and AI Frameworks

The Two-Layer Test also aligns with wider international approaches to people-centred justice and responsible AI. Across these frameworks, a common principle emerges: technology should be assessed not simply by what it can do, but by whether it responds to people's justice needs, expands meaningful access to justice, and operates with appropriate institutional and regulatory safeguards. For governments and public institutions, the Two-Layer Test is not just an ethical framework, it is a practical compliance tool for meeting international legal and development obligations.

Pillar 2 of the OECD Recommendation on Access to Justice and People-centred Justice Systems emphasises understanding people's justice needs and designing services that are appropriate, accessible and responsive to different people and circumstances.[15] Pillar 3 recognises that effective people-centred services also depend on enabling governance and infrastructure, including institutional coordination, technology, data and appropriate resources. These priorities are reflected across the Two-Layer Test: Layer 1 focuses on the value an AI-enabled service creates for justice seekers, while Layer 2 examines the institutional, technical and financial conditions needed to sustain that value.

This approach also supports the ambition of UN Sustainable Development Goal 16.3 to promote the rule of law and ensure equal access to justice for all. Applied to AI, this means that successful deployment is not an end in itself. The relevant question is whether the technology helps people understand their options, access appropriate support and move towards resolving their justice problems. It also means considering who may be excluded. Barriers related to language, literacy, connectivity or digital access can persist, or take new forms, when justice services are digitised. This way, the framework can help institutions assess whether investments in AI are contributing to greater access to justice rather than simply digitising existing barriers or inequalities.

The EU Artificial Intelligence Act provides a complementary risk-based perspective. Its requirements vary according to how an AI system is used and the risks associated with that use. Its high-risk categories include AI used to assist judicial authorities or in alternative dispute resolution[16], and AI used by or for public authorities to evaluate eligibility for essential public assistance benefits and services[17]; some legal-aid eligibility and triage systems may fall within the latter, depending on how they are deployed and by whom.

The Two-Layer Test's **AI Role Taxonomy - Guide, Bridge, Coach and Doer - does not replace these legal classifications**, but can help institutions identify the function an AI system performs within a justice service and the consequences that may follow when it gets that function wrong. This provides a practical basis for considering the regulatory obligations and safeguards that may apply to a particular use of AI, including requirements relating to transparency, risk management and human oversight.

15 OECD, Recommendation of the Council on Access to Justice and People-Centred Justice Systems, OECD/LEGAL/0498 (2023).
16 European Commission, AI Act Service Desk. "Administration of justice and democratic processes" (Annex III, point 8). https://ai-act-service-desk.ec.europa.eu/it/node/468
17 European Commission, AI Act Service Desk. "Essential services" (Annex III, point 5(a)). https://ai-act-service-desk.ec.europa.eu/es/node/465

# 5. Applying the Two-Layer Test: The AI Justice Scorecard

The scorecard translates the Two-Layer Test into a practical assessment for institutions procuring, designing, funding or deploying AI-enabled justice services. It should be used to identify where an initiative is ready to proceed and where further safeguards or redesign may be required.

***Instructions for Evaluators:***

To pass, an initiative must score a ● **Green Light** across all six dimensions.

A single ▶ **Red Flag** represents an unacceptable failure condition, indicating that the project is unsafe, unsustainable, or unhelpful in its current form.

## LAYER 1: JUSTICE VALUE

### Does AI create meaningful value for justice seekers?

| | ▶ | ● |
|---|---|---|
| 1. NON-AI BASELINE<br><br>Actionability and outcomes: does the service give people concrete actionable next steps and help them move towards resolution? | AI is deployed simply to modernise or automate an existing service, without assessing whether a simpler or lower-cost alternative could address the justice problem as effectively. | AI provides demonstrable additional value over the best feasible non-AI alternative, such as improving accessibility, navigation, personalisation or the ability of justice providers to serve people effectively. |
| 2. RIGHT OF CONTESTABILITY<br><br>Dignity and agency of the justice user | Justice seekers cannot tell when AI has influenced a consequential output, challenge that output or obtain human review where needed. | Justice seekers understand the role and limits of AI and have a clear, accessible way to challenge consequential outputs or request human review. |
| 3. CALIBRATED RISK BASED ON ROLE<br><br>Verification and escalation proportionate to role. | Safeguards are applied without considering what role the AI performs or the consequences for justice seekers if it gets that task wrong. | The AI's role - Guide, Bridge, Coach or Doer - is clearly identified, and safeguards reflect the function it performs, the potential consequences of error and the need for human review. |

## LAYER 2: SERVICE VIABILITY

Can the institution reliably sustain that value over time?

### 4. INFRASTRUCTURE AND CHANNELS

Channel fallbacks; local-language testing; in-country capacity.

▶ The service assumes reliable connectivity, devices or digital literacy; performs poorly for relevant languages or users; or depends on external technical expertise to remain operational.

● The service works under real-world access conditions, provides appropriate channel fallbacks, performs reliably across relevant languages and communication contexts, and can be managed by capable national or local staff.

### 5. DATA AND TECHNICAL SOVEREIGNTY

Named owner for legal content; data-use terms, including model training.

▶ Legal information is outdated or lacks clear ownership; the system operates as an isolated technology silo; or data and critical technical functions are controlled in ways that prevent effective institutional oversight or portability.

● Legal sources are authoritative, current and clearly governed; the system can interoperate with relevant services; and appropriate arrangements are in place for data security, sovereignty and technical portability.

### 6. INSTITUTIONAL AND FINANCIAL GOVERNANCE

Decommissioning and continuity plan.

▶ Responsibility for the service is unclear; critical functions depend on a provider that cannot readily be replaced; or financing covers deployment without adequately planning for continued operation, monitoring, maintenance or transition.

● A named owner and oversight structure have clear authority; critical functions can be transferred where necessary; and financing accounts for the full lifecycle of the service, including operation, monitoring, maintenance and transition or exit.

# 6. The Decommissioning Budget: A Funding Prerequisite

As part of Full Lifecycle Financing, institutions should plan not only for the cost of running an AI-enabled justice service, but also for the cost of safely replacing or discontinuing it. One of the most overlooked risks in public-sector technology is what happens when the money runs out, or a pilot ends. In the justice sector, the consequences can be particularly serious. If an AI-enabled service used for triage or referrals suddenly disappears, justice seekers may lose access to support while their justice problems continue. This includes, in some cases, time-sensitive procedures or deadlines.

A **Decommissioning Budget** ensures an institution has the financial resources to safely turn the AI off. It should ensure that institutional knowledge can be preserved or transferred, justice seekers can continue to access support, and their data remains appropriately protected throughout the transition. **If an institution cannot afford the cost of safely unplugging the machine, it cannot afford to turn it on**.

Policymakers must mandate that funds are ring-fenced upfront to execute a three-stage managed shutdown process:

1. **Preserving Public Knowledge**: Budget for data egress fees charged by cloud providers to extract custom legal prompts, decision trees, and citizen data. Pay an independent engineer to convert this proprietary vendor data into open-source formats (like JSON or CSV) so years of legal engineering are not lost.
2. **Maintaining Human Fall-back Services**: Provide short-term "surge staffing" funds to pay human operators, paralegals, or call-centre staff to manually handle the intake volume that the AI was previously processing. Redirect web traffic and phone lines back to a static baseline service.
3. **Safeguarding community Data**: Hire an independent third party to verify that the outgoing vendor has permanently and completely deleted all data pertaining to affected communities, ensuring it is not retained to train future commercial AI models.



# 7. Targeted Policy Recommendations

To successfully move from technological hype to sustainable justice outcomes, stakeholders must act within their specific spheres of authority.

## A. For Governments & Ministries of Justice

*Sovereignty, Public Procurement, and Systemic Oversight*

1. **Employ the Two-Layer Test in Public Procurement:** Enact administrative rules requiring every public AI proposal to undergo a dual assessment: proving both direct value to the justice seeker (Layer 1) and long-term institutional, financial, and technical viability (Layer 2) before receiving deployment approval.
2. **Contract for Portability:** Prohibit proprietary "black box" lock-in in public tech contracts. Mandate open APIs, standardised data export formats, model-substitution rights, and explicit protection against "prompt lock-in." Explicitly ban vendors from using interaction data to train commercial models.
3. **Employ Decommissioning Budgets for Digital Public Infrastructure:** Require every government AI initiative to ring-fence a decommissioning and data-extraction budget prior to launch. If the ministry cannot afford the cost of safely off-ramping vulnerable populations and extracting data, the initiative must not be approved.
4. **Assign clear responsibility:** Assign clear responsibility to a designated office or official for monitoring service quality, logging failure incidents, managing vendor contracts, and handling redress for justice seekers.

## B. For Justice Providers, Bar Associations & Civil Society

*Quality Control, User Agency, and Local Knowledge Governance*

1. **Use AI to strengthen trusted justice providers:** Prioritise using AI to empower paralegals, court clerks, and legal aid lawyers rather than deploying unguided tools for justice seekers facing high-harm legal issues. Human oversight must remain standard practice for consequential decisions.
2. **Validate Legal Knowledge:** Ensure that source databases, retrieval systems, and prompts are co-designed and validated with local communities, customary authorities, and marginalised groups. Technical accuracy must not replace community legitimacy.
3. **Define clear thresholds for errors:** Define explicit failure thresholds for critical tasks (e.g., missed statutory deadlines, incorrect eligibility rejection, or hallucinated statutes). If an AI tool violates these safety gates, it must automatically route the justice seeker to an immediate human fallback.
4. **Measure Actionability, not just Accuracy:** Evaluate tools based on whether the justice seeker successfully completes a necessary procedural step (e.g., submitting a form, reaching a referral, attending a hearing) rather than merely receiving correct legal text.

## C. For International Donors & Philanthropies

*Funding Criteria, Sector Infrastructure, and Risk Mitigation*

1. **Ask for a Non-AI Comparison:** Require applications to show why a proposed AI tool outperforms a simpler, cheaper alternative (e.g., a structured SMS flow, an updated service directory, or a static guide) before disbursing tech development funds.
2. **Fund Full Lifecycle Costs:** Shift grant allocations from software licensing toward essential ecosystem infrastructure: empowering frontline justice workers, continuous human evaluation, multilingual dataset annotation, and fallback staffing.
3. **Invest in Open-Access Sector Infrastructure:** Allocate funding toward shared public goods - such as regional evaluation benchmarks, standardised legal task rubrics, and open-source models - so grassroots providers do not have to waste resources building redundant tech stacks from scratch.
4. **Condition Disbursements on Funded Exit Strategies:** Require grantees to include an explicit, cost decommissioning budget in their proposal to ensure that vulnerable communities are not left stranded if donor funding terminates.

This Policy Brief was written by:

**Viviana Brun**
Program Manager Innovation Labs

**Zainab Malik**
Senior Policy and Advocacy Advisor

**Adedolapo Alabi**
Policy Advisor

Learn more about HiiL:
www.hiil.org

*The illustrations in this policy brief were generated using AI.*

HiiL
Dedicated to people-centred justice