MALG
MONEY ADVICE LIAISON GROUP

Working together to improve the lives of people in debt since 1987

Safe and Effective AI in Regulated Debt Advice

Safe-Use Framework

A practical framework for organisations considering or deploying AI tools within regulated, free-to-client debt advice environments

Safe-Use Framework v1.0
March 2026

Contents

1 Introducing the Debt Advice AI Roadmap 3
2 How to Use This Framework 8
3 Level 1: Adviser Assistance 10
4 Level 2: Client Self-Serve: Information and Assistance 12
5 Level 3: Agentic AI: Taking Action and Sustained Relationships 17
6 Operational Readiness Assessment 20
7 Cross-Cutting Requirements 25
8 Possible Implementation Roadmap 28

1 Introducing the Debt Advice AI Roadmap

How the Roadmap emerged

When this research began in January 2026, the starting question was relatively narrow: how can generative AI-enabled (Gen AI) self-serve tools be used safely within regulated debt advice? The assumption was that the answer would centre on chatbot-style tools interacting with clients. What the research revealed was something broader and more nuanced.

Across 28 interviews with debt advice professionals, a consistent pattern emerged. Participants did not talk about AI as a single thing. They described fundamentally different types of applications, each with different implications. An AI tool that transcribes case notes is a different proposition from one that triages clients. A tool that drafts creditor letters for adviser review is different from one that sends correspondence autonomously. A tool providing general information about Debt Relief Orders (DROs) is different from one maintaining an ongoing advisory relationship with a client over months.

These are not just differences of degree. They are differences of kind. The risk profile, the safeguard requirements, the governance implications, the technological readiness and the sector's appetite for each are fundamentally distinct. Treating them as a single category, "AI in debt advice," obscures the decisions organisations actually need to make.

The Debt Advice AI Roadmap emerged from this evidence as a way of giving the sector a shared vocabulary for these distinctions. It was developed iteratively: the initial structure was drawn from the exploratory interviews, then tested and refined through validation-phase interviews with practitioners from different organisational types. The three-level structure reflects the natural groupings revealed by the evidence, not an imposed taxonomy.

Why three levels

The three levels reflect a fundamental distinction in the type of AI involved and the degree of influence over client outcomes.

Level 1: Adviser Assistance covers tools that support adviser workflows using Gen AI. The AI produces content - drafts, summaries, lookups, analysis - that the adviser reviews before use. The adviser remains in control. The AI does not interact with clients. If it gets something wrong, the adviser catches it. This is the lowest-risk category and the one where the sector expressed the strongest and most consistent support: 82% of sector survey respondents rated adviser-support tools as suitable for use now, the highest endorsement of any application type.

Level 2: Client Self-Serve: Information and Assistance covers tools that interact directly with clients or support them alongside the advice journey, using Gen AI. Risk increases progressively as influence over client decisions grows. At the lower end, a tool providing general factual information about debt solutions carries relatively low risk. At the upper end, a tool that gathers detailed financial information for adviser review falls within the core territory of regulated advice. Sector confidence mapped closely to this gradient: 82% rated general information tools as suitable, dropping to 53% for triage and data gathering. The sector drew clear red lines: 71% said AI should never make final solution recommendations, 71% said it should never assess vulnerability, and 65% said it should never determine insolvency eligibility.

Level 3: Agentic AI: Taking Action and Sustained Relationships covers a fundamentally different category. Where Levels 1 and 2 involve AI that generates content and information, Level 3 involves AI that acts: filing forms, sending correspondence, managing processes, and maintaining sustained evolving relationships with clients. This requires Agentic AI rather than purely Generative AI. The research found no organisations currently deploying agentic AI in regulated debt advice, and sector appetite was very cautious. This level is largely forward-looking but is included because technology is advancing rapidly in adjacent regulated sectors, and the debt advice sector needs a structured basis for assessing these developments as they emerge.

The evidence behind the structure

The Roadmap is grounded in three data sources that consistently reinforce each other.

28 sector expert interviews conducted across two phases with frontline advisers, supervisors, quality assurance staff, service managers, compliance specialists, digital leads and creditor representatives. The collective experience of participants exceeded 250 years in the debt advice sector. The interviews revealed the structural complexity of debt advice, the prevalence and hiddenness of vulnerability, the professional judgement involved in financial statement construction and the critical importance of the advice relationship. Every participant, regardless of their overall attitude to AI, identified human interaction as essential for complex or vulnerable cases.

A YouGov survey of 2,060 UK adults established a population-level baseline for attitudes to AI and financial guidance. 72% were not confident in AI-generated financial advice. 81% preferred human advice. Only 2% would follow AI advice exactly as given. The strongest trust-building factor was knowing AI is regulated by financial authorities (32%), followed by human expert review (27%). These findings reinforced the graduated approach: public trust increases when human oversight is visible and when AI operates within regulated structures.

Where figures are from YouGov Plc - Total sample size was 2060. Fieldwork was undertaken between 11th-12th March 2026. The survey was carried out online. The figures have been weighted and are representative of all UK adults (aged 18+).

A sector-professional survey (n=17) providing directional data, triangulated with the qualitative evidence. The survey confirmed the risk gradient: adviser-support tools were most frequently top-ranked, personalised advisory consistently ranked last. The survey also identified the sector’s red lines on what AI should never do, which informed the boundaries between levels.

What the Roadmap is and what it is not

The Roadmap is a guide to possible applications. It describes where AI might be deployed in regulated debt advice, what type of AI each application requires, what risk it carries and what level of appetite the sector currently has for it. It provides a shared vocabulary that organisations, regulators and technology developers can use to have structured conversations about AI adoption.

The Roadmap is not a prescription. It does not say that every organisation should progress through all three levels. It does not assume that more AI is better. It does not suggest that the technology is ready for every application it describes. For some organisations, Level 1 may represent the full and appropriate extent of AI adoption. For others, certain Level 2 applications may genuinely extend reach and improve access. Level 3 is included so the sector has a framework for assessment as the technology evolves, not because the evidence supports deployment now.

The Roadmap also acknowledges that boundaries between levels are not always clean. Some tools at the upper end of Level 2, particularly those maintaining ongoing client engagement, share characteristics with Level 3 applications. The distinction rests on whether the AI generates content or takes actions. Where a specific tool sits on this boundary depends on its actual functionality rather than how it is marketed.

The central principle: proportionality

The single most important principle underpinning the Roadmap is proportionality. As the influence of an AI tool on client outcomes increases, so must governance, human oversight, regulatory scrutiny, transparency, and sector involvement in its design and testing.

A transcription tool that helps an adviser write case notes faster requires data protection compliance, accuracy testing and adviser review. These are important but proportionate safeguards for a low-influence application. A structured triage tool that screens for vulnerability and routes clients to services requires conservative escalation thresholds, evidence-based vulnerability screening, auditable outcomes, regular accuracy reviews, and monitoring of false-negative rates as a priority safety metric. The safeguards are more demanding because the consequences of getting it wrong are more serious.

This principle runs through the entire Safe-Use Framework. The organisational readiness assessment, the minimum standards for each application type, the escalation triggers, the cross-cutting requirements and the implementation roadmap are all calibrated by the degree of influence the AI tool has over what happens to the client.

The Roadmap at a glance

The following table summarises the three levels, their application types and the sector's current appetite for each.

Level	Application types	Risk and sector appetite
Level 1 **Adviser Assistance**	**1.1** Administrative assistance **1.2** Knowledge and compliance support **1.3** Operational intelligence	**Risk:** Low to moderate **Appetite:** Strong. 82% of sector survey respondents rated adviser-support tools as suitable now—the strongest endorsement of any application. Several organisations are already piloting. **Technology:** Generative AI. Adviser reviews all output before use.
Level 2 **Client Self-Serve: Information and Assistance**	**2.1** General information **2.2** Contextual guidance **2.3** Structured triage **2.4** Pre-appointment data gathering **2.5** Communication support **2.6** Documentation support **2.7** Ongoing engagement	**Risk:** Low to high, increasing with influence over client decisions **Appetite:** Graduated. 82% support general information tools; drops to 53% for triage and data gathering. 71% say AI should never make solution recommendations. 71% say AI should never assess vulnerability **Technology:** Generative AI interacting directly with clients. Current accuracy (~75%) falls short of the 85–90% threshold identified by the sector.
Level 3 **Agentic AI: Taking Action and Sustained Relationships**	**3.1** Personalised advisory **3.2** Adviser-directed actions **3.3** Semi-autonomous actions **3.4** Autonomous process management	**Risk:** High to critical **Appetite:** Very cautious. Personalised advisory rated suitable by only 18% (non-priority) and 12% (all debts). The research found no organisations currently deploying agentic AI in debt advice. **Technology:** Agentic AI that takes actions, not just generates content. Largely forward-looking. 3.4 Autonomous process management may not currently be appropriate for regulated debt advice

How the Roadmap connects to this framework

The remainder of this document provides the detailed safe-use guidance for each level. For every application type within the Roadmap, the framework sets out:

- why the application type exists and what user need it addresses
- the risk level and what the research found about sector appetite
- the minimum standards and safeguards required
- escalation triggers where applicable

- where the technology is not yet ready, and what would need to change

Following the level-specific guidance, the framework includes an Organisational Readiness Assessment: seven dimensions that any organisation should assess before deploying any AI tool, regardless of level. These dimensions are designed to ensure that the fundamental questions, what problem are you solving, who is accountable, where does data go, are your people ready, who do you serve, and can you support the technology, are answered before technology decisions are made.

> *The Roadmap describes what is possible. The following framework describes what is safe. Both are needed. An organisation that understands the possibilities but ignores the safeguards puts clients at risk. An organisation that understands the safeguards but has no structured way to assess applications may either avoid AI entirely or adopt it without adequate governance. The Roadmap and the framework work together.*

2 How to Use This Framework

Purpose

This framework is a practical tool for organisations considering or deploying AI tools within regulated debt advice. It defines when AI-enabled tools are appropriate, for whom, at what level and with what safeguards.

The framework is structured around the Debt Advice AI Roadmap, a three-level approach to understanding where AI might be applied in regulated debt advice, described above. The Roadmap recognises that AI is not a single category of activity. It spans fundamentally different types of application, from tools that help advisers write case notes more efficiently through to autonomous systems that could theoretically manage multi-step processes without human involvement. Each carries a different risk profile, requires different safeguards and sits at a different point of technological readiness.

Every organisation is different.

This framework does not assume that every organisation will or should progress through all three levels. The debt advice sector is diverse. Organisations vary in size, funding, technical capacity, client base and the complexity of the cases they handle.

The right level of AI adoption depends on the needs of both clients and internal users. An organisation primarily serving clients in acute crisis with low digital confidence, complex, intersecting vulnerabilities, and limited English proficiency may determine that client-facing AI tools would not serve those clients well. That is a legitimate and well-evidenced conclusion, not a failure to innovate. Conversely, an organisation with a digitally confident client base presenting with relatively straightforward debt may find that Level 2 tools genuinely extend reach and improve access.

The framework is designed to support both conclusions equally. It helps organisations make informed, evidence-based decisions about where AI fits within their specific context rather than prescribing a single path.

Why use cases matter

A recurring theme in the evidence underpinning this framework is that organisations considering AI often focus on the technology first: which platform to use, what model to deploy, how to integrate with existing systems. The research found that a more productive starting point is the use case. What specific problem are you trying to solve? What user need are you trying to meet? What does the current process look like, and where does it fall short?

Without clarity on use cases, technology decisions happen in a vacuum. An organisation may invest in a sophisticated chatbot without having established whether its clients would use one, or deploy a triage tool without having mapped the decision logic that experienced advisers apply.

The framework, therefore, begins with organisational readiness, including a dedicated dimension on understanding your use cases, before setting out the requirements for each level.

This approach also recognises that some use cases may not require AI at all. If the problem is that clients cannot reach an adviser quickly enough, the solution might be additional staffing, better scheduling or improved telephony rather than a chatbot. AI is one tool among many, and the framework encourages organisations to consider it within that broader context.

Overlap between levels

The boundaries between the three levels are not always clean. Some tools at the upper end of Level 2, particularly those maintaining ongoing client engagement (2.7), share characteristics with agentic applications in Level 3. The distinction is that Level 2 tools generate content and information, whereas Level 3 tools take actions or maintain sustained autonomous relationships. Where a specific tool sits on this boundary will depend on its actual functionality rather than its marketing description. Organisations should assess the tool's behaviour, not its label.

Who is this for

- **Advice providers** can use this framework to assess whether specific AI tools are appropriate for their organisation and client base, identify required safeguards and structure governance
- **Commissioners and funders** can set expectations for AI use within funded services
- **Regulators** can develop proportionate oversight calibrated by level
- **Technology developers** can understand where tools sit within the sector's risk landscape and what standards are expected

Key principles

Six principles underpin this framework

- **Proportionality:** Safeguards proportionate to influence over client outcomes
- **Human oversight:** Decisions materially affecting clients require human professional review
- **Transparency:** Clients must understand when AI is involved
- **Sector involvement:** AI tools must be developed, tested and governed with sector participation
- **Conservative escalation:** When in doubt, escalate to a human adviser
- **Client-centred design:** AI should improve outcomes for clients, not reduce costs at the expense of quality

3 Level 1: Adviser Assistance

Level 1 covers Gen AI tools that support advisers and organisational functions without directly interacting with clients. These tools use Gen AI to produce content, summaries, drafts and analysis that the adviser reviews before use. The adviser remains in control at all times.

This is the lowest-risk level and the appropriate starting point for most organisations. The evidence for Level 1 is the strongest of any in the Roadmap: 82% of sector survey respondents rated adviser-support tools as suitable for use now, and several organisations in the research are already piloting these tools with positive results.

The rationale for starting here is straightforward. Level 1 tools operate behind the scenes. If an AI tool produces an inaccurate case note summary or a poorly drafted letter, the adviser catches it before it goes anywhere. The consequence of error is wasted time, not harm to clients. This creates a safe environment for organisations to build experience, confidence and governance capability before considering higher-risk applications.

Administrative assistance (1.1)

Risk level: Low

Applications include transcription, case note summarisation, letter drafting, and document formatting. The adviser reviews all output before it is used.

Why this application type exists: Administrative burden was identified by virtually every research participant as the single most time-consuming aspect of their work after direct client interaction. Advisers reported spending hours writing up case notes, with some estimating that documentation takes as long as the advice session itself. AI tools at this level address a clear, well-defined need with minimal risk.

Minimum standards

- ☐ Advisers always review before adding to a case, sending to clients, etc
- ☐ AI content is marked as draft until reviewed by an adviser
- ☐ Data protection compliance should be confirmed
- ☐ Accuracy tested with sector-specific terminology
- ☐ Feedback mechanism for error reporting
- ☐ The tool does not store or learn from client data without authorisation

Knowledge and compliance support (1.2)

Risk level: Low–Moderate

Applications include policy lookups, legislation decoding, regulatory guidance, and compliance checking from approved sources. Advisory to the professional, not directly to the client.

Why this application type exists: Debt advice operates across multiple legal jurisdictions, complex and frequently changing legislation and overlapping regulatory requirements. The research found that experienced advisers spend a great deal of time staying current. AI tools at this level can help advisers navigate this complexity more efficiently, but the professional must still assess relevance to the specific case. The risk increases slightly because incorrect compliance information could lead to inappropriate advice if it goes undetected.

Minimum standards

- ☐ Draws from approved, curated sources rather than the open internet
- ☐ Version-controlled and regularly updated
- ☐ Correctly distinguishes jurisdictions (England/Wales, Scotland, NI)
- ☐ Output includes source references for verification
- ☐ Clear update process for case law and legislative changes
- ☐ Acknowledges uncertainty rather than guessing

Operational intelligence (1.3)

Risk level: Moderate

Applications include caseload trend analysis, early warning systems, QA call analysis, training needs identification, and workload optimisation.

Why this application type exists: Organisations need to understand patterns across their caseloads to allocate resources effectively, identify training needs and maintain quality. AI can process large volumes of case data to surface trends that would be invisible to manual review. However, this level introduces new considerations around data protection, algorithmic bias and the risk of automated decisions about individual clients based on pattern matching.

Minimum standards

- ☐ Data anonymised or aggregated where possible
- ☐ Appropriate legal bases for individual-level data
- ☐ Findings reviewed by humans before action
- ☐ No automated decisions about individual clients from pattern matching
- ☐ Algorithmic bias assessed
- ☐ QA findings verified through human sample checking

4 Level 2: Client Self-Serve: Information and Assistance

Level 2 covers AI tools that interact directly with individuals seeking help with debt, as well as tools that support clients throughout the advice journey. These tools use Gen AI to provide information, answer questions, assist with practical tasks, and, at the higher end, gather pre-appointment information for adviser review. Risk increases significantly because the tool is no longer operating behind the scenes with an adviser reviewing every output.

The research found that the sector is cautiously open to client-facing AI tools, but with strong conditions. Confidence mapped closely to the risk gradient within this level: 82% of survey respondents rated general information tools (2.1) as suitable now, but this dropped to 53% for structured triage (2.3) and pre-appointment data gathering (2.4). The sector was clear about where AI should not operate: 71% said it should never make final solution recommendations, 71% said it should never assess vulnerability, and 65% said it should never determine insolvency eligibility.

Requirements within Level 2 are cumulative for the client-facing application types (2.1–2.4). Each builds on the requirements of those below it. A tool capable of structured triage (2.3) must also meet all the standards for general information (2.1) and contextual guidance (2.2), as it will inevitably handle queries at those levels as well.

Technology readiness: The research found that current AI accuracy may be insufficient for some applications within Level 2. Purpose-built tools tested by participants achieved approximately 75% accuracy against a threshold of 85–90%. General-purpose AI tools returned outdated information, US-sourced content for English queries and fabricated legal citations. Organisations considering Level 2 deployment should satisfy themselves that the specific tool they are evaluating meets the accuracy requirements for the application type in question.

General information (2.1)

Risk level: Low

Applications include factual Q&A, terminology, and debt process walkthroughs with no personal data requirement. It does not provide recommendations.

Why this application type exists: Many people approaching debt advice services need basic information before they are ready for a full advice session. What is a DRO? How does an IVA work? What happens if I do not pay a council tax bill? These are factual questions that can be answered from curated, approved sources without personal data or individual assessment. Providing this information via AI frees advisers' time for cases that genuinely require professional judgement.

Minimum standards

- ☐ Approved, curated knowledge sources only
- ☐ Jurisdictional awareness (asks the country before providing information)

- ☐ Informative tone, not authoritative, avoids framing as personalised advice
- ☐ Clear disclaimers: general information, not a substitute for professional advice
- ☐ Signposts to professional advice services
- ☐ Content at an appropriate reading age
- ☐ Knowledge base reviewed quarterly, updated within a minimum number of working days of any significant changes
- ☐ No collection or storage of personally identifiable information

Escalation

- Crisis indicators (bailiff, court, eviction): display emergency contacts, recommend immediate human contact
- Questions requiring personalised assessment: explain why professional advice is needed

Contextual guidance (2.2)

Risk level: Low–Moderate

Applications include individualised responses using limited context: debt prioritisation, creditor contact guidance, and letter drafting.

Why this application type exists: Some queries require the tool to use limited contextual information provided by the user. A person might describe a specific type of debt and ask how to prioritise it, or request help drafting a letter to a specific creditor. This is more than general information because the individual's circumstances shape the response, but it stops short of assessment or recommendation.

Additional standards

- ☐ Responses based only on user-provided information with clear caveats
- ☐ Conservative framing in crisis scenarios
- ☐ No sensitive data beyond immediate query needs
- ☐ Does not assess eligibility or recommend solutions

Escalation

- All 2.1 triggers plus: multiple priority debts, vulnerability indicators, enforcement action → recommend professional advice
- Distress, confusion, frustration → offer human contact

Structured triage (2.3)

Risk level: Moderate–High

Applications include structured questions to identify urgency, complexity and risk. Routes users to appropriate services—key risk: failure to escalate.

Why this application type exists: Many organisations struggle with demand management. Clients wait weeks for an appointment, during which their situation may deteriorate. A well-designed triage tool could identify urgency, route high-risk cases to immediate human contact and help lower-complexity cases prepare for their appointment. However, this is the point at which the consequences of getting it wrong become serious. A triage tool that fails to identify vulnerability, misclassifies urgency or gives false reassurance could cause real harm.

Additional standards

- ☐ Triage logic developed with experienced debt advisers
- ☐ Conservative escalation thresholds
- ☐ Evidence-based vulnerability screening beyond self-report
- ☐ Jurisdiction-specific triage pathways
- ☐ Outcomes logged and auditable
- ☐ Regular accuracy review comparing AI against professional assessment
- ☐ False negative rates monitored as priority safety metric
- ☐ Never tells the user their situation is not serious

Mandatory escalation triggers (2.3 and above)

Immediate human escalation is required for certain triggers. These should be based on the organisation's own client needs, but the following are those suggested by the research:

- Bailiff involvement or threat
- Court proceedings (possession, charging order, bankruptcy petition)
- Rent arrears threatening tenancy
- Mental health disclosure or significant distress
- Financial coercion indicators
- Repeated contradictory responses
- Low digital confidence
- Client confusion
- Physical health issues affecting capacity
- Previous debt solutions affecting options
- Business debts or self-employment
- Any domestic abuse or safeguarding indication

Pre-appointment data gathering (2.4)

Risk level: High

Applications include the collection of detailed circumstances and financial information for adviser review, within the core regulated advice territory.

Why this application type exists: A significant proportion of adviser time is spent gathering basic financial information that the client could potentially provide in advance. If a tool can guide clients through initial budget preparation, flag inconsistencies and present a draft financial statement for adviser review, it could substantially reduce appointment time and allow advisers to focus on interpretation and advice rather than data collection. The risk is significant because errors in data gathering can cascade through the entire advice process. These types of tools are already in use to different degrees, with different levels of success. Usability and client appeal are key to the implementation of these tools; otherwise, they risk failure.

Additional standards

- ☐ All information is positioned as preliminary, subject to professional review
- ☐ Does not calculate or present debt solution eligibility
- ☐ Challenges unrealistic entries but does not override user input
- ☐ Financial statement framed as a draft for adviser review
- ☐ Adviser review is mandatory before any advice
- ☐ Flags gaps, inconsistencies and areas for follow-up
- ☐ Open banking subject to specific consent, treated as supplementary
- ☐ Highest data security standards
- ☐ Does not result in additional overwhelm for the client

Communication support (2.5)

Risk level: Low–Moderate

Applications include drafting creditor letters, conversation coaching, and production of standard correspondence.

Why this application type exists: Communicating with creditors is a practical task that many clients find intimidating. AI can help draft standard holding letters, prepare clients for difficult conversations and provide templates for routine correspondence. The key distinction is between standard communications (which carry lower risk) and consequential communications such as statute-barred debt assertions or formal complaints (which could affect the client’s legal position and require adviser review). Many organisations already provide templates to clients that would form a good basis for implementing these tools. As with 2.4 applications, usability and client appeal are key.

- ☐ Standard holding letters appropriate for unsupervised use with disclaimers
- ☐ Consequential communications (statute-barred, complaints, enforcement) require adviser review
- ☐ Does not negotiate or commit to payment arrangements
- ☐ Jurisdiction-specific templates regularly reviewed
- ☐ Accessible language, avoiding unexplained jargon

Documentation support (2.6)

Risk level: Low

Applications include helping clients organise paperwork, understanding correspondence, and preparing for appointments.

Why this application type exists: Many clients arrive at advice appointments with disorganised paperwork and a limited understanding of the correspondence they have received. A tool that helps clients sort, categorise and understand their documents before an appointment could improve the efficiency of the advice session and reduce client anxiety. As with 2.4 applications, usability and client appeal are key.

- ☐ Does not interpret legal or financial significance, focuses on organisation
- ☐ Flags enforcement or court-related documents as urgent
- ☐ Does not store document copies without consent

Ongoing engagement (2.7)

Risk level: Moderate

Applications include reminders, progress tracking, repayment prompts, and re-engagement.

Why this application type exists: Debt advice is rarely a single interaction. Clients may need to make regular payments, respond to creditor correspondence, attend court hearings or complete steps within a debt solution. Engagement tools can support clients through this ongoing journey. However, the tone and approach must be carefully calibrated. Clients in financial difficulty are often already experiencing stress, and the last thing they need is intrusive or pressuring automated contact.

Note: There is a natural overlap between 2.7 and Level 3. Simple scheduled reminders and progress prompts sit within Level 2. If the tool begins adapting its behaviour based on client responses, maintaining context across sessions or taking actions autonomously, it crosses into Level 3 territory. This tool could also be considered as part of Level 1 if managed by advisers.

- ☐ Sensitive frequency and tone, not intrusive or pressuring
- ☐ Adapts to non-engagement rather than escalating contact
- ☐ Re-engagement offers help rather than demanding action
- ☐ Easy opt-out mechanism
- ☐ Engagement data identifies the need for support, not to penalise

5 Level 3: Agentic AI: Taking Action and Sustained Relationships

Level 3 addresses a fundamentally different category of AI. While Levels 1 and 2 use Gen AI to produce content and information, Level 3 involves Agentic AI: systems that take autonomous actions, execute multi-step workflows, use external tools, and maintain sustained, evolving relationships with clients. This is AI that does things, not just says things.

The research found no organisations currently deploying agentic AI in regulated debt advice. This level is largely forward-looking. However, agentic capabilities are advancing rapidly across the wider technology landscape, including in adjacent regulated sectors. The sector will encounter these applications and needs a structured basis for assessment.

The critical concern at Level 3 is that actions taken by an AI system can have real-world consequences that are difficult or impossible to reverse. An incorrectly filed form, correspondence sent to the wrong creditor, or a payment commitment made without proper authorisation could cause significant harm to the client. The safeguards required at this level reflect that reality.

Personalised advisory (3.1)

Risk level: Highest

Applications include ongoing personalised support using stored information as well as scenario modelling, progress tracking, sustained advisory influence.

Why this application type exists: This represents the most ambitious client-facing application. A tool operating at 3.1 would maintain an ongoing relationship with the individual, store their information across sessions and provide personalised guidance that evolves. The evidence suggests this is not currently appropriate for most organisations. The technology is not sufficiently accurate, the regulatory position on the information-advice boundary is not sufficiently clear, and the required safeguards are substantial. Personalised advisory tools were rated as suitable by only 18% of survey respondents for non-priority debts and 12% for all debts.

Standards

- ☐ Regulatory and legal review of information-advice boundary position
- ☐ Consumer Duty compliance assessed and documented
- ☐ Human oversight through regular sampling, audit and outcome monitoring
- ☐ Does not make final solution recommendations
- ☐ Outcome monitoring against human-delivered advice
- ☐ Same QA processes as human advice
- ☐ Regular jurisdictional and legislative updates
- ☐ Clear accountability for AI-generated guidance
- ☐ Human advice is accessible at any point

Not currently recommended for

- Organisations serving primarily highly vulnerable clients
- Insolvency, competing priority debts or enforcement cases
- Organisations without experience at lower levels
- Any context with an unclear accountability framework

Adviser-directed actions (3.2)

Risk level: High

AI executing tasks under direct adviser instruction, such as filing forms, sending pre-approved correspondence, and updating records.

Why this application type exists: Advisers perform many routine actions that follow predictable patterns. Filing a standard form, sending a holding letter to a creditor, or updating a case management record are tasks that AI could perform under direct instruction, freeing the adviser to focus on judgement and client interaction. This is the Level 3 application type most likely to emerge in the near future. The critical principle is that the adviser initiates every action and confirms before execution.

Standards

- ☐ Every action requires explicit adviser initiation
- ☐ Confirmation step before execution of any client-affecting action
- ☐ Full audit trail: who instructed, what executed, outcome
- ☐ Cancel/reverse capability
- ☐ Two-step confirmation for actions affecting legal position

Semi-autonomous actions (3.3)

Risk level: Very High

AI acts within pre-approved boundaries, with adviser approval for consequential actions.

Why this application type exists: It envisions AI operating within defined parameters without requiring explicit instructions for each action, but with approval gates for anything consequential. An example might be a system that automatically sends appointment reminders, generates routine correspondence or updates case records within pre-approved rules, but requires an adviser's sign-off before sending anything to a creditor or filing a statutory form.

Standards

- ☐ Action boundaries explicitly defined, documented, and reviewed
- ☐ Consequential actions require adviser approval before execution
- ☐ Clear log of all actions and triggering rules
- ☐ Override capability
- ☐ Outcome monitoring against human-directed equivalents
- ☐ Limited caseload piloting mandatory

Autonomous process management (3.4)

Risk level: Critical

AI independently manages multi-step processes. May not currently be appropriate for regulated debt advice.

Why this application type exists: 3.4 is included for completeness and to provide the sector with a framework for assessing fully autonomous capabilities if and when they become viable. The research evidence is clear that this is not currently appropriate. The combination of legal complexity, the prevalence of vulnerability, jurisdictional variation, and the potential for irreversible harm means that autonomous process management would require a level of reliability, accountability, and regulatory clarity that does not yet exist.

Would require before consideration

- ☐ Regulatory approval or explicit guidance
- ☐ Extensive piloting demonstrating equivalent outcomes
- ☐ Full liability/accountability frameworks
- ☐ Automatic suspension if outcomes fall below thresholds
- ☐ Client opt-out to human advice without delay
- ☐ Sector-wide agreement on use cases and boundaries

6 Operational Readiness Assessment

Introduction

Before deploying any AI tool, organisations should assess their readiness across seven dimensions. This assessment is not a one-off exercise. As organisations introduce new tools or consider moving between levels, readiness should be reassessed. The dimensions are interconnected: technical readiness without governance is dangerous, governance without workforce readiness is ineffective and any deployment without a clear understanding of the use case risks solving the wrong problem.

Use case and user need clarity

This dimension comes first for a reason. Before selecting a tool, choosing a platform or assessing technical requirements, an organisation needs to be clear about what it is trying to achieve and for whom. The research found that the most productive AI deployments in the sector started with a specific, well-defined problem rather than a general aspiration to "use AI."

A clear use case articulates what process or task is being considered for AI, why the current approach falls short, who the intended users are (advisers, clients, managers), what outcome a successful deployment would achieve and how that outcome would be measured. Without this clarity, organisations risk deploying technology that solves the wrong problem, duplicates existing capability or introduces risk without commensurate benefit.

This dimension also recognises that not every problem requires an AI solution. If the underlying issue is insufficient staffing, inadequate training, poor data quality or process inefficiency, technology may mask rather than resolve the problem.

- ☐ Specific use case or problem clearly articulated
- ☐ Current process for handling this use case is documented and understood
- ☐ Intended users identified (advisers, clients, operational teams)
- ☐ User needs to drive the use case
- ☐ Success criteria and measurable outcomes defined
- ☐ Alternative non-AI solutions considered and assessed
- ☐ Use case mapped to the appropriate level of the Debt Advice AI Roadmap
- ☐ Proportionality assessed: does the scale of the problem justify the investment and risk of an AI solution

Governance readiness

Effective AI governance requires clear accountability, documented policies and proportionate oversight. For many debt advice organisations, this does not mean creating entirely new governance structures. It means extending existing governance to account for AI-specific considerations. The key question is whether the organisation has a named individual who is accountable for AI use, understands the regulatory obligations involved and has the authority to pause or withdraw tools if they are not performing as expected.

For organisations subject to the Senior Managers and Certification Regime, AI governance should sit within existing accountability structures. For smaller organisations, this may simply mean a board-level decision confirming who is responsible and which approval processes apply before a new tool is introduced.

- ☐ Named individual accountable for AI use
- ☐ AI use policy approved at the appropriate level
- ☐ Policy distinguishes Level 1 (adviser tools) from Levels 2 and 3 (client-facing and agentic)
- ☐ Process for reviewing and approving new AI tools before deployment
- ☐ Process for monitoring performance after deployment
- ☐ Consumer Duty obligations understood in relation to AI
- ☐ Senior managers' regime responsibilities considered where applicable

Data protection readiness

AI tools process information, and in a debt advice context, that information is often highly sensitive. It can include details of income, expenditure, health conditions, housing status, relationships and legal proceedings. Data protection readiness is not simply about ticking compliance boxes. It requires a genuine understanding of where data goes when it enters an AI tool, what happens to it during processing and whether it is retained, shared or used for purposes beyond the immediate interaction.

Many commercially available AI tools process data through cloud infrastructure that may be hosted outside the UK. Some use interaction data to improve their models, meaning client information could influence future outputs for other users. Organisations must understand these data flows before deployment, not after. The preparation of a Data Protection Impact Assessment (DPIA) is both a legal requirement for tools that process personal data at scale and a practical activity that prompts the right questions to be asked early.

- ☐ DPIA completed or planned for each AI tool
- ☐ Legal basis assessed for client data processing
- ☐ Data processing agreements with third-party providers
- ☐ Consent mechanisms in place where required
- ☐ Understanding of where data is stored, processed and whether it leaves the UK
- ☐ Staff understand what client information can and cannot be entered into AI tools
- ☐ Data breach response plan accounts for AI-related incidents

Technical readiness

Technical readiness is about more than having the right infrastructure. It includes understanding how an AI tool integrates with existing systems, what happens when it fails and whether the organisation has the technical knowledge to evaluate whether a tool is performing as expected. For many debt advice organisations, particularly smaller ones, this may be the most challenging dimension, as they rely on external providers for technology and may lack in-house technical expertise.

This does not mean every organisation needs a technology team. It means understanding which questions to ask technology providers, which service level agreements should be in place, and what contingency plans exist if a tool becomes unavailable. Cyber Essentials certification, or progress towards it, provides a baseline for information security but should be seen as a minimum rather than sufficient assurance for AI deployment.

- ☐ Infrastructure requirements assessed
- ☐ Integration with existing systems understood
- ☐ Cyber Essentials in place or being worked towards
- ☐ Testing plan before full deployment
- ☐ Fallback procedures for when AI tools are unavailable

Workforce readiness

The introduction of AI tools affects everyone in the organisation. Advisers need to understand what tools do, what they do not do and how to identify when outputs are wrong. Managers need to understand how AI changes workflows and what new oversight responsibilities arise. Quality assurance teams need to know how to assess AI-assisted work alongside human-only work.

Equally important is the emotional and professional dimension. Sector research found significant anxiety among some advisers about job security, alongside enthusiasm among others about the potential to reduce administrative burden. Honest, direct communication about how AI will and will not affect roles is essential. Organisations that introduce AI tools without addressing workforce concerns risk undermining trust and adoption. The research also identified a specific risk to junior adviser development. If AI handles too much of the cognitive work too early, newer advisers may not develop the depth of expertise that experienced colleagues possess.

- ☐ Staff informed about AI plans and their role
- ☐ Training provided on use, review and error identification
- ☐ Professional accountability is understood as remaining with the human
- ☐ Job security concerns addressed through direct communication
- ☐ Impact on junior adviser development considered

Client base assessment

The suitability of AI tools, particularly client-facing ones, depends fundamentally on who the organisation serves. The research consistently found that there is no one-size-fits-all answer on whether clients can or should engage with AI tools. Organisations serving individuals in acute crisis with significant vulnerability, low digital confidence and complex intersecting issues will need to take a very different approach from those whose clients are digitally confident and present with relatively straightforward debt.

Multiple research participants estimated that approximately 50% of their clients would struggle to engage with digital self-serve tools of any kind. This proportion varies significantly by organisation type. The client base assessment should be honest and evidence-based. It should identify not just who might benefit from AI tools, but also who might be disadvantaged, excluded, or harmed, and what alternative provision will be maintained for those individuals.

- ☐ Digital confidence of client base assessed
- ☐ Proportion suitable for digital self-serve estimated
- ☐ Proportion for whom self-serve is inappropriate identified
- ☐ Plan for how excluded clients continue to access services
- ☐ Accessibility requirements considered, including reading age, language, and digital exclusion

Individual Client Suitability Assessment

Beyond organisational readiness, individual client suitability must be assessed at the point of access for any client-facing tool. Not every client who could use a self-serve tool should use one, and not every client who seems confident is well served by one. The following criteria provide a structured basis for real-time assessment.

Self-serve is likely appropriate when

- ☐ Individual demonstrates digital confidence
- ☐ Query relates to general information or straightforward circumstances
- ☐ No enforcement or crisis indicators
- ☐ Individual voluntarily chooses self-serve
- ☐ Responses are consistent and coherent

Self-serve should be offered with caution when

- ☐ Some digital skills, but hesitation or uncertainty
- ☐ Moderate complexity (multiple debts, no active enforcement)
- ☐ Previous AI use created expectations needing reassessment

Self-serve is not appropriate when

Route to a human adviser immediately for*:

- Active enforcement: bailiffs, court proceedings, possession, charging orders
- Crisis: imminent eviction, disconnection, benefit sanction
- Vulnerability: mental health distress, cognitive difficulty, domestic abuse, coercion
- Low digital confidence: struggling with the tool, frustration, confusion
- Complex intersecting issues: debt with housing crisis, benefit appeals, immigration
- Insolvency considerations: DRO, IVA or bankruptcy assessment
- Business or self-employment debt
- Recent/current debt solutions affecting eligibility
- Cannot provide consistent information or understand questions
- Individual requests human contact at any point

*Note, these will vary by organisation but are a starting point suggested by the research.

> *The overriding principle: self-serve should make support easier to access, not become a barrier. If there is any doubt, the route to human contact should be clear.*

7 Cross-Cutting Requirements

The following requirements apply across all levels. Regardless of whether an organisation is deploying a simple transcription tool at Level 1 or considering a personalised advisory interface at Level 3, these cross-cutting requirements represent the baseline standards that should be in place. They reflect the regulatory environment, the vulnerability of the client population and the professional obligations that organisations in the debt advice sector already hold.

Data protection and privacy

Data protection is not an optional add-on to AI deployment. It is a legal requirement and a fundamental safeguard for the individuals that debt advice organisations serve. The data processed in debt advice is among the most sensitive that exists: income, expenditure, health conditions, family circumstances, housing status and legal proceedings. AI tools that process this data must meet the same standards as any other system handling personal information, with additional scrutiny for the specific risks that AI introduces, including data being used for model training, processed outside the UK, or retained beyond the immediate purpose.

- ☐ DPIA completed for each tool processing personal data
- ☐ Appropriate legal basis for processing
- ☐ Data processing agreements with third parties
- ☐ Client data not used for general model training without informed consent
- ☐ Data minimisation applied
- ☐ Retention periods defined and enforced
- ☐ International transfer requirements met
- ☐ Staff understand data handling rules for each tool

Transparency with clients

Transparency is both a regulatory expectation and a trust-building requirement. The research found that the strongest factor in public trust of AI for financial guidance was knowing that it is regulated by financial authorities (32%), followed by knowing a human expert reviews the output (27%). Clients have a right to know when AI is playing a role in their advice journey, what that role is and how human oversight is maintained. Transparency should be proactive, clear and accessible, not buried in terms and conditions.

- ☐ Clients informed when AI is involved in their journey
- ☐ AI role explained in plain language
- ☐ Clients understand that a human adviser oversees their case
- ☐ Opt-out without disadvantage available
- ☐ AI use included in privacy notices

Accountability and governance

AI does not create an exception to existing accountability frameworks. If an AI tool contributes to a client outcome, the organisation remains responsible for that outcome. The FCA has made it clear that existing regulatory frameworks, including Consumer Duty obligations, fully apply to services that involve AI. Organisations cannot delegate accountability to a technology system. This means maintaining a register of AI tools in use, monitoring their performance, investigating errors and ensuring that lessons are shared rather than repeated.

- ☐ Named individual accountable for AI governance
- ☐ Register of all AI tools with level classification
- ☐ Regular reviews of tool performance and outcomes
- ☐ AI error/detriment incidents recorded and investigated
- ☐ Lessons shared across organisation and sector
- ☐ Consumer Duty met for all AI-mediated processes

Knowledge base management

The quality of any AI tool is only as good as the information it draws from. In debt advice, information changes frequently through case law, legislation, regulatory interpretation and policy updates. Scotland operates an entirely separate legal and debt solutions framework. The research found that maintaining currency is one of the sector's most significant ongoing challenges. Any AI tool, whether it supports advisers or interacts with clients, must draw from curated, approved and regularly updated sources rather than the open internet. Knowledge base governance is not a one-off setup task; it is an ongoing operational commitment.

- ☐ All tools draw from curated, approved sources
- ☐ Version-controlled with change logs
- ☐ Case law updates within five working days
- ☐ Jurisdictional differences maintained as separate knowledge sets
- ☐ Quarterly accuracy review through expert assessment
- ☐ Source attribution available for verification

Accessibility and inclusion

The debt advice client population includes individuals with low literacy, limited English, cognitive difficulties, mental health conditions, visual and hearing impairments and very low digital confidence. Any AI tool that interacts with clients must be designed with this population in mind, not as an afterthought. The research found that reading age is a particularly important consideration. Accessibility is not simply about compliance with WCAG standards; it is about ensuring that tools genuinely work for the people who need them most.

- ☐ Client-facing tools at an appropriate reading age
- ☐ Tested with a range of digital confidence levels

- ☐ Alternative channels remain available
- ☐ Language support available or signposted
- ☐ Design accounts for users in distress

8 Possible Implementation Roadmap

The following roadmap describes a possible sequencing for AI adoption. It is not prescriptive. Every organisation is different, and the pace, scope and extent of implementation will vary depending on organisational size, technical capacity, funding, client base and the specific use cases relevant to it. Some organisations may spend considerable time in Phase 1 and determine that this is where AI adds the most value for their context. Others may progress more quickly based on their readiness and client needs.

The critical point is that progression should be evidence-based. Moving from one phase to the next should be driven by demonstrated success, adequate governance and a clear user need, not by a desire to keep pace with other organisations or by external pressure to adopt technology.

Phase 1: Foundation

Recommended starting point for all organisations

- Deploy 1.1 tools (transcription, case notes, letters)
- Complete organisational readiness assessment across all six dimensions
- Develop and approve AI use policy
- Train staff
- Establish feedback and monitoring

> *For some organisations, Phase 1 may represent the full extent of AI adoption that is appropriate for their client base and operational context. This is a legitimate outcome.*

Phase 2: Expansion

When Phase 1 is established, and evidence supports progression

- Deploy 1.2 (knowledge/compliance support)
- Deploy 2.5–2.6 (client communication/documentation support)
- If the client base is suitable, pilot 2.1 with monitoring
- Develop knowledge base governance for client-facing tools

Phase 3: Structured self-serve

Evidence-based progression only

- Pilot 2.2–2.3 with limited caseloads and close monitoring
- Deploy 1.3 operational intelligence
- Deploy 2.7 ongoing engagement
- Establish outcome monitoring comparing AI-assisted and human-only

Phase 4: Advanced applications

Only with robust evidence, regulatory clarity and demonstrated governance capability

- Consider 2.4 based on pilot evidence
- Consider 3.2 adviser-directed agentic tools
- 3.1, 3.3 and 3.4 only with regulatory clarity and extensive piloting