ISSN: 3033-4136

DOI: https://doi.org/10.17868/strath.00095854

Agentic AI for Scaling Targeted Support:

A Governance Framework for the FCA Advice–Guidance Boundary

FRIL
Financial Regulation Innovation Lab

FinTech Scotland

University of Glasgow

University of Strathclyde Glasgow

Agentic AI for Scaling Targeted Support: A Governance Framework for the FCA Advice–Guidance Boundary

Mark Cummins, James Bowden, Hao Zhang, Kushagra Jain

Strathclyde Business School, University of Strathclyde, and Financial Regulation Innovation Lab, 199 Cathedral Street, Glasgow, G4 0QU, UK

26 March 2026

We acknowledge funding from Innovate UK, award number 10055559.

Corresponding authors:

Email: mark.cummins@strath.ac.uk

Email: james.bowden@strath.ac.uk

Email: hao.zhang@strath.ac.uk

Email: kushagra.jain@strath.ac.uk

Financial Regulation Innovation Lab

Who are we?

The Financial Regulation Innovation Lab (FRIL) is an industry-led collaborative research and innovation programme focused on leveraging new technologies to respond to, shape, and help evolve the future regulatory landscape in the UK and globally, helping to create new employment and business opportunities, and enabling the future talent.

FRIL provides an environment for participants to engage and collaborate on the dynamic demands of financial regulation, explore, test and experiment with new technologies, build confidence in solutions and demonstrate their ability to meet regulatory standards worldwide.

FRIL is part of the Glasgow City Region Innovation Accelerator programme, funded through Innovate UK on behalf of UK Research and Innovation. The Innovation Accelerator programme is investing £130 million in 26 transformative R&D projects to accelerate the growth of three high-potential innovation clusters, including the Glasgow City Region.

What is Actionable Research?

FRIL will integrate academic research with an industry relevant agenda, focused on enabling knowledge on cutting-edge topics such as generative and explainable AI, advanced analytics, advanced computing, and earth-intelligent data as applied to financial regulation. The approach fosters cross sector learning to produce a series of papers, actionable recommendations and strategic plans that can be tested in the innovation environment, in collaboration across industry and regulators.

Locally-led Innovation Accelerators delivered in
partnership with DSIT, Innovate UK and City Regions

Department for Science, Innovation & Technology

Innovate UK

GLASGOW CITY REGION

FRIL White Paper Series

Agentic AI for Scaling Targeted Support: A Governance Framework for the FCA Advice–Guidance Boundary

Mark Cummins* James Bowden* Hao Zhang * Kushagra Jain*

* University of Strathclyde

26 March 2026

Abstract

The Advice–Guidance Boundary Review (AGBR) introduces targeted support as a new regulated activity intended to address the persistent financial advice gap in the UK. While generative AI technologies offer the potential to scale accessible financial support, doing so within the advice–guidance boundary introduces significant governance challenges. Compliance requires structural control over segmentation logic, boundary monitoring, knowledge governance, vulnerability detection, and audit transparency. This white paper proposes an agentic AI governance framework that embeds these regulatory functions within the architecture of AI-enabled financial support systems. The framework distributes responsibility across specialised agents responsible for segmentation governance, boundary monitoring, vulnerability detection, knowledge management, and supervisory audit. By embedding compliance functions as interacting agents surrounding a multimodal generative AI interface, the proposed architecture transforms regulatory compliance from a behavioural expectation into a structural property of the system. The framework provides a conceptual foundation for scaling targeted pensions support safely and transparently under the FCA's AGBR while supporting responsible innovation in AI-enabled financial services.

Table of Contents

1. Introduction

The Advice Guidance Boundary Review (AGBR), led jointly by HM Treasury and the Financial Conduct Authority (FCA),[1] represents a significant reconfiguration of how financial support may be delivered to consumers in the United Kingdom. The review responds to a persistent structural advice gap, with millions of consumers underserved by existing markets. In particular, long-term financial decisions such as pensions present acute challenges, combining complexity, inertia, and low engagement with high long-run consequences. *Targeted support*, introduced by the FCA under AGBR as a new regulated activity, seeks to address the gap between generic guidance and regulated advice, where the former is insufficiently tailored to the specific needs of the consumer, while the latter is expensive and delivered by a limited pool of financial advisors. Targeted support enables authorised firms to provide appropriate suggestions to defined *groups of consumers*, based on limited but relevant information, without undertaking a full individualised suitability assessment.

The FCA has emphasised that firms must define consumer segments that are sufficiently precise to produce appropriate suggestions, while avoiding segmentation that effectively replicates an individual suitability assessment. It has also highlighted the importance of clear communications, appropriate treatment of vulnerable customers, and robust governance frameworks underpinning targeted support. These expectations imply that compliance cannot be treated as an afterthought. Rather, compliance must be embedded structurally within the delivery of targeted support.

Zhang et al. (2026) explores how frontier multimodal generative artificial intelligence could potentially serve as a delivery interface for scaling targeted support. That work provided preliminary evidence that speech-enabled, audio-visual digital advisors can improve accessibility, enhance consumer engagement, recognise vulnerability, and maintain discipline at the advice-guidance boundary. By combining curated knowledge bases with fine-tuned large language models, the proposed system illustrated how targeted support communications can be structured, auditable, and aligned with regulatory expectations. The contribution of that paper lay in showing how multimodal conversational architectures can translate regulatory intent into consumer-facing interactions that are compliant by design.

However, the introduction of targeted support as a regulated activity raises a further and deeper operational challenge that extends beyond the design of a single conversational interface. Scaling targeted support safely and consistently requires more than merely generating appropriate responses to individual queries. Rather, it requires the systematic management of consumer segmentation logic, advice-guidance boundary enforcement, knowledge base governance, consumer vulnerability monitoring, and audit logging and supervisory oversight. Each of these functions carries regulatory

[1] https://www.fca.org.uk/firms/advice-guidance-boundary-review

significance in its own right. If segmentation becomes too granular, it risks drifting toward individualised advice. If boundary discipline relies solely on prompt phrasing, it may become fragile under edge-case interactions. If knowledge sources are not continuously governed and version-controlled, factual accuracy and regulatory alignment may degrade over time. And if interactions cannot be transparently reconstructed and reviewed, firms may struggle to evidence compliance under supervisory scrutiny.

This defines the core problem addressed in this second white paper. Any viable next-stage solution must therefore reconcile four interrelated demands: high-volume, low-friction delivery; structural enforcement of the advice-guidance boundary; continuous knowledge and regulatory governance; and regulator-ready auditability. These demands move the discussion from the level of conversational intelligence to the level of workflow and systems design.

This problem framing motivates the exploration undertaken in the remainder of this paper. We propose that agentic AI architectures – composed of specialised, interacting agents responsible for segmentation, boundary monitoring, knowledge management, vulnerability detection, and audit generation – offer a coherent governance framework for scaling targeted support under the AGBR. An agentic system distributes responsibility across defined components, enabling compliance to become an architectural property of the system itself. The sections that follow develop this agentic AI governance framework, outline its regulatory implications, and position it as a conceptual foundation for scaling targeted support through AI innovation.

2. Solution Framework

2.1 Agentic AI in Finance

The rapid evolution of large language models (LLMs) has catalysed a transition from model-centric artificial intelligence toward agentic AI systems capable of autonomous perception, planning, memory, and tool use. In financial services, this shift is particularly significant. Whereas earlier AI deployments focused on predictive analytics and decision-support tools embedded within static workflows, agentic AI introduces systems that can coordinate multi-step reasoning processes, interact with heterogeneous data sources, collaborate with other agents, and adapt dynamically to regulatory and market environments.

Recent scholarship has begun to formalise this emerging paradigm. Okpala et al. (2025) demonstrate how specialised agent "crews" can collaboratively execute modelling and model risk management tasks. Their architecture distributes responsibility across role-defined agents. This work highlights a crucial insight: compliance and human oversight can be embedded explicitly within the system itself, rather than imposed externally.[2]

From a governance perspective, Kurshan, Balch, and Byrd (2025) argue that generative and multi-agent AI systems behave as complex adaptive systems that cannot be effectively governed through traditional static model risk frameworks. They propose a modular, layered governance architecture comprising: (i) self-regulation modules embedded beside each model, (ii) firm-level governance blocks that aggregate system activity data and enforce policy, (iii) regulator-hosted monitoring agents, and (iv) independent audit blocks. This layered structure is particularly instructive for regulated financial contexts. It suggests that oversight mechanisms must mirror the decentralised, adaptive properties of the systems they supervise. Governance must therefore become architectural rather than procedural.

At an implementation level, Jagannathan et al. (2025) introduce the ADAPT framework (Assess, Design, Assemble, Pilot, Transform) as a structured pathway for deploying agentic AI responsibly within highly regulated environments. Their roadmap emphasises strategic alignment, data governance, regulatory compliance, pilot testing, and controlled scaling. This contribution underscores that agentic systems in finance must be embedded within institutional governance and risk-management frameworks, rather than deployed purely as technological artefacts.

These works collectively reveal three core themes shaping the development of agentic AI in financial services:

1. **Workload Distribution**: AI systems increasingly decompose complex tasks into structured, multi-step processes coordinated across specialised agents.

2. **Embedded Governance:** Oversight, validation, and critique functions can be implemented as agents within the system itself.

3. **Architectural Compliance:** Traditional model risk management assumptions (static models, periodic validation, fixed datasets) are inadequate for adaptive, interacting agents; governance must evolve into modular, layered, and continuous supervision.

However, the AGBR context differs materially from the domains addressed in the above literature. Much of the current agentic finance research focuses on quantitative modelling and model risk management. Applications often involve high-frequency data, institutional decision-making, or internal control environments. By contrast, targeted support under the AGBR operates at the consumer interface, within a legally sensitive advice-guidance boundary. The primary risks are misclassification of support as advice,

[2] For further reading, interested readers are directed to Scott Cunningham's SubStack series on Claude Code

inappropriate segmentation, inadequate vulnerability recognition, and failure to evidence compliance. In this setting, the agentic challenge is structural boundary enforcement, segment discipline, consumer comprehension, and regulator-ready auditability. The system must demonstrate that it does not cross into personalised advice, even as it adapts dynamically to user interactions. It must also preserve clear lineage between regulatory policy, segmentation logic, and delivered suggestions. This distinction defines the contribution of the present white paper.

2.2 Agentic AI Types

Agentic AI systems differ in capability and architectural structure. Before distinguishing between single-agent and multi-agent architectures, it is useful to clarify what is meant by an "agent".[3] Although the term is widely used in current industry discussions, it can refer to slightly different concepts depending on the perspective adopted. Classical AI literature typically categorises agents according to their decision-making capabilities, whereas more recent industry frameworks, particularly those associated with large language model developers such as OpenAI and Anthropic, tend to describe agents in terms of system architecture and task orchestration.

A widely cited practitioner taxonomy, summarised by IBM (see footnote 3), identifies five principal types of AI agents: simple reflex agents, model-based reflex agents, goal-based agents, utility-based agents, and learning agents. While this taxonomy originates from classical AI research, it remains useful for understanding how different agent capabilities may be implemented within modern LLM-based systems.

Simple reflex agents operate using condition–action rules without maintaining memory. They respond directly to current inputs and are suitable for well-defined, repetitive tasks. **Model-based reflex agents** extend this structure by maintaining an internal representation of the environment, allowing decisions to incorporate contextual and historical information. **Goal-based agents** introduce explicit planning, selecting actions that advance defined objectives. **Utility-based agents** evaluate competing outcomes according to a utility function, balancing trade-offs across multiple criteria. **Learning agents** adapt over time by incorporating feedback and updating internal models.

In practice, modern LLM-enabled systems often combine elements of these paradigms. A conversational system may be model-based in its contextual reasoning, goal-oriented in its instruction-following, and partially adaptive through reinforcement or supervised fine-tuning. Within multi-agent systems, different agents may embody different types simultaneously.

This taxonomy is important because it clarifies that "agentic AI" is not a single, uniform concept. Different agent types introduce different governance implications. Reflex agents are predictable but inflexible. Goal- and utility-based agents introduce

[3] https://www.ibm.com/think/topics/ai-agent-types

optimisation capabilities but increase behavioural complexity. Learning agents introduce adaptability but also raise concerns regarding drift and explainability. With this foundation in place, architectural structure becomes central.

Single-Agent Architectures

A single-agent system represents the most compact form of agentic AI. In this configuration, one LLM-powered agent is responsible for perceiving inputs, planning responses, invoking tools, maintaining memory, and generating outputs. The system may include retrieval layers, guardrails, and internal reflection mechanisms, but these functions are orchestrated within a unified decision loop.

Such architectures are attractive because they are comparatively simple to deploy and monitor. A single agent can be equipped with structured prompts, access to a curated knowledge base, and defined constraints on output formatting. When combined with retrieval-augmented generation and runtime guardrails, this model can generate compliant, well-structured responses in many use cases. However, the concentration of responsibility within a single agent introduces structural limitations. Oversight becomes internal rather than independent. From a governance perspective, this can create difficulties in evidencing separation of duties or demonstrating that compliance checks operate independently of the generative process. In domains such as content summarisation or internal productivity tools, these limitations may be acceptable. In regulated financial contexts, they are materially more consequential, particular those involving legally sensitive distinctions such as the advice–guidance boundary.

The multimodal Digital Pensions Advisor presented in Zhang et al. (2026) can be understood as a structured single-agent architecture. Although augmented with retrieval mechanisms, guardrails, and multimodal interface layers, governance functions such as segmentation discipline, suggestion formulation, and boundary compliance were orchestrated within a unified reasoning loop. The system demonstrated that compliant targeted support delivery is feasible through careful design. However, as deployment scales, the concentration of governance functions within a single agent may limit the structural separability and audit clarity required for supervisory robustness.

Multi-Agent Architectures

Multi-agent systems distribute responsibility across multiple specialised agents that collaborate to achieve a shared objective. Each agent is assigned a defined role, with explicit task boundaries and communication protocols. Rather than one model performing all functions, the workflow is decomposed into interacting components.

This architecture mirrors institutional organisational structures. Financial institutions routinely separate customer engagement, risk assessment, compliance oversight, and audit functions. Multi-agent systems can replicate this separation digitally. Multi-agent design enables this mirroring. Multi-agent systems also allow different IBM-style agent types to coexist and collaborate within a single governance framework.

Structural Coordination

Multi-agent systems themselves may adopt different coordination forms. Some are sequential, in which agents execute tasks in a predefined order. Others are hierarchical, with supervisory or “manager” agents overseeing subordinate agents and validating outputs before progression. More advanced systems employ hybrid or asynchronous structures, allowing agents to operate concurrently, triggering interventions when predefined conditions are met.

The choice of structure determines how authority, escalation, and intervention operate. A purely sequential system provides clarity and audit simplicity but may lack responsiveness. A hierarchical system introduces explicit oversight authority but risks concentrating decision-making in a single supervisory node. Hybrid architectures distribute monitoring functions across parallel agents, increasing robustness but also coordination complexity.

In regulated financial environments, structural clarity is often as important as technical capability. Supervisors require reconstructible workflows and clearly defined responsibility boundaries. Therefore, architectural simplicity and traceability must be weighed against adaptability and autonomy.

Implications for the AGBR Context

In the context of targeted support under the AGBR, these architectural distinctions take on particular significance. The core regulatory risk is the inadvertent crossing of the advice–guidance boundary, inappropriate segmentation granularity, inadequate vulnerability recognition, or insufficient evidencing of compliance. A single conversational agent, however well designed, concentrates activities within one reasoning loop. By contrast, a multi-agent architecture permits segmentation, suggestion formulation, boundary monitoring, vulnerability detection, and audit logging to be separated into specialised, interacting functions. This separation does more than improve technical robustness. It enables the system to demonstrate that targeted support remains segment-level rather than individualised advice; that compliance monitoring is independent of suggestion generation; and that regulatory obligations are operationalised as structural constraints rather than informal guidelines.

For these reasons, this white paper proceeds on the premise that multi-agent architectures are more appropriate for scaling targeted support within the advice–guidance boundary. The next subsection builds on this foundation by identifying the specific agent roles and governance functions required to operationalise targeted support under the AGBR.

2.3 Agentic AI for Scaling Targeted Support

The multimodal Digital Pensions Advisor presented in Zhang et al. (2026) demonstrated that generative AI can deliver targeted support through a controlled, retrieval-augmented interface. However, as discussed in Section 2.2, that architecture concentrated segmentation logic, suggestion generation, vulnerability handling, and boundary discipline within a unified reasoning loop. While sufficient for a prototype or limited deployment, scaling targeted support across millions of consumers requires a more distributed and governance-oriented design.

An agentic architecture for scaling targeted support therefore begins with the principle of functional separation. Distinct regulatory obligations should correspond to distinct system roles. Rather than a single agent interpreting segmentation rules, generating suggestions, monitoring compliance, and logging audit data simultaneously, these responsibilities can be distributed across specialised agents operating within a coordinated framework.

At a conceptual level, at least five core governance functions emerge in the AGBR context.

1. Segmentation Governance

Targeted support is explicitly defined as operating at the level of consumer segments rather than individuals. The integrity of segmentation is therefore foundational. An agentic system must ensure that: segments are defined using limited, relevant characteristics; segmentation granularity does not drift toward personalised suitability assessment; and changes in segmentation logic are documented and version-controlled.

A **Segmentation Agent** would operationalise these requirements. Its responsibility would be to define, validate, and apply segment criteria based on approved policy parameters. This creates a structural boundary between identifying a consumer’s segment and determining the appropriate suggestion for that segment. Such separation mirrors internal governance practices within financial institutions, where customer classification, risk scoring, and advisory recommendations are often managed by distinct functions. In an agentic framework, this separation becomes programmable and auditable.

IBM Agent Type Mapping: The Segmentation Agent most closely resembles a **simple reflex agent**, potentially augmented with limited model-based context. Its function is rule-constrained and policy-driven, with the segmentation task being repetitive. It applies predefined segmentation criteria to observed inputs without pursuing independent optimisation goals. Predictability and determinism are regulatory virtues here. By constraining segmentation logic to rule-based or tightly model-bound behaviour, the system reduces the risk of granularity drift toward individualised advice. In this context, flexibility is less important than traceability.

2. Boundary Monitoring and Advice Drift Prevention

The advice–guidance boundary remains the central regulatory risk. Even if segmentation and suggestion generation are well-designed, language drift or contextual interaction could inadvertently move toward personalised advice. An agentic architecture therefore benefits from an independent **Boundary Agent**. This agent would review generated outputs prior to delivery, assessing whether language, framing, or reasoning implies individualised suitability assessment. Unlike guardrails embedded within the generative agent, this monitoring function would be structurally distinct, with the authority to request revision, constrain phrasing, or escalate to human oversight.

IBM Agent Type Mapping: The Boundary Agent aligns most closely with a **model-based reflex agent**. It must maintain an internal representation of regulatory expectations and contextual conversational cues. Unlike a simple reflex agent, it cannot rely solely on static condition–action rules; advice drift often emerges subtly in phrasing and cumulative interaction context. The Boundary Agent therefore requires contextual interpretation while remaining constrained by regulatory objectives. Its objective is solely regulatory containment.

3. Vulnerability and Consumer Protection

The FCA's behavioural research emphasises the importance of clarity, consumer understanding, and the appropriate handling of vulnerability.[4] A scaled targeted support system must therefore continuously monitor interaction signals for indicators of confusion, distress, or vulnerability. A dedicated **Vulnerability Agent** could operate in parallel with the interaction flow, analysing linguistic cues and behavioural signals. Where thresholds are met, the system could modify its communication style, provide additional explanatory context, or escalate to human support. Separating this function from suggestion generation reduces the risk that performance optimisation objectives conflict with consumer protection priorities. **IBM Agent Type Mapping:** The Vulnerability Agent most naturally reflects a **learning agent**, potentially combined with model-based reasoning. Vulnerability detection often depends on pattern recognition across language use, sentiment, interaction dynamics, and behavioural signals. A learning-based structure allows the agent to improve detection thresholds over time, subject to governance controls. However, because learning introduces drift risk, its adaptive capacity must itself be monitored and version-controlled. In this sense, the Vulnerability Agent embodies adaptive sensitivity, but within bounded governance.

4. Knowledge Governance and Regulatory Traceability

Targeted support suggestions must be grounded in approved and up-to-date regulatory policy and interpretations. In a scaled environment, regulatory change is continuous. Without structured oversight, knowledge bases may become outdated or inconsistent. A **Knowledge Agent** would manage retrieval sources, enforce version control, and log the

[4] https://www.fca.org.uk/firms/advice-guidance-boundary-review

policy basis for each generated suggestion. This creates traceability between regulatory source material and delivered communication. Over time, such traceability becomes essential for supervisory review and audit reconstruction.

IBM Agent Type Mapping: The Knowledge Agent most closely resembles a **model-based reflex agent**, operating with structured memory and environmental representation. It maintains an internal model of approved policy sources and regulatory interpretations, ensuring that retrieval is constrained to validated materials. Its function is to preserve informational integrity. In environments where regulatory updates occur, the Knowledge Agent may also incorporate limited goal-based elements, e.g. ensuring alignment with current policy states, but its core design remains model-based and governance-oriented rather than optimisation-driven.

5. Audit and Supervisory Interface

Finally, scaling targeted support requires regulator-ready transparency. Supervisors must be able to reconstruct how a particular suggestion was generated, which segment logic applied, which boundary checks were performed, and which knowledge sources were invoked. An **Audit Agent** can consolidate information from other agents, generating structured logs and supervisory summaries. Rather than treating audit as a downstream reporting exercise, it becomes an integrated, real-time function.

IBM Agent Type Mapping: The Audit Agent aligns most closely with a **utility-based agent**, potentially augmented by learning capabilities. It evaluates interactions against multiple criteria: boundary compliance, segmentation integrity, knowledge traceability, and anomaly detection. Where deviations are detected, it may escalate or flag for review. Its “utility function” is specified in terms of governance robustness and supervisory transparency. Over time, anomaly detection mechanisms may incorporate learning components, improving sensitivity to emerging risk patterns while preserving reconstructibility.

Architectural Implication

Mapping these five governance functions to recognised agent types underscores the importance of a multi-agent architecture. Reflex-style determinism governs segmentation. Model-based contextual reasoning governs boundary and knowledge oversight. Learning-based adaptation enhances vulnerability detection. Utility-style evaluation supports audit integrity. By aligning agent type with regulatory function, governance becomes structurally embedded. Targeted support under the AGBR is therefore delivered by a coordinated system of specialised agents whose behavioural characteristics are deliberately matched to their regulatory responsibilities. Note, however, that in the framework proposed here, these agents operate within a largely flat coordination structure rather than a defined hierarchy. Each agent performs a distinct governance function and interacts with the others through defined interfaces. This design emphasises separation of duties and transparency over centralised decision authority. The need for a hierarchical structure can be considered by the organisation deploying the framework.

The next section builds upon this conceptual architecture to outline how such an agentic framework could be mapped onto the local multimodal AI deployment described in Zhang et al. (2026), illustrating how governance functions and interface delivery can be integrated within a coherent operational workflow.

3. Agentic AI Governance Framework for Scaling Targeted Pensions Support

The multimodal Digital Pensions Advisor presented in Zhang et al. (2026) established an important proof of concept: targeted pensions support can be delivered through a locally deployed (or cloud deployed), retrieval-augmented multimodal generative AI system while remaining within the advice–guidance boundary. The workflow – live voice input, speech-to-text conversion, avatar-mediated interaction, LLM response generation, conditional retrieval from a curated knowledge base, response merging, and text-to-speech output – demonstrated that compliant support can be structured at the interface layer.

This section reinterprets the original multimodal workflow as the interaction layer of a broader governance architecture rather than as a complete system. The objective is to position the Digital Pensions Advisor within an agentic framework in which compliance obligations are distributed across specialised roles and continuously monitored, in line with the agent suggestions of the previous section.

In its original configuration, the local deployment workflow functioned as a largely linear processing chain. Consumer input entered via voice, was converted to text, passed into a generative model augmented by retrieval, and returned to the user through an avatar interface. Governance mechanisms – segmentation discipline, boundary definition, vulnerability monitoring and knowledge base control – were embedded within the generative layer through prompt design and curated retrieval sources. At the centre of the agentic governance framework lies the principle of functional separation. Building the Segmentation Agent, Boundary Agent, Vulnerability Agent, Knowledge Agent and Audit Agent around the Multimodal AI Advisor, provides a stronger governance infrastructure around any consumer-avatar interaction. Figure 1 provides a summary visualisation of the agentic governance framework.

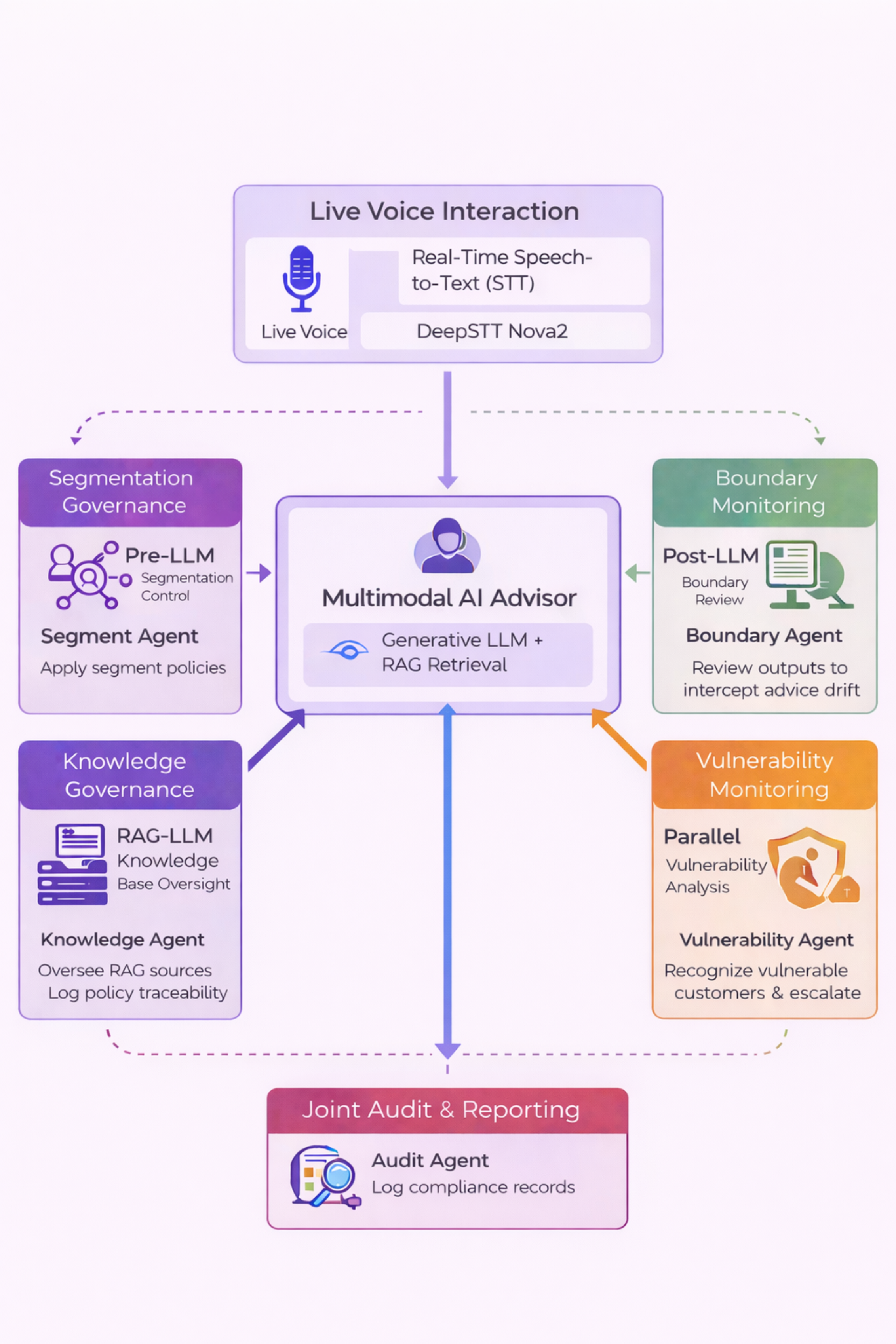

Figure 1 Agentic Governance Framework: Multimodal AI Pensions Advisor

It is important to emphasise that this governance framework does not displace the multimodal Digital Pensions Advisor. The live voice interface, speech-to-text processing, avatar-mediated delivery, generative LLM core, retrieval layer, response merging, and text-to-speech output remain intact. The consumer-facing experience does not fundamentally change. What changes is the architectural embedding of governance around that experience. The original workflow demonstrated that, in principle, targeted support can be achieved. The agentic framework ensures that compliance is monitored, and evidenced across scale.

In effect, the Digital Pensions Advisor becomes the visible surface of a deeper system. Beneath the conversational interface, segmentation is validated before suggestion, boundary discipline is verified before delivery, vulnerability is monitored continuously, knowledge sources are controlled systematically and updated regularly, and audit logs are generated in real time. Compliance shifts from being a property of prompt design to a property of system structure.

4. Conclusion

The AGBR represents one of the most significant structural reforms in the delivery of retail financial support in recent decades. By introducing targeted support as a regulated activity, the FCA has created an opportunity to expand meaningful decision assistance to millions of consumers who are currently underserved by the traditional advice market. Yet this opportunity is inseparable from a governance challenge. Scaling targeted support requires not only technological capability, but structural discipline.

Zhang et al. (2026) demonstrated that multimodal generative AI can serve as a promising delivery interface for targeted pensions support. Through a locally deployed, retrieval-augmented architecture, the Digital Pensions Advisor showed that conversational AI can operate within the advice-guidance boundary while maintaining clarity, traceability, and behavioural sensitivity. That work established feasibility at the interface layer.

This white paper has extended that foundation by addressing the architectural layer. Rather than concentrating segmentation logic, suggestion generation, boundary monitoring, vulnerability detection, and audit logging within a single reasoning loop, we have proposed an agentic governance framework that distributes these functions across specialised agents. In doing so, compliance ceases to be an implicit property of prompt design and becomes an explicit property of system structure.

The proposed framework rests on three core principles.

First, functional separation ensures that segmentation, suggestion formulation, boundary monitoring, knowledge governance, and vulnerability detection operate as distinct responsibilities. This mirrors institutional governance structures and reduces the risk of advice drift or segmentation creep.

Second, architectural oversight embeds compliance within the workflow itself. Monitoring agents operate alongside the generative core rather than within it, and audit inputs are produced in real time rather than reconstructed retrospectively. Governance becomes observable.

Third, scalable traceability connects each interaction to its segment classification, policy source, boundary review outcome, and audit record. This enables regulator-ready transparency as targeted support expands across consumer populations.

This paper stops short of full technical implementation. The objective has been to articulate a design framework aligned with the FCA's AGBR principles and to demonstrate how agentic architectures can support responsible innovation in pensions. Future work may explore empirical validation, stress-testing of boundary-monitoring agents, supervisory sandbox experimentation, and formal integration with firm-level compliance systems.

5. References

Jagannathan, S., Sridhar, S., Gulkotwar, N., Baskar, P. and Tambe, A. (2025) ‘A roadmap for agentic AI in financial services industry’, SSRN working paper. Available at: http://dx.doi.org/10.2139/ssrn.5392281.

Kurshan, E., Balch, T. and Byrd, D. (2025) ‘The agentic regulator: risks for AI in finance and a proposed agent-based framework for governance’, arXiv preprint, arXiv:2512.11933. Available at: https://doi.org/10.48550/arXiv.2512.11933.

Okpala, I., Golgoon, A. and Kannan, A.R. (2025) ‘Agentic AI systems applied to tasks in financial services: modeling and model risk management crews’, arXiv preprint, arXiv:2502.05439. Available at: https://doi.org/10.48550/arXiv.2502.05439.

Zhang, H., Bowden, J. and Cummins, M. (2026) ‘Multimodal AI for scaling targeted support: navigating the FCA advice–guidance boundary’, FRIL White Paper Series, University of Strathclyde. Available at: https://www.strath.ac.uk/media/departments/accountingfinance/fril/whitepapers/Multimodal_AI_.pdf.

6. About the Authors

Professor Mark Cummins is Professor of Financial Technology at the Strathclyde Business School, University of Strathclyde, where he leads the FinTech Cluster as part of the university's Technology and Innovation Zone leadership and connection into the Glasgow City Innovation District. As part of this role, he is driving collaboration between the FinTech Cluster and the other strategic clusters identified by the University of Strathclyde, in particular the Space, Quantum and Industrial AI Clusters. Professor Cummins is the lead investigator at the University of Strathclyde on the newly funded (via UK Government and Glasgow City Council) Financial Regulation Innovation Lab initiative, a novel industry project under the leadership of FinTech Scotland and in collaboration with the University of Glasgow. He previously held the posts of Professor of Finance at the Dublin City University (DCU) Business School and Director of the Irish Institute of Digital Business. Professor Cummins has research interests in the following areas: financial technology (FinTech), with particular interest in Explainable AI and Generative AI; quantitative finance; energy and commodity finance; sustainable finance; model risk management. Professor Cummins has over 50 publication outputs. He has published in leading international discipline journals such as: European Journal of Operational Research; Journal of Money, Credit and Banking; Journal of Banking and Finance; Journal of Financial Markets; Journal of Empirical Finance; and International Review of Financial Analysis. Professor Cummins is co-editor of the open access Palgrave title *Disrupting Finance: Fintech and Strategy in the 21st Century*. He is also co-author of the Wiley Finance title *Handbook of Multi-Commodity Markets and Products: Structuring, Trading and Risk Management*.

Dr James Bowden is Senior Lecturer in Financial Technology at Strathclyde Business School, University of Strathclyde, where he is the programme director of the MSc Financial Technology. Prior to this, he gained experience as a Knowledge Transfer Partnership (KTP) Associate at Bangor Business School, and he has previous industry experience within the global financial index team at Russell Investments (now FTSE Russell). Dr Bowden's research focuses on different areas of financial technology (FinTech), and his published work involves the application of text analysis algorithms to financial disclosures, news reporting, and social media. More recently he has been working on projects incorporating audio analysis into existing financial text analysis models and investigating the use cases of satellite imagery for the purpose of corporate environmental monitoring. Dr Bowden has published in respected international journals, such as the European Journal of Finance, the Journal of Comparative Economics, and the Journal of International Financial Markets, Institutions and Money. He has also contributed chapters to books including "Disruptive Technology in Banking and Finance", published by Palgrave Macmillan. His commentary on financial events has previously been published in The Conversation UK, the World Economic Forum, MarketWatch and Business Insider, and he has appeared on international TV stations to discuss financial innovations such as non-fungible tokens (NFTs).

Dr. Hao Zhang is a Research Associate at the Financial Regulation Innovation Lab (FRIL), University of Strathclyde. He holds a PhD in Finance from the University of Glasgow, Adam Smith Business School. Hao held the position of Senior Project Manager at the Information Center of the Ministry of Industry and Information Technology (MIIT) of the People's Republic of China. His recent research has focused on asset pricing, risk management, financial derivatives, intersection of technology and data science.

Dr. Kushagra Jain is a Research Associate at the Financial Regulation Innovation Lab (FRIL), Strathclyde Business School, University of Strathclyde. His research interests include artificial intelligence, machine learning, financial technology, regulatory technology, international finance, risk management, and asset pricing, among others. He is a recipient of doctoral scholarships from the Financial Mathematics and Computation Cluster (FMC), Research Ireland (formerly Science Foundation Ireland (SFI)), Higher Education Authority (HEA) and Michael Smurfit Graduate Business School, University College Dublin (UCD). Previously, he worked in wealth management and as a statutory auditor. He completed his doctoral studies in Business (Banking and Finance) from UCD, and obtained his MSc in Finance from UCD, his Accounting Technician professional accreditation from the Institute of Chartered Accountants of India and his undergraduate degree from Bangalore University. He is an FMC Database Management Group Data Manager, and was formerly Research Assistant, PhD Representative and Teaching Assistant for undergraduate, graduate and MBA programmes at UCD.

Get in touch

FRIL@FinTechScotland.com

FinTech Scotland®

University of Glasgow

University of Strathclyde Glasgow