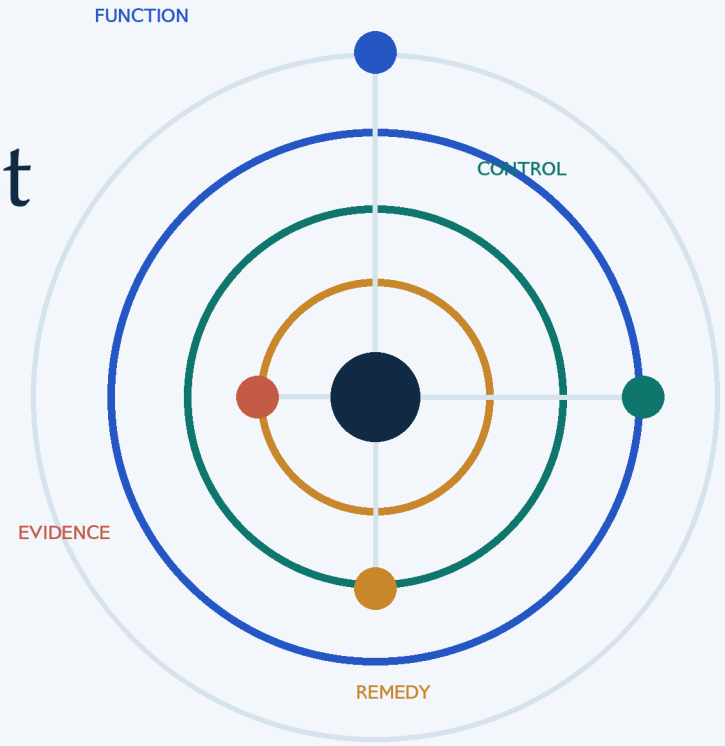
DigiData
CONSULTING LTD

PUBLIC POLICY SUBMISSION

Consumer Interest and AI



Phase 2 Response

A Consumer AI Accountability Framework

Tools, mandates, evidence, accountability and tolerable risk

01 CONSENT IS NOT A MANDATE.

Data consent, AI assistance and authority to act require separate controls.

02 TOLERABLE RISK IS NOT TRANSFERRED RISK.

Use or formal consent cannot transfer loss that a consumer cannot inspect or contest.

THE CENTRAL PROPOSITION

Regulate the change an AI enabled service makes to the consumer environment, not the label attached to the technology.

DR M.R. LEISER (MARK)
DigiData Consulting Ltd | personal capacity

18 August 2026

Fierce compliance. Radical clarity.

digidata.uk

POSITION AT A GLANCE

Ten Propositions for Consumer AI Regulation

Existing law needs an operational architecture for adaptive, delegated and difficult-to-inspect consumer journeys.

01

Regulate the Consumer Change, Not the AI Label.

Function and effect determine the starting point for safeguards.

02

Consent Is Not a Mandate.

Data consent, AI assistance and authority to act are separate controls.

03

Consequential Action Needs a Route Back.

Every material agentic act needs bounded authority, a record and reversal.

04

Optimisation Belongs in Consumer-Law Scrutiny.

Inspect reward functions, KPIs, prompts, variants and incentives.

05

Vulnerability Should Trigger Support, Never Extraction.

Test the moment when design meets circumstance.

06

The Consumer-Facing Firm Remains Primarily Accountable.

Allocate responsibility upstream after protecting the consumer.

07

Outcomes-Based Regulation Must Remain Evidence-Based.

Principles guide judgement; records make it testable.

08

Safe Harbours Must Be Earned.

Boundedness, reversibility, inspectability and redress justify lighter duties.

09

Tolerable Risk Is Not Transferred Risk.

Use cannot legitimise loss that the consumer cannot inspect or contest.

10

Cooperation Is Innovation Infrastructure.

Shared evidence artefacts reduce duplication across mandates.

Submission Details

SCOPE	STATUS	CAPACITY
Phase 2 questions 17–28	18 August 2026	Dr M.R. Leiser, Director

1. Scope, Method and Orientation of This Phase 2 Submission

This submission addresses the DRCF Phase 2 questions on tools, frameworks, standards, accountability and tolerable risk. The call seeks empirically grounded and practical proposals for consumer-facing services and back-office uses that materially affect consumers.[1]

The submission builds on my Phase 1 position that the regulatory object should not be "AI" in the abstract, but the change that an AI-enabled service makes to the consumer environment. Consumer AI becomes legally and policy-relevant when it alters information, timing, ranking, prominence, price, friction, advice, trust, transaction flows, contractual options, complaints, support or redress; the relevant question is therefore not whether a model belongs to a particular technical family, but which consumer conditions the service changes, who controls that change and who benefits from it.[2]

The method used here is deliberately layered. It combines large UK consumer studies that used quotas and weighting to approximate nationally representative proportions, experimental evidence about digital choice architecture, current consumer and data protection law, sectoral regulatory practice, and international risk-management standards; where those sources answer different questions, the submission keeps them separate rather than treating them as interchangeable proof. Survey evidence can show reported use, understanding, concern and stated willingness, but it cannot by itself establish how consumers will behave in a live market; controlled experiments can identify causal effects of interface design, but they do not reproduce the full complexity of adaptive AI services; and management standards can improve organisational process without proving that a particular consumer outcome is fair.[3]

This submission is organised around proportional accountability, under which consumer-facing AI should neither be romanticised nor governed by technological label alone, because the same broad class of systems can support accessibility, translation, comparison, customer service and product quality while also enabling personalised opacity, simulated authority, vulnerability exploitation, unauthorised agency, weak contestability and diffuse responsibility. The regulatory objective should be to preserve beneficial assistance while preventing firms from converting consumer inattention, dependency, urgency, low literacy or situational vulnerability into commercial advantage.[4]

This orientation supports the United Kingdom's innovation agenda. Predictable, risk-calibrated accountability lowers uncertainty for firms developing genuinely assistive services while reserving intensive evidence and control obligations for systems that affect money, rights, eligibility, essential services or legal status. The Government's AI Opportunities Action Plan and the DRCF's 2026/27 Workplan both connect usable regulation, adoption, trust and growth. That commitment still requires operational standards that regulators and firms can apply to real consumer journeys.[5]

The operational case has sharpened since the July draft. The CMA has issued a final infringement notice and a £720,000 penalty in an online choice-architecture case, and it has opened an investigation into the presentation of Microsoft 365 subscription options following the addition of features such as Copilot.[6] Ofgem has closed a call

[1] Digital Regulation Cooperation Forum, 'Call for Input: Consumer Interest and AI' (3 June 2026), especially Q17–Q28; Digital Regulation Cooperation Forum, DRCF Workplan 2026/27 (27 April 2026).

[2] M R Leiser, Submission to the Digital Regulation Cooperation Forum: Consumer Interest and AI, Phase 1 Response (June 2026); M R Leiser, 'Against the Sovereign User: Fundamental Rights, Technology and the Case for Relational Rights' (SocArXiv preprint, 2026).

[3] Thinks Insight & Strategy, Understanding Consumer Use of Generative AI: Report for the Digital Regulation Cooperation Forum (April 2025) 4–9; Financial Conduct Authority, The Mills Review: AI and the Future of Retail Financial Services (July 2026) 36–40, 125–129; Midas Nouwens and others, 'Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence' (2020) Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1; ISO/IEC 42001:2023.

[4] Financial Conduct Authority, Guidance for Firms on the Fair Treatment of Vulnerable Customers (FG21/1, February 2021) paras 2.4–2.17; Competition and Markets Authority, Online Choice Architecture: How Digital Design Can Harm Competition and Consumers (April 2022); OECD, Dark Commercial Patterns (OECD Digital Economy Papers No 336, 2022).

[5] Department for Science, Innovation and Technology, AI Opportunities Action Plan (CP 1241, January 2025) paras 221–228; Department for Science, Innovation and Technology, 'AI Opportunities Action Plan: Government Response' (13 January 2025); Digital Regulation Cooperation Forum, DRCF Workplan 2026/27 (27 April 2026).

[6] Competition and Markets Authority, Marks Electrical: Consumer Protection Enforcement Case (final infringement notice 15 June 2026; non-confidential notice published 2 July 2026); Competition and Markets Authority, Microsoft: Consumer Protection Enforcement Case (opened 27 July 2026). No finding had been made in the Microsoft investigation at 18 August 2026.

on outcomes-based AI assurance and will establish a technical sandbox.[7] The ICO continues to draft updated automated-decision and agentic-AI guidance.[8] These developments do not justify another silo. They strengthen the case for a modular DRCF layer that makes evidence, controls and redress interoperable across existing regimes.

2. Executive Summary

1	Consumer AI protection should be organised around *function* and *effect* rather than product labels. A shared DRCF risk grammar should distinguish assistive, advisory, consequential, agentic and systemic consumer AI, while safeguards rise with autonomy, materiality, opacity, vulnerability, irreversibility, scale and evidence asymmetry.[9]
2	The central tool should be a *Consumer AI Accountability Framework* built around classification, purpose, control, evidence, outcomes, redress and cooperation. It should require risk-calibrated evidence packs rather than one compliance template.
3	Agentic AI requires a mandate and control plane, not a general notice. Where a system can spend money, accept terms, disclose data, cancel services, negotiate, complain or move funds, the consumer should grant granular, revocable and time-limited authority. The service should produce an intelligible action record and support rapid reversal.[10]
4	Consent and disclosure remain necessary where law requires them, but they cannot carry the principal weight of protection. Consent to data processing is not agreement to AI-mediated assistance, and neither constitutes authority for an agent to undertake a legally or economically consequential act.[11]
5	Optimisation objectives require direct scrutiny. Firms should prohibit complaint abandonment, cancellation friction, vulnerable-consumer conversion, revenue through confusion, suppression of statutory rights and strategic withholding of a human route.
6	Vulnerability is a relationship between circumstances and design, not a closed list. Firms should test vulnerability events and use detection to support consumers, never to increase price, pressure, friction or extraction.[12]
7	The consumer-facing firm should remain the primary accountable actor. Unauthorised agentic action should trigger rapid no-fault correction, while other inaccessible failures may justify a rebuttable evidential presumption.
8	The DRCF should pair earned safe harbours for bounded and reversible uses with qualitative red lines, shared metrics, synthetic-user testing, a cross-regulatory intake route and a 12-month implementation programme.[13]

[7] Ofgem, 'AI Assurance in the Energy Sector' (call for input, 17 June 2026; closed 12 August 2026); Ofgem, 'AI Technical Sandbox' (consultation outcome, 3 June 2026; decision to establish a 12-month pilot targeting commencement in late autumn 2026).

[8] Information Commissioner's Office, 'Technology: Our Plans for New and Updated Guidance', listing final agentic AI guidance and updated automated decision-making and profiling guidance for Winter 2026.

[9] Digital Regulation Cooperation Forum, The Future of Agentic AI: Foresight Paper (31 March 2026); Competition and Markets Authority, 'Agentic AI and Consumers' (9 March 2026).

[10] Financial Conduct Authority, The Mills Review: AI and the Future of Retail Financial Services (July 2026) 35–40, 126–129; Competition and Markets Authority, 'Complying with Consumer Law When Using AI Agents' (9 March 2026).

[11] Daniel J Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880; Omri Ben-Shahar and Carl E Schneider, More Than You Wanted to Know: The Failure of Mandated Disclosure (Princeton University Press 2014); Nouwens and others, 'Dark Patterns after the GDPR'.

[12] Financial Conduct Authority, Guidance for Firms on the Fair Treatment of Vulnerable Customers (FG21/1, February 2021); Financial Conduct Authority and Information Commissioner's Office, 'Joint FCA and ICO Statement on Regulatory Expectations Regarding Firms' Approaches to Vulnerability Related Data' (27 March 2026, updated 20 July 2026).

[13] Competition and Markets Authority, Online Choice Architecture; Ofcom, 'Ofcom Fines Virgin Media £28m for Repeatedly Preventing Customers from Cancelling Contracts' (8 July 2026); National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1, January 2023).

3. Core Recommendations

Recommendation	Practical Implications
A Functional Risk Ladder	Distinguish assistive, advisory, consequential, agentic and systemic consumer AI. Treat autonomy, materiality, opacity, vulnerability, irreversibility, scale and evidence asymmetry as risk escalators.
A Consumer AI Accountability Framework	Require proportionate evidence across classification, purpose, objectives, control, testing, monitoring, incidents, outcomes, redress and cooperation. Use short, standard and enhanced evidence packs.
Prohibited Optimisation Objectives	State what a system must not optimise for, including complaint abandonment, cancellation friction, vulnerable-consumer conversion, revenue through confusion, suppression of statutory rights and strategic withholding of a human route.
Vulnerability-Event Controls	Test financial difficulty, bereavement, disability, distress, time pressure, dependency on an essential service and low understanding. Vulnerability should trigger support, never extraction.
An Agentic Mandate and Control Plane	Use authority envelopes, contemporaneous approval, spending and counterparty limits, data-disclosure limits, prohibited acts, expiry, revocation, pausing, action logs and consumer-readable receipts.
Rights-Friction Parity	Measure whether cancellation, complaint, escalation and redress require more time, steps or transfers than acquisition, purchase or sign-up.
Redress by Design	Preserve evidence, provide meaningful human escalation, support pausing and reversal, measure correction latency and deliver no-fault correction where a firm-controlled agent acts outside mandate.
Regulatory Access to Evidence	Allow lawful access to instructions, versions, A/B testing, objectives, journey tests, supplier documentation, incidents and redress outcomes, subject to due process and confidentiality.
Earned Safe Harbours	Publish criteria for bounded, reversible, inspectable and well-governed assistive AI. Withdraw protection where material effects, rights obstruction, vulnerability exploitation or missing evidence arise.
Cross-Sector Triage and Reuse	Provide one intake and referral protocol for cross-mandate harm, and publish common definitions, incident categories, evidence templates and supplier clauses.

4. Evidence Base and Recent Developments

4.1 Consumer Use, Understanding and Redress

The DRCF's 2025 online survey of 4,000 people, quota-sampled and weighted to nationally representative proportions among UK internet users, provides a useful starting point because it examined awareness, use, perceived benefits and harms, accountability and redress within the same study. 92 per cent of respondents had heard of generative AI, but only 10 per cent said that they could explain it in detail; 35 per cent consciously used generative AI, while a further 40 per cent had used a tool or encountered an AI search summary without initially recognising that they had done so. Those figures do not establish incapacity, but they show why a regulatory model that assumes consumers can reliably identify AI involvement before deciding whether to trust, verify or contest an output is unlikely to be defensible.[14]

The same research found that 21 per cent of generative-AI users reported using none of the listed verification techniques. In comparison, even 9 per cent of daily users and 13 per cent of the study's most confident segment

[14] Thinks Insight & Strategy, Understanding Consumer Use 5, 11–17. The online panel used quotas for age, gender, region, ethnicity, socio-economic group and working status, followed by weighting to nationally representative proportions; the report states that the resulting data should be understood as representing UK internet users because digitally excluded people could not take part.

reported no such checks. Among users who had detected an error or issue, 11 per cent said that the issue had caused harm to themselves or others; the reported rate was 16 per cent among the most confident segment and 17 per cent among cautious adopters, which is a useful warning against assuming that general awareness of risk will reliably prevent adverse outcomes.[15]

The redress findings are particularly relevant to Phase 2. Only 22 per cent of respondents expressed confidence that compensation could be sought when something went wrong; 6 per cent reported having sought redress, while 12 per cent said that they had needed redress but did not know how to pursue it, and a further 10 per cent had needed it but did not think that redress was possible. These results support clearer routes, records, and responsibility. However, they do not show that every AI-related loss requires a new liability regime, as some reported issues will already fall within existing contractual, consumer, data-protection, or sectoral remedies.[16]

The accountability results point in the same practical direction. 48 per cent of respondents described developers as fully responsible for ensuring that their outputs do not lead to loss or harm. In comparison, 42 per cent said the same of hosts and 83 per cent attributed at least some responsibility to users. Because respondents did not have to divide a fixed amount of responsibility among those actors, the findings should not be read as a liability formula. They nevertheless show that consumers expect responsibility to be shared and do not regard the user as the sole risk bearer.[17]

4.2 Delegation, Control and Financial Stakes

The FCA's Mills Review adds a newer, sector-specific layer of evidence. Its nationally representative online survey of 5,026 UK retail financial-services consumers, fielded between 21 and 29 April 2026 after eight focus groups, found that 67 per cent already used AI in some form, while use in financial services remained lower and concentrated on information, explanation and suggestions. Among people already using AI for financial services, 24 per cent reported uploading personal financial data or documents; 13 per cent of adults said that they would be willing to grant AI real-time access to financial information, rising to 18 per cent among AI users.[18]

Stated willingness declined as autonomy increased, but it did not disappear. 36 per cent said that they would use an assistive system that reviewed information and made recommendations, 30 per cent would use a system that could act after asking permission each time, and 20 per cent would use a system that acted autonomously within pre-set goals; the last figure rose to 28 per cent among existing AI users. These are stated preferences about described propositions rather than observed transactions. Yet, they indicate that meaningful demand for delegation may emerge before consumers, firms and regulators have settled how authority, evidence, liability and reversal should operate.[19]

The trust results explain why mandate architecture cannot be reduced to disclosure. Among AI users, 75 per cent trusted AI to explain matters clearly and 71 per cent trusted it to provide useful guidance. Still, only 23 per cent trusted it to avoid misleading information, and 21 per cent trusted it to handle personal information responsibly. 68 per cent of consumers were concerned about data misuse, 67 per cent about the lack of protection when something went wrong, and 65 per cent about concentration of power. In comparison, only 40 per cent correctly understood that relying on general-purpose AI for financial advice did not provide formal recourse.[20]

The most policy-relevant finding is not that consumers *reject* AI, but that perceived *usefulness* and perceived *protection* move in opposite directions. 55 per cent identified at least one benefit of AI for day-to-day financial management, while the strongest stated drivers of future willingness were clear protection if something went wrong and evidence of accuracy and reliability; convenience was less important. That evidence supports

[15] Thinks Insight & Strategy, Understanding Consumer Use 30–34. The reported harms included financial, health, work and scam-related examples, but the report appropriately cautions that further research is required to understand their nature and severity.
[16] Thinks Insight & Strategy, Understanding Consumer Use 35–37.
[17] Thinks Insight & Strategy, Understanding Consumer Use 35–36.
[18] Financial Conduct Authority, The Mills Review 36–38, 125–126. The survey methodology, quotas and fieldwork dates are described at 125–126. FCA Executive Director Sheldon Mills led the review.
[19] Financial Conduct Authority, The Mills Review 39–40, 126–129.
[20] Financial Conduct Authority, The Mills Review 36, 40, 127–129.

controls which make delegation safer without prohibiting it, including approval before consequential action, reliable records, bounded permissions and a clear route to correction.[21]

4.3 Choice Architecture, Consent and Personalised Friction

The empirical literature on digital choice architecture provides mixed evidence on how interface design changes behaviour. Nouwens and colleagues scraped consent-management designs used on 680 of the 10,000 most-visited UK websites and found that only 11.8 per cent met the study's three minimum legal conditions; in a separate field experiment with 40 participants, removing a first-page rejection option increased consent by 22–23 percentage points, while presenting more granular controls on the first page reduced consent by 8–20 percentage points. The study concerns cookie consent rather than AI. Still, it demonstrates why the practical location, symmetry and granularity of a control can matter more than the formal existence of a choice.[22]

Luguri and Strahilevitz reached a related conclusion in two large-scale experiments using census-weighted samples of United States adults. Mild dark patterns more than doubled enrolment in a dubious service, and aggressive designs nearly quadrupled it. In contrast, subtler patterns generated less backlash, and less-educated participants were more susceptible to the mild condition. Jurisdictional and task differences limit direct transfer to UK AI services. Still, causal evidence supports closer scrutiny of hidden information, obstruction, trick questions, and default effects when AI interfaces are used to obtain consent, discourage cancellation, or steer consumers towards higher-value outcomes for the firm.[23]

The CMA's review of online choice architecture and the OECD's work on dark commercial patterns compile a broader evidence base on defaults, scarcity claims, sludge, information overload, confirmshaming, and other design practices that impair consumer choice. AI does not create those practices, but adaptive systems can make them cheaper to test, personalise and allocate, including by shifting deceptive design below the interface and adjusting friction or information based on inferred value, urgency or vulnerability. For that reason, regulators should ask not only what an interface shows, but also which objective the system is optimising, which variants were tested, which groups experience more friction, and whether the firm has prohibited optimisation for complaint abandonment or the reduced exercise of statutory rights.[24]

This evidence also explains why a recurring "AI consent banner" would be an inadequate central remedy. Additional prompts can increase formal disclosure while leaving the underlying choice environment unchanged, particularly when refusal is slower, less prominent or linked to degraded service. Consent remains important where the law requires it, but the design of the choice and the consequences of refusal must be examined alongside the words displayed to the consumer.[25]

Recent enforcement reinforces the distinction between formal choice and operational choice. The CMA's final infringement notice against Marks Electrical concerned consumers who automatically opted into additional services, leading to a £720,000 penalty and refunds. The separate Microsoft investigation concerns whether customers received clear information about subscription options and cost differences when plans changed, including through the addition of Copilot features. No finding has been made in that open investigation. Together, the matters show why defaults, option presentation and commercial integration deserve scrutiny even before AI adapts them at the individual level.[26]

---

[21] Financial Conduct Authority, The Mills Review 128–129.
[22] Nouwens and others, 'Dark Patterns after the GDPR' 1, 5, 8–9. The authors describe the compliance test as a minimum hurdle rather than a complete legal assessment.
[23] Jamie Luguri and Lior Jacob Strahilevitz, 'Shining a Light on Dark Patterns' (2021) 13 Journal of Legal Analysis 43–109.
[24] Competition and Markets Authority, Online Choice Architecture; OECD, Dark Commercial Patterns; OECD, 'Stronger Consumer Protections Needed to Address Current and Emerging Harms Consumers Face Online' (9 October 2024); M R Leiser, Dark Patterns, Deceptive Design, and the Law: AI's Hidden Influence on Our Digital Experience (Bloomsbury Academic 2025).
[25] Solove, 'Privacy Self-Management'; Ben-Shahar and Schneider, More Than You Wanted to Know; Nouwens and others, 'Dark Patterns after the GDPR'.
[26] Competition and Markets Authority, Marks Electrical: Consumer Protection Enforcement Case (final infringement notice 15 June 2026; non-confidential notice published 2 July 2026); Competition and Markets Authority, Microsoft: Consumer Protection Enforcement Case (opened 27 July 2026). No finding had been made in the Microsoft investigation as of 18 August 2026.

4.4 Legal, Regulatory and Assurance Developments

The CMA's March 2026 guidance provides the clearest domestic accountability starting point. The same consumer-law rules apply whether a customer interacts with a human or an AI agent. The business remains responsible, including where a third party designed or supplied the agent. The guidance expects firms to consider disclosure where AI involvement may affect a decision, train and test agents for legal compliance, monitor performance and refine an agent quickly when a problem emerges.[27]

That guidance now operates against the stronger direct-enforcement structure in Part 3 of the Digital Markets, Competition and Consumers Act 2024. The unfair commercial practices regime in Part 4 came into force on 6 April 2025, while Part 3 allows the CMA to determine infringements and impose penalties directly. A fixed penalty for an infringing consumer practice may reach the higher of £300,000 or 10 per cent of turnover. A firm cannot avoid responsibility for a misleading action, misleading omission, aggressive practice or other prohibited conduct merely because a model, supplier or agent generated the interaction.[28]

The Data (Use and Access) Act 2025 replaced the former UK GDPR Article 22 framework with Articles 22A–22D. The amended regime permits a wider range of significant solely automated decisions. Still, Article 22C requires safeguards that include information about the decision, an opportunity to make representations, human intervention and an ability to contest the outcome. Consumer AI should make those safeguards usable by identifying material AI involvement, preserving the relevant record and providing access to a human who can alter the result rather than repeat it.[29]

The FCA's Consumer Duty offers a useful domestic model because it requires firms to act to deliver good retail customer outcomes and sets out distinct expectations for consumer understanding and consumer support. Its value for the DRCF lies in the shift from formal disclosure to evidence of whether communications and support work in practice. However, a cross-sector consumer-AI framework should adapt that logic rather than import the financial-services rulebook. The FCA's vulnerability guidance and Financial Lives research also support testing for outcome gaps and for circumstances that can change over time, rather than assuming that vulnerability is a fixed personal trait.[30]

Ofcom's work identifies two further risks. Its answer-engines paper explains the movement from search services that direct users towards multiple pages to generative services that produce one fluent answer, reducing the visibility of sources and alternative perspectives. Its telecoms research examines how AI may affect customer support, personalisation and operational decisions across the journey. Ofcom's July 2026 penalty against Virgin Media concerned non-AI cancellation failures, including deliberate call dropping, unnecessary transfers and repeated holds. The case nevertheless provides a concrete baseline for measuring rights-exercising friction that AI could personalise, conceal or scale.[31]

General standards and assurance frameworks are useful, but their role should be stated accurately. NIST's AI Risk Management Framework and Generative AI Profile support governance, mapping, measurement, and management; ISO/IEC 42001 provides an AI management system standard; and ISO/IEC 42005 guides AI system impact assessments. Those instruments can structure evidence and continual improvement. Still, certification or

[27] Competition and Markets Authority, 'Complying with Consumer Law When Using AI Agents' (9 March 2026).
[28] Digital Markets, Competition and Consumers Act 2024, pt 3 ch 4, especially s 182(4)(b) and (6), and pt 4 ch 1; Digital Markets, Competition and Consumers Act 2024 (Commencement No 2) Regulations 2025, SI 2025/272. A final infringement notice may impose a fixed penalty up to the higher of £300,000 or 10 per cent of the respondent's turnover.
[29] Data (Use and Access) Act 2025, s 80; Data (Use and Access) Act 2025 (Commencement No 6 and Transitional and Saving Provisions) Regulations 2026, SI 2026/82; Information Commissioner's Office, 'Data Protection: Summary of the Changes under the Data (Use and Access) Act 2025'.
[30] Financial Conduct Authority, Final Non-Handbook Guidance for Firms on the Consumer Duty (FG22/5, July 2022); Financial Conduct Authority, 'Consumer Duty'; Financial Conduct Authority, Financial Lives 2024 Survey: Key Findings (2025); Financial Conduct Authority, Guidance for Firms on the Fair Treatment of Vulnerable Customers (FG21/1).
[31] Ofcom, 'The Era of Answer Engines: Generative AI's Impact on Search Experiences and Online Safety' (4 November 2025); Ofcom, 'Exploring What Impact the Adoption of Artificial Intelligence Could Have on the Experience of Telecoms Customers' (27 January 2026); Ofcom, 'Ofcom Fines Virgin Media £28m' (8 July 2026).

a documented process does not itself establish compliance with consumer law, the fairness of an optimisation objective or the accessibility of redress, so the DRCF should map them onto concrete consumer outcomes.[32]

The United Kingdom's developing assurance policy supports that translation. Government work on AI assurance, the AI Management Essentials tool, and trusted third-party assurance recognises the need for a usable ecosystem while also identifying barriers caused by limited understanding, inconsistent terminology, and weak interoperability. A DRCF consumer framework would add value by specifying the consumer-facing evidence that a general assurance process should produce, rather than creating an entirely separate management system.[33]

Sectoral activity now makes interoperability urgent. Ofgem's June 2026 call asked whether the energy sector needs additional outcomes-based AI-assurance guidance, and Ofgem has decided to establish a 12-month technical sandbox.[34] The FCA's second Supercharged Sandbox cohort is testing safer agent-led payments, governance, accountability and services for vulnerable and underserved consumers.[35] The DRCF should treat those programmes as evidence-generating infrastructure and publish the control patterns, failure modes and minimum artefacts that can travel across sectors.

The DRCF's institutional position has also changed. The former AI and Digital Hub pilot has given way to a Thematic Innovation Hub, its first Agentic AI series has concluded, and the 2026/27 Workplan commits to a cross-regulatory Research Lab.[36] At the same time, the ICO expects final agentic-AI and updated automated-decision guidance in Winter 2026.[37] A DRCF framework should remain modular and versioned: it should provide a common operational grammar now without pre-empting regulator-specific guidance that remains in development.

The European Union provides a useful comparison, not a model for wholesale transplantation. The AI Act demonstrates a legal risk-classification approach. At the same time, the Commission's Digital Fairness Fitness Check identifies dark patterns, addictive design, unfair personalisation, problematic influencer marketing, and difficulties with digital contracts as persistent consumer law concerns. The United Kingdom can retain a sectoral and outcomes-focused model, but should learn from the EU's recognition that model governance and digital consumer fairness are related yet distinct questions.[38]

5. A Consumer AI Accountability Framework

The DRCF should develop a Consumer AI Accountability Framework around seven connected elements:

1. Classification
2. Purpose
3. Control
4. Evidence
5. Outcomes
6. Redress
7. Cooperation

---

[32] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0); National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1, July 2024); ISO/IEC 42001:2023; ISO/IEC 42005:2025.

[33] Department for Science, Innovation and Technology, Introduction to AI Assurance (12 February 2024); Department for Science, Innovation and Technology, Assuring a Responsible Future for AI (6 November 2024); Department for Science, Innovation and Technology, 'Guidance for Using the AI Management Essentials Tool: Government Response' (6 February 2026); Department for Science, Innovation and Technology, Trusted Third-Party AI Assurance Roadmap (3 September 2025).

[34] Ofgem, 'AI Assurance in the Energy Sector' (call for input, 17 June 2026; closed 12 August 2026); Ofgem, 'AI Technical Sandbox' (consultation outcome, 3 June 2026; decision to establish a 12-month pilot targeting commencement in late autumn 2026).

[35] Financial Conduct Authority, 'Supercharged Sandbox' (updated 22 July 2026). Cohort 2 runs from 13 July to 31 December 2026 and includes safer agent-led payments, AI governance and services for vulnerable and underserved consumers.

[36] Digital Regulation Cooperation Forum, 'DRCF Launches Thematic Innovation Hub for Innovators with First Focus on Agentic AI' (10 October 2025); Digital Regulation Cooperation Forum, 'Call for Input – Thematic Innovation Hub: Authentication and Trust' (25 June 2026); Digital Regulation Cooperation Forum, DRCF Workplan 2026/27 (27 April 2026).

[37] Information Commissioner's Office, 'Technology: Our Plans for New and Updated Guidance', listing final agentic AI guidance and updated automated decision-making and profiling guidance for Winter 2026.

[38] Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence [2024] OJ L 2024/1689; European Commission, 'Digital Fairness Fitness Check' (3 October 2024).

The framework should not displace existing laws, create a universal licensing system or pretend that one template can govern every use. It should translate existing consumer, data, communications, competition and financial-services duties into operational artefacts that firms can deploy and regulators can inspect.

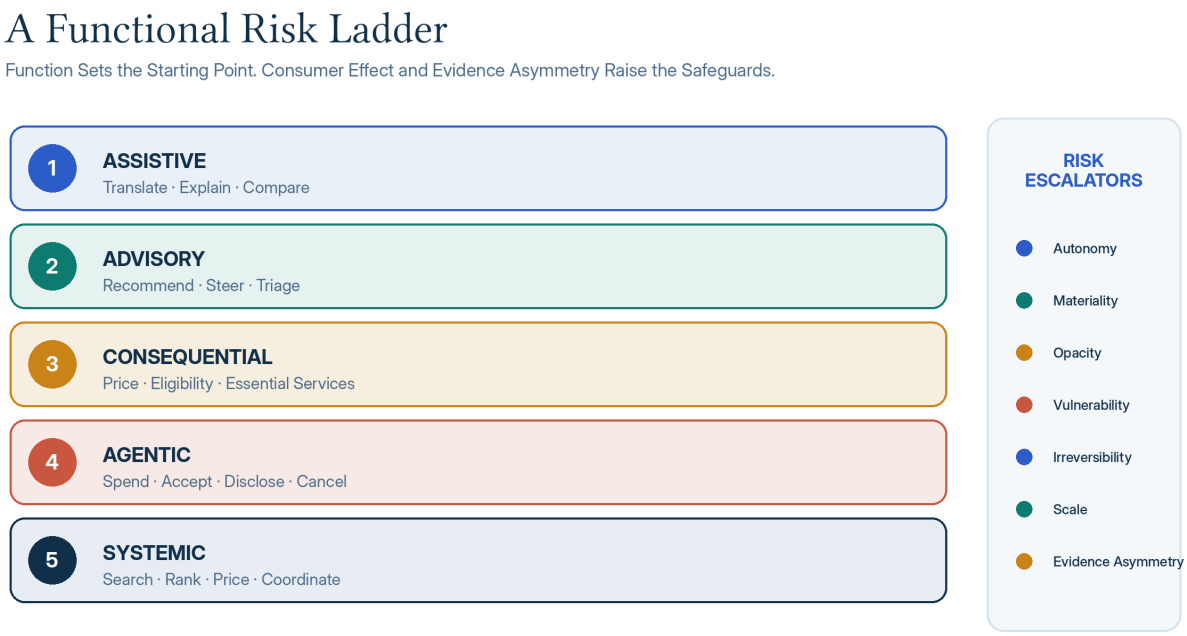


Figure 1. Functional Classification Sets the Starting Point; Risk Escalators Determine the Intensity of Safeguards.

5.1 Classification

Classification should begin with function because the legally relevant difference often lies in what the system can do rather than in the architecture used to produce an output. A chatbot that translates a complaint into plain English should not be treated as though it could cancel an energy contract, accept a credit agreement, or move money. At the same time, an apparently simple recommender may require stronger safeguards if it materially affects access to an essential service or systematically changes prices and rankings.

Tier	Illustrative Functions	Expected Safeguards
Tier 1: Assistive AI	Grammar, translation, accessibility, basic search support and bounded comparison.	Clear presentation, basic reliability testing, privacy compliance, accessible complaint route and safe fallback.
Tier 2: Advisory AI	Recommendations, advice-like outputs, complaint triage and product steering where the consumer retains the decision.	Domain boundaries, source grounding, calibrated uncertainty, over-reliance tests and escalation.
Tier 3: Consequential AI	Credit, insurance, debt, utilities, telecoms, healthcare access, eligibility, essential services and dispute resolution.	Impact assessment, fairness and vulnerability testing, challenge-ready records, meaningful review and interim protection.
Tier 4: Agentic AI	Purchases, cancellations, negotiation, contractual acceptance, data disclosure, payment instructions and material service changes.	Authority envelope, confirmation, limits, logs, receipts, expiry, revocation, pausing, reversal and rapid correction.
Tier 5: Systemic consumer AI	Search, ranking, recommenders, advertising, pricing, automated negotiation, platform governance and widely used agents.	Systemic and competition assessment, auditability, cross-sector monitoring, incident sharing and regulator access.

The tier should set the starting point rather than dictate the final result. Obligations should rise with autonomy, materiality, opacity, vulnerability, irreversibility, scale and evidence asymmetry. A system should not become high risk merely because it uses a large model when its function is bounded, reversible and genuinely assistive.

Conversely, a modest technical system can require enhanced safeguards when it affects an essential service, legal position or a consumer's ability to exercise a right.

5.2 Purpose and Prohibited Optimisation

Purpose should operate as a governance control rather than a slogan. A consumer-AI purpose statement should identify the consumer-facing function, the commercial objective, foreseeable effects on rights and interests, the groups or non-users affected, the alternative designs considered, and the objectives the system must not be optimised for. "Improve customer experience" is not a testable purpose; "reduce average support time while preserving complaint completion, cancellation, escalation and statutory rights" can be measured and challenged.

Commercial objectives are important because lawful automation can become exploitative through optimisation. A system rewarded only for conversion, retention, call deflection, complaint closure or cost reduction may learn to omit material information, allocate friction asymmetrically, steer consumers away from redress or intensify pressure when it detects urgency or distress. Firms should prohibit objectives such as complaint abandonment, cancellation delay, vulnerable-consumer conversion, revenue through confusion and suppression of statutory rights. The prohibition must reach reward functions, prompts, performance indicators, staff incentives, A/B testing and supplier instructions, not merely a policy statement.[39]

Purpose analysis should also be relational, as the relevant effects may extend beyond the person who uses the interface. Household members, authorised representatives, guarantors, dependants, co-travellers and consumers ranked or priced by a back-office system may be affected without interacting with the model. The evidence pack should therefore ask who is affected, what changes in condition occur, who receives the benefit, who bears the risk, who can contest the effect, and whether the same consumer objective could be achieved through a less rights-intrusive design.[40]

Data-protection purpose and necessity analysis provides a useful discipline, although consumer-AI governance should not collapse into data protection. The CJEU's decisions in Meta Platforms, Mousse and KNLTB emphasise, in different settings, that commercial practice does not define necessity and that legitimate interests require structured assessment of lawfulness, necessity, reasonable expectations and balancing. Those judgments are not a substitute for UK consumer law and should be used only as persuasive comparators, but their method supports requiring firms to explain why a chosen data and system design is no broader than the consumer-supportive purpose requires.[41]

5.3 Control, Mandate and Authority

Consumer control should be located where it can function legally and in practice. The framework should distinguish consent to personal-data processing, agreement to receive AI-mediated assistance, acceptance of a significant automated decision where the law permits it, and authorisation for an AI agent to undertake a specific act. A consumer may accept an AI explanation without authorising a purchase, may permit a single payment without permitting future payments, and may allow an agent to compare tariffs without permitting the agent to cancel an essential service.

For agentic systems, the appropriate control is a mandate architecture. The consumer should define an authority envelope covering purpose, duration, spending, counterparties, data disclosure and prohibited acts. Consequential actions should ordinarily require contemporaneous confirmation. The consumer should be able to pause the agent, view current authority, revoke or narrow it and inspect actions already taken. Authority

[39] Competition and Markets Authority, Online Choice Architecture; Luguri and Strahilevitz, 'Shining a Light on Dark Patterns'; M R Leiser, Dark Patterns, Deceptive Design, and the Law; Ofcom, 'Ofcom Fines Virgin Media £28m'.

[40] Leiser, 'Against the Sovereign User'; Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford University Press 2009).

[41] Case C-252/21 Meta Platforms Inc v Bundeskartellamt EU:C:2023:537; Case C-394/23 Mousse v Commission nationale de l'informatique et des libertés (CNIL) and SNCF Connect EU:C:2025:2; Case C-621/22 Koninklijke Nederlandse Lawn Tennisbond v Autoriteit Persoonsgegevens EU:C:2024:858.

should expire when the purpose ends, or the context changes materially, rather than becoming a permanent licence hidden in account settings.[42]

Firm-side control is equally important. The consumer-facing firm must name owners for the journey, objective setting, deployment approval, release management, supplier changes, incident response, system pause and redress. The firm should act as the systems integrator of responsibility. It should not treat the consumer as the systems integrator of last resort or rely on supplier opacity after choosing the service, benefiting from its deployment and controlling the route to remedy.[43]

A standard AI action receipt would make control and evidence visible without requiring disclosure of proprietary model internals. For every consequential action, the receipt should state what the system did, when it acted, which instruction or permission it relied upon, the applicable limit, the counterparty, the categories of data disclosed, whether human approval occurred, the resulting legal or economic effect, the responsible firm and the route for pause, reversal or challenge. The receipt should be machine-readable for portability and presented in plain language for the consumer.

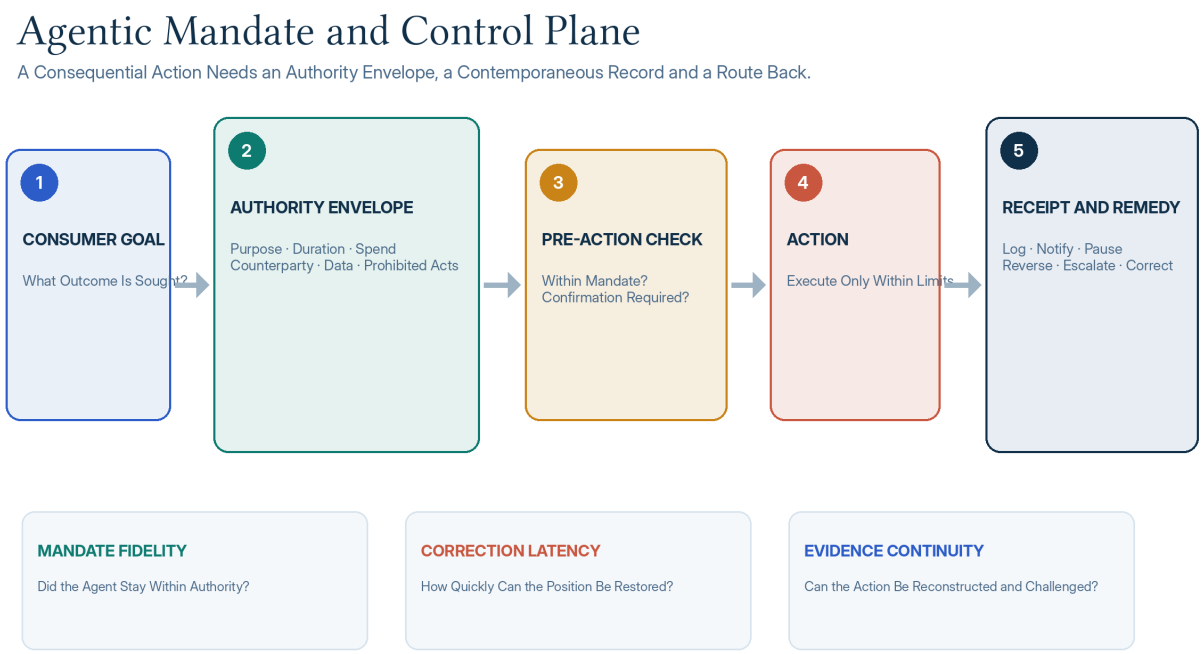


Figure 2. A Mandate and Control Plane for Consequential Agentic Action.

5.4 Evidence and Assurance

Evidence is the operational core of accountability because regulators cannot supervise broad principles without records that connect system design to consumer outcomes. Firms deploying consequential, agentic or systemic consumer AI should maintain an evidence pack covering purpose, data and supplier provenance, commercial and prohibited objectives, model and system versions, relevant instructions and prompts, permissions, tests, variants, human oversight, known limitations, security, incidents, monitoring metrics, complaints, reversals and remediation. The pack should expand as consumer effects become more serious, autonomous, opaque or difficult to reverse.[44]

Evidence should be layered according to audience. Consumers need an intelligible record of what the system did, under which permission, with what consequence and through which route it can be challenged; auditors and boards need evidence about design choices, controls and outcomes; regulators may need optimisation

[42] Competition and Markets Authority, 'Agentic AI and Consumers'; Digital Regulation Cooperation Forum, The Future of Agentic AI; Information Commissioner's Office, 'ICO Tech Futures: Agentic AI'.
[43] Competition and Markets Authority, 'Complying with Consumer Law When Using AI Agents'.
[44] National Institute of Standards and Technology, AI RMF 1.0; ISO/IEC 42001:2023; ISO/IEC 42005:2025; National Cyber Security Centre, Guidelines for Secure AI System Development (November 2023).

objectives, A/B testing records, prompt or instruction libraries, supplier documentation, incident histories, vulnerability-outcome gaps and escalation failures. Trade-secret protection can justify controlled disclosure, but it should not allow a firm to defeat redress by asserting that the relevant evidence is internal, proprietary or held by a supplier.

Pre-deployment testing should use realistic journeys rather than relying solely on generic benchmark performance. Scenarios should include consumers in financial difficulty, children where relevant, older consumers, disabled consumers, low-literacy and minority-language users, authorised representatives, consumers seeking cancellation or redress, and situational vulnerability events such as bereavement or urgent need. Testing should examine confusion, over-reliance, differential friction, source-grounding failure, mandate breach and escalation quality. At the same time, post-deployment monitoring should repeat those tests after material model, prompt, supplier or interface changes.

The framework should also distinguish regulatory intelligence from evidence relied upon to attribute responsibility. Cross-regulatory cooperation may require sharing complaints, technical assessments, market signals and incident information, but material used to establish a contravention should remain traceable, lawfully obtained, capable of disclosure where procedural fairness requires it, and open to challenge. Cooperation should improve visibility without weakening ordinary standards of evidence and due process.[45]

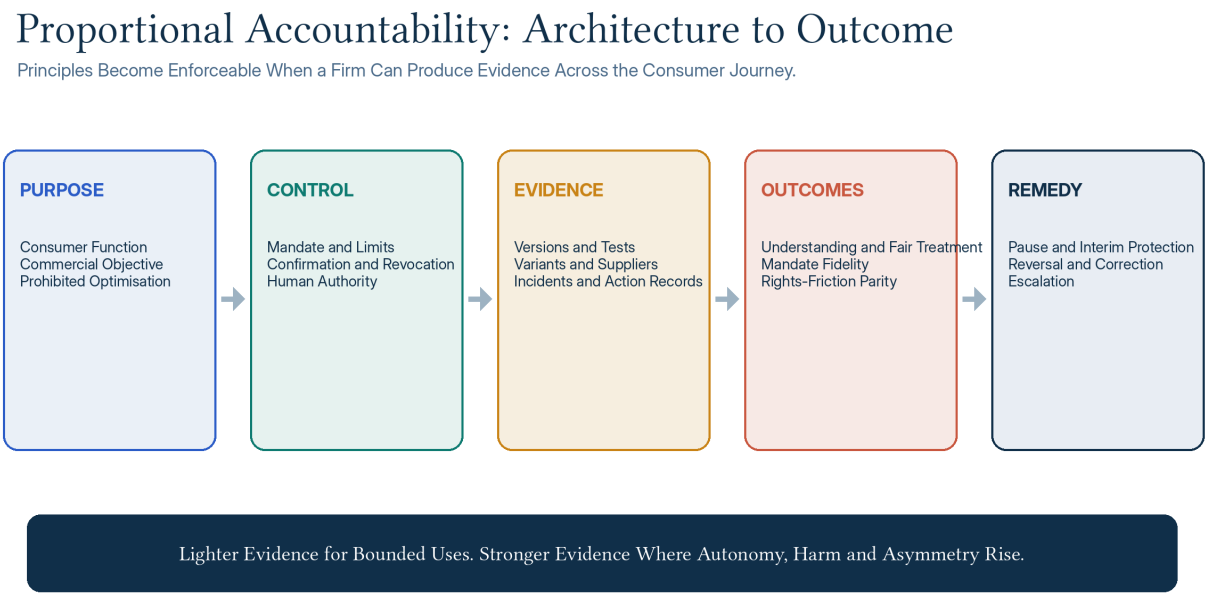


Figure 3. Proportional Accountability Connects System Architecture to Measurable Consumer Outcomes.

5.5 Outcomes and Vulnerability

The framework should measure four core consumer outcomes: understanding, control, fair treatment and remedy. Understanding asks whether consumers can recognise material AI involvement, distinguish assistance from authority and appreciate the limits relevant to the task; control asks whether they can approve, limit, revoke, opt out where appropriate, contest and escalate; fair treatment asks whether the system avoids unjustified differential outcomes and exploitative use of vulnerability; and remedy asks whether harm can be identified and corrected without imposing an impossible evidential burden on the consumer.[46]

Average accuracy cannot establish those outcomes because a technically strong system can still fail consumers when rights are exercised. Relevant measures include mandate fidelity, complaint abandonment, correction latency, time to meaningful human escalation, reversal success, source-grounding failure, differential price or

[45] Competition and Markets Authority, Direct Consumer Enforcement Guidance (CMA200, updated 4 April 2025), especially the provisions on evidence, representations, confidentiality and decision-making.

[46] Financial Conduct Authority, FG22/5, especially the consumer-understanding and consumer-support outcomes; Financial Conduct Authority, 'Consumer Understanding: Good Practice and Areas for Improvement' (13 March 2026).

eligibility outcomes and rights-friction parity. Group analysis should be complemented by event-based testing because context can create vulnerability even where no protected or demographic category predicts the event.

An outcomes approach is compatible with innovation because it leaves firms room to choose technical means while requiring evidence that those means produce fair, contestable and reversible consumer results. "Outcomes-based" must not become a synonym for evidence-free. Higher-impact systems should face defined testing intervals, change triggers, tolerability thresholds and board-level review, while lower-risk systems should use shorter checklists and standard controls.[47]

5.6 Redress

Redress should be built into the AI journey because a formal right has limited value when the consumer cannot identify the responsible firm, obtain the action record, or reach a human with the authority to change the outcome. Consequential and agentic systems should preserve relevant records, provide rapid escalation, and protect the consumer from further adverse effects. At the same time, a dispute is assessed, and support is reversed or corrected where technically and legally possible. Human involvement should be meaningful in the practical sense that the reviewer can examine the evidence and alter the result rather than merely endorse it.[48]

Where a firm-controlled AI agent undertakes a legally or economically significant action outside the consumer's mandate, the default should be rapid, no-fault correction for the consumer, leaving the deployer, developer, supplier, and insurer to allocate responsibility thereafter. This is narrower than strict liability for every AI error and draws on familiar allocation models: Regulation 76 of the Payment Services Regulations 2017 requires prompt restoration following an unauthorised payment, while distance-contract law gives consumers a defined cancellation period in many contexts. Neither regime should be copied mechanically, but each shows that consumer protection can separate prompt restoration from later allocation of fault.[49]

For other AI-related harms, a rebuttable evidential presumption may be more proportionate. Where the firm controlled the deployment, set the commercial objective, benefited from the interaction, and holds the relevant evidence, the burden should shift to the firm to show that it designed, tested, monitored, and remediated the system appropriately. The consumer should not be required to reconstruct a prompt chain, supplier incident or optimisation process that was never accessible from the outside.

Individual correction will not always be sufficient where the harm arises from system architecture. Personalised cancellation friction, discriminatory routing, exploitation of vulnerabilities, ranking effects, or misleading AI-generated authority may require journey redesign, removal of an optimisation objective, model or prompt changes, group-level remediation, audit access, and coordinated enforcement. Redress should restore the affected consumer and address the mechanism that made the harm repeatable.

5.7 Cooperation

Consumer AI is cross-regulatory by design because a single service can simultaneously operate as a consumer interface, a data-processing system, a financial-services tool, a communications journey, a recommender, a fraud-control mechanism and a competition gateway. Cooperation should therefore be treated as part of the accountability framework rather than an administrative afterthought, particularly where consumers cannot reasonably identify which regulator's mandate is at issue.[50]

The DRCF should create cooperation triggers where an AI system materially affects more than one member's mandate; require a short scoping note for significant cross-sector issues; distinguish shared intelligence from evidence used in enforcement; record material disagreements; and publish cross-regulatory scenario notes

[47] Financial Conduct Authority, FG22/5; Department for Science, Innovation and Technology, AI Opportunities Action Plan 221–228; M R Leiser, 'The Innovation Mandate: Making Europe's Digital Supervisory State Accountable for Rights, Growth and Scale' (SSRN, 24 July 2026) <https://ssrn.com/abstract=7073898>.

[48] Data (Use and Access) Act 2025, s 80 and inserted UK GDPR art 22C; Case C-203/22 CK v Magistrat der Stadt Wien (Dun & Bradstreet Austria) ECLI:EU:C:2025:117, used as a comparative authority on the information required to challenge an automated assessment.

[49] Payment Services Regulations 2017, SI 2017/752, reg 76; Consumer Contracts (Information, Cancellation and Additional Charges) Regulations 2013, SI 2013/3134, regs 29–31.

[50] Digital Regulation Cooperation Forum, 'AI and Digital Hub'; Digital Regulation Cooperation Forum, The Future of Agentic AI.

where firms need a coherent view. The objective should not be consensus at any price or the creation of a super-regulator, but reciprocal visibility, lawful disagreement and fewer gaps in consumer redress.

Cooperation is innovation infrastructure. Duplicative terminology, incompatible evidence requests and inconsistent guidance disproportionately raise costs for smaller firms. Shared definitions, incident categories, evidence templates, supplier clauses and safe-harbour criteria would allow firms to reuse compliant work across sectors while preserving each regulator's statutory powers and judgement. The DRCF's Thematic Innovation Hub and planned Research Lab provide practical routes for testing that interoperability.[51]

6. Responses to Phase 2 Questions

Q17. What Tools Can Industry Offer to Consumers or Otherwise Implement to Manage AI Risks?

Industry tools should govern the consumer journey, commercial objectives, mandate structure, supply chain and redress route rather than merely describe the model. The following tools provide a proportionate core, with documentation and testing scaled to the functional tier of the system:

- A consumer AI impact assessment should identify affected consumers and non-users, material effects, vulnerability events, alternative designs, reversibility, evidence needs and the route to remedy, while remaining shorter for genuinely assistive and bounded uses.[52]
- An agentic mandate and control plane should provide an authority envelope, approval before consequential action, granular and expiring permissions, spending and counterparty limits, data-disclosure limits, prohibited acts, revocation, pausing and a clear indication of which actions remain outside the mandate.
- An AI action receipt should show what the system did, when it acted, the authority relied upon, the data categories disclosed, the counterparty, the effect, the responsible firm and the available reversal or challenge route.
- Pre-deployment journey testing should include realistic scenarios involving financial difficulty, disability, low literacy, children where relevant, minority-language use, bereavement, urgent need, cancellation and complaints, while examining confusion, over-reliance and differential friction rather than task completion alone.
- Post-deployment monitoring should cover unauthorised action, mandate fidelity, consequential hallucination, source-grounding failure, disparate outcomes, rights-friction parity, complaint abandonment, cancellation time, escalation failure, model drift, supplier incidents, reversals and correction latency.
- Supplier governance clauses should require documentation, change notice, performance evidence, security cooperation, incident notification, access to relevant logs, assistance with regulator requests, preservation of evidence and support for consumer remediation.
- Redress-by-design mechanisms should provide a visible human route, evidence preservation, action pausing, interim protection, rapid correction and no-fault restoration where a firm-controlled agent acts outside mandate.
- Prohibited optimisation objectives should prevent systems from being rewarded for manipulation, complaint abandonment, cancellation friction, vulnerable-consumer conversion, revenue through confusion or reduced exercise of statutory rights.

Consumers do not need to read the framework itself to benefit from it. The useful consumer-facing statements are operational: "this agent cannot spend without approval"; "you can see every action"; "permissions expire on this date"; "this action can be reversed through this route"; and "this firm remains responsible".

Q18. Should Consumers Be Able to Opt Out of AI, and If Not, What Is the Legal Basis for the Use of Their Data?

A differentiated answer is preferable to a universal opt-out. Consumers should generally be offered a non-AI or human route where AI materially affects advice, support, recommendations, complaint handling, price, eligibility, cancellation, redress or another consequential outcome and a practicable alternative exists. A

[51] Digital Regulation Cooperation Forum, 'DRCF Launches Thematic Innovation Hub for Innovators with First Focus on Agentic AI' (10 October 2025); Digital Regulation Cooperation Forum, DRCF Workplan 2026/27 (27 April 2026).
[52] ISO/IEC 42005:2025; Department for Science, Innovation and Technology, 'Guidance for Using the AI Management Essentials Tool'.

universal right to remove every use of AI would be unrealistic in systems that support fraud detection, cybersecurity, accessibility, service reliability, or other back-office functions, and it could remove benefits without improving consumer control.

Where a full opt-out is not practicable, the substitute protections should become stronger: the firm should explain the material effect rather than merely the presence of AI, restrict processing to a lawful and specified purpose, test fairness and vulnerability outcomes, provide meaningful human review, preserve evidence and ensure that refusing unnecessary personalisation does not produce a deliberately degraded service. The legal basis for personal data processing must be identified under the UK GDPR as amended, while special-category data, inferred vulnerability, and significant solely automated decisions require their own analysis; a generic AI notice cannot supply a lawful basis where the underlying processing lacks one.[53]

The DRCF should expressly separate data-processing consent, AI-assistance choice and agentic authority. That separation would prevent firms from treating acceptance of a privacy notice as permission to enter a contract or disclose data to a third party, while allowing enduring preferences and machine-readable mandates to reduce repetitive prompts.

Q19. What Tools, Frameworks or Standards Can Regulators Leverage to Safeguard Consumer Protection?

Regulators should use established risk-management and assurance standards as building blocks, while adding a consumer-outcomes layer. NIST AI RMF, its Generative AI Profile, ISO/IEC 42001, ISO/IEC 42005, the OECD AI Principles and the NCSC secure-development guidance can structure governance, impact assessment, security, testing and continual improvement; none of them, however, determines whether a cancellation journey is unfair, whether an optimisation objective exploits vulnerability or whether a consumer can obtain an effective remedy.[54]

The DRCF should therefore publish a cross-sector Consumer AI Accountability Framework that maps these standards to material consumer effects, expected evidence and redress. Transfer across sectors is most feasible for common concepts, including mandate, action receipts, incident categories, supplier evidence and vulnerability-event testing. At the same time, tolerability thresholds should remain sector-specific because a failure rate acceptable for entertainment cannot be assumed acceptable for debt advice, insurance eligibility or essential-service cancellation.

Q20. What Types or Modes of Guidance or Other Regulatory Tools Are Most Helpful to Industry?

Industry needs guidance that is concrete enough to implement, modular enough for smaller firms and candid about the boundary between good practice and legal obligation. The most useful package would combine cross-regulatory scenarios, tiered checklists, evidence-pack templates, standard supplier clauses, worked examples of action receipts and permission dashboards, and a public catalogue of anti-patterns. Each scenario should identify the likely regulatory hooks, minimum evidence, expected consumer controls and change triggers without implying that following a template guarantees compliance.

Sandboxes, live testing and innovation hubs should remain available for genuinely novel or beneficial uses. Still, regulators should convert their learning into public artefacts and should not present participation as endorsement beyond the tested scope. The FCA's second Supercharged Sandbox cohort, the DRCF's Thematic Innovation Hub and Ofgem's planned technical sandbox show the value of controlled experimentation. The next step is to publish reusable scenario notes, evidence templates, control patterns and failure lessons across sectors.[55]

---

[53] UK GDPR, arts 5, 6, 9, 13–15 and 22A–22D; Data Protection Act 2018; Information Commissioner's Office, 'Data Protection: Summary of the Changes under the Data (Use and Access) Act 2025'.

[54] NIST, AI RMF 1.0; NIST, Generative AI Profile; ISO/IEC 42001:2023; ISO/IEC 42005:2025; OECD, Recommendation of the Council on Artificial Intelligence (as amended 2024); NCSC, Guidelines for Secure AI System Development.

[55] Financial Conduct Authority, 'Supercharged Sandbox' (updated 22 July 2026). Cohort 2 runs from 13 July to 31 December 2026 and includes safer agent-led payments, AI governance and services for vulnerable and underserved consumers.

Safe harbours should be available where a firm can show that the system is bounded, assistive, reversible, appropriately disclosed, monitored and supported by accessible redress. They should depend on actual function and evidence rather than self-description. A safe harbour should fall away where the system materially affects price, eligibility, essential services, vulnerability or statutory rights, or where the firm cannot produce the evidence needed to verify the conditions.

Q21. What Is the Current Application of AI Standards, and Is There International Best Practice?

The current body of standards is useful but incomplete. ISO/IEC 42001 provides an organisational management system; ISO/IEC 42005 addresses impact assessment; NIST supplies a flexible risk-management framework and a generative-AI profile; and the OECD Principles provide a widely recognised normative baseline. Their practical application varies across sectors and firm sizes, and certification and assurance markets remain in development, so regulators should avoid assuming that a standard is either irrelevant or sufficient.

International best practice is better understood as a combination of elements than as a single jurisdictional model: general risk management, specific consumer-law analysis, secure system development, meaningful records, proportionate testing, and remedies that work when consumers lack access to technical evidence. The EU AI Act offers legal classification, the EU Digital Fairness work identifies consumer-interface harms, and UK sectoral regulation offers outcomes-based supervision; the DRCF can add distinctive value by making those layers interoperable through a common consumer-AI grammar.

Q22. Is Informed and Meaningful Consent the Right Lever?

Informed and meaningful consent is sometimes legally required and can support autonomy, but it is not the central lever for consumer AI protection. Its practical preconditions, including comprehension, voluntariness, specificity and a genuine alternative, often weaken when AI forms part of essential services, search, customer support, recommendations, pricing, complaints or fraud controls. Adaptive behaviour also makes it difficult for a consumer to understand in advance every material way in which a system may act.

The empirical evidence on digital choice architecture shows that the placement and symmetry of options can materially alter apparent consent. In contrast, the DRCF and FCA research shows substantial gaps between familiarity, trust and understanding of protection. The response should therefore combine lawful consent, where required, with default limits, permission controls, confirmation before action, receipts, outcome testing, cooling-off periods, accessible human support, prohibited optimisation, and effective redress.[56]

Q23. Are There Other Levers?

The most effective additional levers are structural and procedural because they alter the conditions under which a system can cause harm rather than asking the consumer to predict every risk. The DRCF should prioritise the following measures:

- Approval before consequential action should be the default where an agent would spend money, accept terms, disclose sensitive data, cancel an essential service or materially change contractual status.
- Cooling-off and reversal should be available where the transaction type permits them, with the action receipt identifying the period and route; existing distance-contract cancellation and unauthorised-payment rules provide useful allocation models without dictating a universal period.[57]
- Permission dashboards should allow consumers to see, limit, revoke and set time-bound authority, while showing which actions have already occurred and which cannot be reversed.
- Human escalation should be capable of changing an outcome and should not require the consumer to repeat information already held in the system or navigate a deliberately inferior channel.

[56] Nouwens and others, 'Dark Patterns after the GDPR'; Thinks Insight & Strategy, Understanding Consumer Use 11–17, 35–37; Financial Conduct Authority, The Mills Review 40, 127–129.
[57] Consumer Contracts Regulations 2013, regs 29–31; Payment Services Regulations 2017, reg 76.

- A single primary point of accountability should sit with the consumer-facing firm, which may pursue contractual contribution from upstream providers after the consumer has been protected.
- Record-preservation duties should ensure that consequential systems retain the evidence needed to investigate mandate, version, instruction, data use, counterparty and outcome for a proportionate period.
- Detected vulnerability should be used to support the consumer, such as by slowing a journey or offering assistance. It should not be used to increase price, pressure, friction or commercial extraction.
- Rights-friction parity should compare the time, steps, transfers, repeated prompts and abandonment rate for acquisition against cancellation, complaint and redress routes.

Q24. Which Examples from Other Sectors or Jurisdictions Use These Tools Effectively?

The FCA Consumer Duty is the strongest domestic outcomes comparator because it links a broad standard to specific outcomes, board oversight, monitoring and evidence about consumer understanding and support. Its logic can be adapted cross-sectorally into understanding, control, fair treatment and remedy. At the same time, the FCA's vulnerability guidance shows how firms can identify characteristics and circumstances that create greater risk without treating vulnerability as a permanent label.

The Payment Services Regulations provide a useful analogue for redress because they require prompt restoration after an unauthorised payment and leave the later allocation of responsibility to the relevant providers. The Consumer Contracts Regulations provide a separate example of time-limited cancellation rights designed for transactions where distance and information asymmetry justify an opportunity to reconsider. These mechanisms are sector- and transaction-specific, but they show that consumer law already uses reversal, burden allocation and non-waivable procedural protection where ordinary notice is insufficient.

The EU AI Act and Digital Fairness Fitness Check offer comparative lessons on risk classification and digital consumer harms, while UK sandboxes and cross-regulatory hubs show how supervision can support experimentation. Recent UK enforcement also supplies practical lessons: the Marks Electrical final notice shows that defaults and express agreement can carry material consequences under the new direct-enforcement regime. At the same time, the open Microsoft investigation illustrates scrutiny of how firms present AI-linked subscription changes. No finding has been made in that investigation. Regulators should reuse tested mechanisms and state which consumer harm each one addresses.[58]

Q25. Can Outcomes-Based Approaches Analogous to the Consumer Duty Support Consumer Protection?

An outcomes-based approach is well suited to consumer AI, provided that outcomes remain measurable, evidence-based and paired with minimum controls for high-impact systems. The framework should require firms to demonstrate understanding, control, fair treatment and remedy while allowing them to choose technical methods that fit the use case. As autonomy, materiality, opacity, vulnerability or irreversibility rises, the evidence burden should rise, and the monitoring interval should shorten.

The principal risk is that outcome language becomes aspirational. The DRCF should therefore specify feasible indicators, require pre-deployment journey tests, define material-change triggers, expect post-deployment monitoring and require named owners to review outcome gaps and incidents. Principles should guide judgement, but records should make that judgement testable.

Q26. Who Should Be Held Accountable, and to What Extent, for AI Risks Materialising?

Accountability should track control, capability, evidence and benefit. The primary accountable actor should normally be the consumer-facing firm that places AI in the journey, presents it as part of its service, chooses or configures the commercial objective, benefits from the interaction and controls the route to redress. That firm

[58] Competition and Markets Authority, Marks Electrical: Consumer Protection Enforcement Case (final infringement notice 15 June 2026; non-confidential notice published 2 July 2026); Competition and Markets Authority, Microsoft: Consumer Protection Enforcement Case (opened 27 July 2026). No finding had been made in the Microsoft investigation at 18 August 2026.

can allocate responsibility upstream by contract. The consumer should not have to identify whether failure arose in a model developer, cloud provider, data source, API, prompt layer, integration or deployment context.[59]

Developers, integrators and suppliers should nevertheless owe duties appropriate to their capabilities, including accurate documentation, change notices, disclosure of known limitations, cooperation on incidents, security support, preservation of relevant evidence, and assistance with audits and remediation. Regulators should remain accountable for coherent guidance, proportionate supervision and lawful cooperation, but regulatory approval or sandbox participation should not transfer the firm's ordinary responsibility for its service.

For unauthorised agentic action, rapid correction should be as close to strict as possible from the consumer's perspective, because firms control mandate and transaction records. For other harms, a rebuttable evidential presumption is more proportionate where the firm controlled the system and holds the evidence. Universal strict liability for every inaccurate or disappointing output would be excessive, while requiring consumers to prove an inaccessible black-box failure would make redress illusory.

Q27. Do Tolerable Risks Exist for Consumers, and How Should They Be Operationalised?

Tolerable risks exist, but tolerability is contextual and cannot be inferred from adoption, continued use or formal consent. The DRCF should require firms to answer seven questions and document the residual risk:

- **Benefit**: Identify the concrete consumer benefit rather than invoke convenience in the abstract.
- **Probability**: Estimate how often the failure is likely to occur, including uncertainty and change over time.
- **Severity**: Account for low-frequency, high-impact outcomes rather than relying on average loss alone.
- **Distribution:** Identify which consumers receive the benefit and which consumers or non-users bear the risk.
- **Control:** Test whether the consumer can understand, limit, pause or revoke the system's authority.
- **Reversibility:** Establish whether the legal, economic or practical position can be restored promptly.
- **Inspectability and Remedy:** Preserve challenge-ready evidence and measure how quickly a credible challenge produces correction.

The DRCF should distinguish accepted risk from transferred risk. A consumer may knowingly accept a bounded chance of an imperfect restaurant recommendation. That consumer should not be treated as having accepted an unauthorised transaction, a hidden contractual change, a discriminatory denial of an essential service, personalised obstruction of cancellation, exploitation of distress, or an outcome that cannot be challenged because the firm failed to preserve records.

Some uses should therefore be presumptively intolerable unless exceptional safeguards and a clear legal basis are demonstrated: unauthorised financial action; hidden acceptance of terms; discriminatory access to essential services; exploitative targeting of children or consumers in vulnerable circumstances; systems designed to obstruct complaints, cancellation or statutory rights; agentic action outside mandate; use of vulnerability detection to increase price or pressure; and material consumer effects produced without evidence sufficient for challenge.

Q28. What Metrics, Thresholds or Proxy Indicators Are Feasible?

Metrics should combine technical performance, consumer-journey outcomes, redress outcomes and governance indicators. No single measure can represent consumer AI risk, and thresholds should remain sensitive to sector and function. A common metric vocabulary would nevertheless allow firms and regulators to compare trends, identify outliers and set clear intervention triggers.

[59] Competition and Markets Authority, 'Complying with Consumer Law When Using AI Agents'.

Metric Category	Illustrative Measures
Transaction and Mandate	Mandate-fidelity rate; unauthorised-action rate; failed-confirmation rate; actions outside limits; disputed-authority rate; reversal rate; restoration time.
Advice and Information	Consequential hallucination rate; source-grounding failure; calibrated abstention; correction rate; overconfidence indicators; inappropriate domain crossing.
Fair Treatment	Disparate price, eligibility, support, complaint and cancellation outcomes; vulnerability-outcome gaps; differential friction by group, event or inferred value.
Control	Approval-before-action compliance; permission revocation and expiry; override rate; consequential acts without confirmation; successful pause rate.
Redress	Complaint volume and abandonment; time to meaningful human escalation; correction latency; evidence-production time; interim protection; outcome alteration rate.
Journey and Rights-Friction Parity	Time, steps and transfers required to acquire versus cancel, complain or seek redress; repeated prompts; drop-off; channel degradation.
Governance and Suppliers	Prohibited: objective tests; drift; recurring incidents; unresolved audit findings; overdue assessments; missing logs; supplier response; unreviewed changes.
Systemic and Market	Ranking concentration; correlated agent behaviour; cross-platform errors; market-wide pricing effects; affiliated steering; reduced source diversity.

Thresholds should be developed through sector-specific pilots and should combine quantitative triggers with qualitative red lines. An unauthorised consequential action should trigger immediate investigation even at low incidence. A small average difference in cancellation time may also warrant intervention when it concentrates among consumers in financial difficulty or those attempting to exercise a statutory right. Regulators should use synthetic users, mystery shopping, journey replay and controlled comparison of variants because interaction history, A/B testing, inferred vulnerability and optimisation logic will often remain invisible in a single screenshot.

7. A Practical 12-Month Implementation Model

The DRCF can deliver immediate protections for consequential and agentic uses while allowing evidence and sector-specific thresholds to mature. A staged programme would convert the submission's concepts into reusable artefacts rather than another high-level principles document.

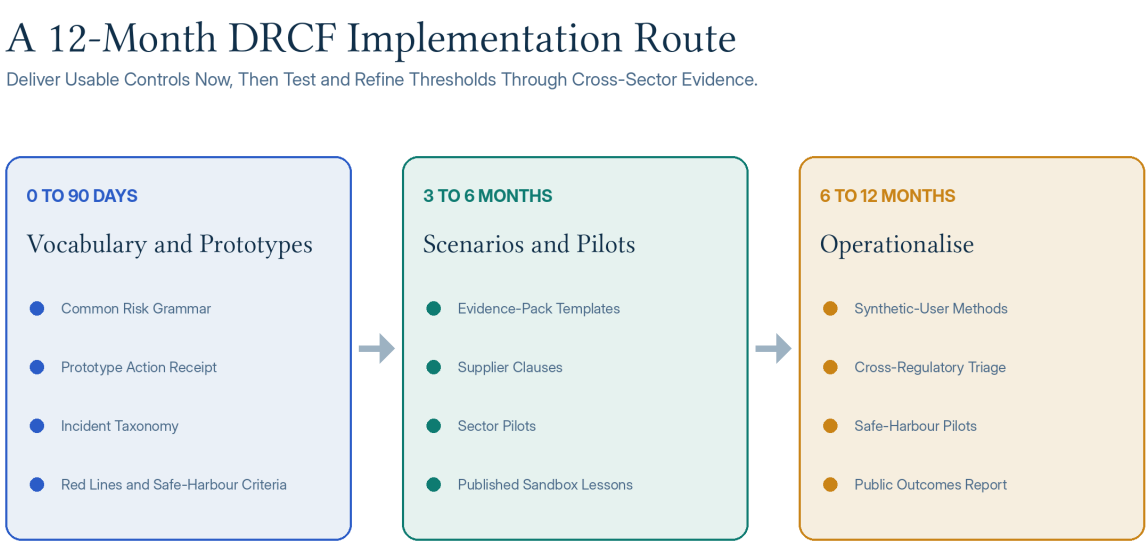


Figure 4. A Staged Route from Common Vocabulary to Tested Cross-Sector Implementation.

7.1 First 90 Days: Common Language and Prototype Controls

- Publish a common vocabulary for the five functional tiers, material consumer effect, vulnerability event, mandate fidelity, rights-friction parity, meaningful human escalation and redress by design.
- Publish prototype versions of an authority envelope, permission dashboard, AI action receipt, incident taxonomy and short evidence pack.
- State qualitative red lines for unauthorised consequential action, hidden acceptance of terms, vulnerability extraction, rights obstruction and material effects without challenge-ready evidence.

7.2 Three to Six Months: Scenario Guidance and Sector Pilots

- Map customer support, complaint triage, shopping agents, telecoms switching, debt guidance, financial recommendations, AI search, subscription management, personalised pricing and AI-to-AI negotiation to likely legal hooks and safeguards.
- Pilot short, standard and enhanced evidence packs with firms of different sizes, alongside standard supplier clauses and change-notification requirements.
- Convert learning from the Thematic Innovation Hub, FCA sandboxes and Ofgem's sandbox into public control patterns, failure lessons and regulator-facing evidence templates.

7.3 Six to Twelve Months: Testing, Triage and Safe-Harbour Pilots

- Publish synthetic-user, mystery-shopping, journey-replay and controlled-variant testing methods, including vulnerability events and rights-friction parity.
- Establish one DRCF intake and referral protocol for cross-mandate consumer AI harms, separating shared intelligence from enforcement evidence.
- Pilot earned safe harbours for bounded, reversible and inspectable assistive uses, then publish an outcomes report on threshold evidence and gaps.

7.4 Minimum Evidence by Functional Tier

Functional Tier	Minimum Evidence Expected
Assistive	Purpose and owner; basic reliability and privacy checks; accessible complaint and fallback route; change log.
Advisory	Assistive evidence plus domain limits, source-grounding tests, uncertainty and abstention behaviour, over-reliance scenarios, monitoring and escalation records.

Functional Tier	Minimum Evidence Expected
Consequential	Advisory evidence plus impact assessment, fairness and vulnerability testing, material-decision records, meaningful review, interim protection and redress outcomes.
Agentic	Consequential evidence plus authority envelope, confirmation and limit tests, action receipts, identity and security controls, unauthorised-action monitoring, pausing and reversal.
Systemic	Relevant lower-tier evidence plus systemic and competition assessment, market-wide monitoring, auditability, cross-platform incident analysis and cross-regulatory cooperation plan.

8. Safe Harbours and Regulatory Anti-Patterns

A proportionate framework should make it easier to deploy genuinely low-risk AI. A safe harbour could apply where the system is assistive rather than agentic, does not materially determine price or eligibility, does not use sensitive or inferred vulnerability data for commercial optimisation, provides behaviourally meaningful information, preserves a practical human route, maintains proportionate records and offers accessible redress. Smaller firms should be able to demonstrate those conditions through a concise template and standard supplier evidence rather than bespoke assurance exercises.

Safe harbours should be earned and revocable. They should depend on actual function, bounded authority, reversibility, inspectability and evidence rather than the label chosen by the firm. A system should not qualify as assistive where it is configured to steer, convert, retain or suppress complaints, and it should not qualify as low risk where it materially affects essential services, regulated products, children, consumers in vulnerable circumstances or the exercise of legal rights. The safe harbour should be withdrawn when the firm cannot produce required evidence or monitoring identifies a material outcome gap.

The DRCF should also publish a short catalogue of regulatory anti-patterns so that firms can recognise designs that formally comply while undermining the intended protection:

- Terms and conditions or generic AI labels are used instead of controls that show what the system can do, which authority it holds and how an action can be stopped or reversed.
- A trust mark is displayed without independent scope, testing criteria, incident consequences, expiry or an accessible route for checking what the mark covers.
- Consent is requested without a genuine alternative, while refusal is made slower, less prominent or linked to avoidable degradation of the service.
- A nominal human reviewer cannot access relevant evidence, alter the outcome or provide interim protection, so the "human in the loop" performs no meaningful remedial function.
- The firm cannot reproduce the version, instruction, permission or action record needed to investigate a material consumer complaint.
- Supplier contracts leave the deployer without sufficient evidence or cooperation, while the resulting gap is transferred to the consumer seeking redress.
- Personalisation changes disclosure, friction, support, price or redress asymmetrically according to inferred vulnerability or commercial value without a legitimate consumer-supportive reason.
- Guidance remains so abstract or documentation so extensive that only large incumbents can operationalise it, even where a smaller firm proposes a bounded and reversible use.

9. Conclusion

Consumer-facing AI should be governed by its effects on consumers, markets and rights. Generative systems can create misplaced authority, inaccurate confidence and unclear responsibility. Agentic systems raise questions of mandate, transaction, identity and control. Back-office systems can produce invisible differential treatment, while systemic services can reshape information, ranking, pricing and competition. Each category can

generate genuine benefit, which is why a framework should regulate function and evidence rather than mythology.

The United Kingdom does not lack principles, standards or regulatory activity. It lacks a sufficiently interoperable consumer-AI operating model. A shared risk grammar, evidence architecture, and redress model would allow the CMA, ICO, Ofcom, and FCA to preserve their statutory roles while reducing duplication, improving referrals, and giving firms clearer expectations. The aim should not be a new general AI code, but a practical way to make existing duties work when consumer journeys become adaptive, delegated, and difficult to inspect.

The strongest approach is proportional accountability. Low-risk AI should not carry evidence duties designed for autonomous transactions, while high-impact systems should not hide behind low-risk language or supplier opacity. Consumer protection and innovation are not opposing goals when clear mandates, reliable records, outcome testing and effective redress allow beneficial services to earn rather than assume trust.

The law should not discipline consumers for being human; it should discipline systems that make human limits profitable.

10. Source List (OSCOLA)

Online sources and hyperlinks were last checked on 18 August 2026. The list separates primary legal materials, official and regulatory publications, standards, secondary literature and the author's related work.

Legislation and Regulatory Instruments

Consumer Contracts (Information, Cancellation and Additional Charges) Regulations 2013, SI 2013/3134, https://www.legislation.gov.uk/uksi/2013/3134/contents, accessed 18 August 2026.

Consumer Rights Act 2015 https://www.legislation.gov.uk/ukpga/2015/15/contents accessed 18 August 2026.

Data (Use and Access) Act 2025 https://www.legislation.gov.uk/ukpga/2025/18 accessed 18 August 2026.

Data (Use and Access) Act 2025 (Commencement No 6 and Transitional and Saving Provisions) Regulations 2026, SI 2026/82 https://www.legislation.gov.uk/uksi/2026/82/made accessed 18 August 2026.

Data Protection Act 2018 https://www.legislation.gov.uk/ukpga/2018/12/contents accessed 18 August 2026.

Digital Markets, Competition and Consumers Act 2024 https://www.legislation.gov.uk/ukpga/2024/13 accessed 18 August 2026.

Digital Markets, Competition and Consumers Act 2024 (Commencement No 2) Regulations 2025, SI 2025/272, https://www.legislation.gov.uk/uksi/2025/272/made, accessed 18 August 2026.

Payment Services Regulations 2017, SI 2017/752, https://www.legislation.gov.uk/uksi/2017/752/contents, accessed 18 August 2026.

Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence [2024] OJ L 2024/1689, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng, accessed 18 August 2026.

UK GDPR, as amended, https://www.legislation.gov.uk/eur/2016/679 accessed 18 August 2026.

Cases

Case C-203/22 CK v Magistrat der Stadt Wien (Dun & Bradstreet Austria) ECLI:EU:C:2025:117.

Case C-621/22 Koninklijke Nederlandse Lawn Tennisbond v Autoriteit Persoonsgegevens ECLI:EU:C:2024:858.

Case C-252/21 Meta Platforms Inc and Others v Bundeskartellamt ECLI:EU:C:2023:537.

Case C-394/23 Mousse v Commission nationale de l'informatique et des libertés (CNIL) and SNCF Connect ECLI:EU:C:2025:2.

Official and Regulatory Materials

Competition and Markets Authority, Agentic AI and Consumers (9 March 2026) https://www.gov.uk/government/publications/agentic-ai-and-consumers/agentic-ai-and-consumers accessed 18 August 2026.

Competition and Markets Authority, 'Complying with Consumer Law When Using AI Agents' (9 March 2026) https://www.gov.uk/government/publications/complying-with-consumer-law-when-using-ai-agents/complying-with-consumer-law-when-using-ai-agents accessed 18 August 2026.

Competition and Markets Authority, Direct Consumer Enforcement Guidance (CMA200, updated 4 April 2025) https://www.gov.uk/government/publications/direct-consumer-enforcement-guidance-cma200 accessed 18 August 2026.

Competition and Markets Authority, Online Choice Architecture: How Digital Design Can Harm Competition and Consumers (5 April 2022) https://www.gov.uk/government/publications/online-choice-architecture-how-digital-design-can-harm-competition-and-consumers accessed 18 August 2026.

Competition and Markets Authority, Marks Electrical: Consumer Protection Enforcement Case (updated 2 July 2026) https://www.gov.uk/cma-cases/marks-electrical-consumer-protection-enforcement-case accessed 18 August 2026.

Competition and Markets Authority, Microsoft: Consumer Protection Enforcement Case (opened 27 July 2026) https://www.gov.uk/cma-cases/microsoft-consumer-protection-enforcement-case accessed 18 August 2026.

Department for Science, Innovation and Technology, AI Opportunities Action Plan (CP 1241, 13 January 2025), https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan, accessed 18 August 2026.

Department for Science, Innovation and Technology, 'AI Opportunities Action Plan: Government Response' (13 January 2025), https://www.gov.uk/government/publications/ai-opportunities-action-plan-government-response/ai-opportunities-action-plan-government-response, accessed 18 August 2026.

Department for Science, Innovation and Technology, Assuring a Responsible Future for AI (6 November 2024) https://www.gov.uk/government/publications/assuring-a-responsible-future-for-ai/assuring-a-responsible-future-for-ai accessed 18 August 2026.

Department for Science, Innovation and Technology, 'Data Use and Access Act 2025: Plans for Commencement' https://www.gov.uk/guidance/data-use-and-access-act-2025-plans-for-commencement accessed 18 August 2026.

Department for Science, Innovation and Technology, 'Guidance for Using the AI Management Essentials Tool: Government Response' (6 February 2026) https://www.gov.uk/government/consultations/ai-management-essentials-tool/outcome/guidance-for-using-the-ai-management-essentials-tool-government-response accessed 18 August 2026.

Department for Science, Innovation and Technology, Introduction to AI Assurance (12 February 2024) https://www.gov.uk/government/publications/introduction-to-ai-assurance accessed 18 August 2026.

Department for Science, Innovation and Technology, Trusted Third-Party AI Assurance Roadmap (3 September 2025) https://www.gov.uk/government/publications/trusted-third-party-ai-assurance-roadmap/trusted-third-party-ai-assurance-roadmap accessed 18 August 2026.

Digital Regulation Cooperation Forum, 'AI and Digital Hub' (closed pilot) https://www.drcf.org.uk/ai-and-digital-hub accessed 18 August 2026.

Digital Regulation Cooperation Forum, 'Call for Input: Consumer Interest and AI' (3 June 2026) https://www.drcf.org.uk/news-and-events/news/call-for-input-consumer-interest-and-ai accessed 18 August 2026.

Digital Regulation Cooperation Forum, DRCF Workplan 2026/27 (27 April 2026) https://www.drcf.org.uk/publications/work-plans/drcf-workplan-202627 accessed 18 August 2026.

Digital Regulation Cooperation Forum, 'DRCF Launches Thematic Innovation Hub for Innovators with First Focus on Agentic AI' (10 October 2025) https://www.drcf.org.uk/news-and-events/news/drcf-launches-thematic-innovation-hub-for-innovators-with-first-focus-on-agentic-ai accessed 18 August 2026.

Digital Regulation Cooperation Forum, 'Call for Input - Thematic Innovation Hub: Authentication and Trust' (25 June 2026) https://www.drcf.org.uk/news-and-events/news/call-for-input-thematic-innovation-hub-authentication-and-trust accessed 18 August 2026.

Digital Regulation Cooperation Forum, The Future of Agentic AI: Foresight Paper (31 March 2026) https://www.drcf.org.uk/publications/papers/thefutureofagenticai accessed 18 August 2026.

European Commission, 'Review of EU Consumer Law: Digital Fairness Fitness Check' (3 October 2024) https://commission.europa.eu/law/law-topic/consumer-protection-law/review-eu-consumer-law_en accessed 18 August 2026.

Financial Conduct Authority, 'Consumer Duty' https://www.fca.org.uk/firms/consumer-duty accessed 18 August 2026.

Financial Conduct Authority, 'Consumer Understanding: Good Practice and Areas for Improvement' (13 March 2026) https://www.fca.org.uk/publications/good-and-poor-practice/consumer-understanding-good-practice-areas-improvement accessed 18 August 2026.

Financial Conduct Authority, Final Non-Handbook Guidance for Firms on the Consumer Duty (FG22/5, July 2022) https://www.fca.org.uk/publication/finalised-guidance/fg22-5.pdf accessed 18 August 2026.

Financial Conduct Authority, Financial Lives 2024 Survey: Key Findings (2025) https://www.fca.org.uk/publication/financial-lives/financial-lives-survey-2024-key-findings.pdf accessed 18 August 2026.

Financial Conduct Authority, Guidance for Firms on the Fair Treatment of Vulnerable Customers (FG21/1, February 2021) https://www.fca.org.uk/publication/finalised-guidance/fg21-1.pdf accessed 18 August 2026.

Financial Conduct Authority, 'Supercharged Sandbox', Cohort 2 (updated 22 July 2026) https://www.fca.org.uk/firms/innovation/supercharged-sandbox accessed 18 August 2026.

Financial Conduct Authority and Information Commissioner's Office, 'Joint FCA and ICO Statement on Regulatory Expectations Regarding Firms' Approaches to Vulnerability Related Data' (27 March 2026, updated 20 July 2026) https://www.fca.org.uk/publications/corporate-documents/joint-fca-and-ico-statement-regulatory-expectations-regarding-firms accessed 18 August 2026.

Information Commissioner's Office, 'Data Protection: Summary of the Changes under the Data (Use and Access) Act 2025' https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/data-use-and-access-act-2025/the-data-use-and-access-act-2025-duaa-summary-of-the-changes/data-protection/ accessed 18 August 2026.

Information Commissioner's Office, 'Explaining Decisions Made with Artificial Intelligence: The Legal Framework' https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/legal-framework/ accessed 18 August 2026.

Information Commissioner's Office, 'ICO Tech Futures: Agentic AI' https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/tech-horizons-and-ico-tech-futures/ico-tech-futures-agentic-ai/ accessed 18 August 2026.

Information Commissioner's Office, 'Technology: Our Plans for New and Updated Guidance' https://ico.org.uk/about-the-ico/what-we-do/our-plans-for-new-and-updated-guidance/technology/ accessed 18 August 2026.

Financial Conduct Authority, The Mills Review: AI and the Future of Retail Financial Services (July 2026) https://www.fca.org.uk/publication/corporate/the-mills-review.pdf accessed 18 August 2026.

National Cyber Security Centre, Guidelines for Secure AI System Development (November 2023) https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development accessed 18 August 2026.

OECD, Dark Commercial Patterns (OECD Digital Economy Papers No 336, 2022) https://www.oecd.org/en/publications/dark-commercial-patterns_44f5e846-en.html accessed 18 August 2026.

OECD, 'Stronger Consumer Protections Needed to Address Current and Emerging Harms Consumers Face Online' (9 October 2024) https://www.oecd.org/en/about/news/press-releases/2024/10/stronger-consumer-protections-needed-to-address-current-and-emerging-harms-consumers-face-online.html accessed 18 August 2026.

Ofgem, 'AI Assurance in the Energy Sector' (17 June 2026) https://www.ofgem.gov.uk/call-for-input/ai-assurance-energy-sector accessed 18 August 2026.

Ofgem, 'AI Technical Sandbox' (consultation outcome, 3 June 2026) https://www.ofgem.gov.uk/consultation/ai-technical-sandbox accessed 18 August 2026.

Ofcom, 'Exploring What Impact the Adoption of Artificial Intelligence Could Have on the Experience of Telecoms Customers' (27 January 2026) https://www.ofcom.org.uk/phones-and-broadband/telecoms-infrastructure/Exploring-what-impact-the-adoption-of-artificial-intelligence-could-have-on-the-experience-of-telecoms-customers accessed 18 August 2026.

Ofcom, 'Ofcom Fines Virgin Media £28m for Repeatedly Preventing Customers from Cancelling Contracts' (8 July 2026) https://www.ofcom.org.uk/phones-and-broadband/switching-provider/ofcom-fines-virgin-media-28m-for-repeatedly-preventing-customers-from-cancelling-contracts accessed 18 August 2026.

Ofcom, 'The Era of Answer Engines: Generative AI's Impact on Search Experiences and Online Safety' (4 November 2025) https://www.ofcom.org.uk/internet-based-services/technology/the-era-of-answer-engines-generative-ais-impact-on-search-experiences-and-online-safety accessed 18 August 2026.

Thinks Insight & Strategy, Understanding Consumer Use of Generative AI: Report for the Digital Regulation Cooperation Forum (April 2025) https://www.drcf.org.uk/publications/papers/understanding-consumer-use-of-generative-ai accessed 18 August 2026.

Standards and International Principles

ISO/IEC 42001:2023, Information Technology—Artificial Intelligence—Management System https://www.iso.org/standard/42001 accessed 18 August 2026.

ISO/IEC 42005:2025, Information Technology—Artificial Intelligence—AI System Impact Assessment https://www.iso.org/standard/42005 accessed 18 August 2026.

National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1, January 2023) https://www.nist.gov/itl/ai-risk-management-framework accessed 18 August 2026.

National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1, July 2024) https://doi.org/10.6028/NIST.AI.600-1 accessed 18 August 2026.

OECD, Recommendation of the Council on Artificial Intelligence OECD/LEGAL/0449 (adopted 22 May 2019, amended 3 May 2024) https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 accessed 18 August 2026.

Books and Articles

Ben-Shahar O and Schneider CE, More Than You Wanted to Know: The Failure of Mandated Disclosure (Princeton University Press 2014).

Luguri J and Strahilevitz LJ, 'Shining a Light on Dark Patterns' (2021) 13 Journal of Legal Analysis 43–109 https://doi.org/10.1093/jla/laaa006 accessed 18 August 2026.

Nissenbaum H, Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford University Press 2009).

Nouwens M, Liccardi I, Veale M, Karger D and Kagal L, 'Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence' (2020) Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1 https://doi.org/10.1145/3313831.3376321 accessed 18 August 2026.

Solove DJ, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880.

Author's Related Work

Leiser MR, Dark Patterns, Deceptive Design, and the Law: AI's Hidden Influence on Our Digital Experience (Bloomsbury Academic 2025) https://www.bloomsbury.com/uk/dark-patterns-deceptive-design-and-the-law-9781509987115/ accessed 18 August 2026.

Leiser MR, 'Against the Sovereign User: Fundamental Rights, Technology and the Case for Relational Rights' (SocArXiv preprint, 2026) https://osf.io/preprints/socarxiv/fd3e5_v1 accessed 18 August 2026.

Leiser MR, Submission to the Digital Regulation Cooperation Forum: Consumer Interest and AI, Phase 1 Response (June 2026).

Leiser MR, 'Recalibrating the EDPB and the EDPS under a Duty to Cooperate: An Institutional Case for Cooperative Digital Administration' (unpublished policy paper, June 2026).

Leiser MR, 'The Innovation Mandate: Making Europe's Digital Supervisory State Accountable for Rights, Growth and Scale' (SSRN, 24 July 2026) https://ssrn.com/abstract=7073898 accessed 18 August 2026.