Submission to the Digital Regulation Cooperation Forum

Consumer Interest and AI

Phase 2 Response: Tools, Frameworks, Accountability and Tolerable Risk

Dr M.R. Leiser (Mark)

DigiData Consulting

Personal capacity

July 2026

1. Scope and orientation of this Phase 2 submission

This submission primarily addresses the DRCF questions for Phase 2 (Q17-Q28), concerning the tools, frameworks, standards, compliance obligations, accountability models, and risk metrics that should be available to consumers, industry, regulators, and policymakers. It builds on my Phase 1 submission, which argued that the regulatory object should not be "AI" in the abstract, but the way AI changes 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. [1][2]

The organising concept remains proportional accountability. Consumer-facing AI should not be romanticised, feared, or regulated by technological label alone. It should be governed according to what it does to consumers, markets and rights. The same AI family can support accessibility, translation, comparison, customer support and product quality; it can also enable personalised opacity, simulated trust, vulnerability exploitation, unauthorised agency, weak contestability and diffuse accountability. The task is to preserve the benefits while preventing the conversion of consumer weakness, inattention, dependency, opacity or vulnerability into commercial advantage. [2]

The Phase 2 call is particularly important because it asks the right second-order question. If consumer protection cannot depend primarily on individual vigilance in adaptive AI markets, what practical tools should sit above and around the consumer? The answer should not be a new layer of generic AI bureaucracy. It should be a cross-sector Consumer AI Accountability Framework that translates existing obligations - fair dealing, data protection, competition, communications, financial services, online safety, professional diligence and redress - into operational duties that firms can implement and regulators can test. [1]

This approach is also aligned with the UK's innovation agenda. The Government's AI Opportunities Action Plan and its response both emphasise that the right regulatory regime should address risks while supporting innovation, trust and adoption. [8][9] Proportional accountability is not anti-innovation. It is innovation infrastructure. It gives low-risk, genuinely assistive AI a clearer route to market while ensuring that high-impact systems cannot hide behind low-risk language.

2. Executive summary

This submission makes five claims.

1. Consumer AI protection should be built around function, not labels. The DRCF should adopt a shared risk grammar that distinguishes among **assistive**, **advisory**, **consequential, agentic**, and **systemic consumer** AI. This should determine the intensity of obligations.
2. The central industry tool should be a Consumer AI Accountability Framework built around classification, purpose, control, evidence, outcomes, redress and cooperation. The framework should operate through risk-calibrated evidence packs, rather than one-size-fits-all compliance.
3. Agentic AI requires a permission and records architecture. Where AI can spend money, accept terms, disclose data, cancel services, negotiate, complain, move funds or interact with other systems, the default should be approval before consequential action, granular permissions, consumer-readable action logs and rapid reversal.
4. Consent and disclosure are necessary but insufficient. Consumer protection cannot be reduced to more notices, more terms or an "AI cookie banner". Consent to data processing is not consent to AI assistance, and consent to AI assistance is not authorisation for AI to act.
5. Regulators need architecture-aware tools. These include common terminology, AI journey testing, synthetic-user and mystery-shopping powers, access to audit trails and optimisation objectives, incident taxonomies, cross-regulatory triage, safe harbours for low-risk AI and red lines for exploitative optimisation.

The submission therefore recommends that the DRCF develop a Consumer AI Accountability Framework, a practical risk ladder, a standardised AI action receipt, an agentic permission architecture, a consumer AI

incident taxonomy, standard supplier clauses, cross-sector scenario guidance and a set of innovation-friendly safe harbours for low-risk AI.

3. Core recommendations

Recommendation	Practical implication
A functional risk ladder	Adopt a shared DRCF risk ladder: assistive AI, advisory AI, consequential AI, agentic AI and systemic consumer AI.
Consumer AI Accountability Framework	Require firms to maintain risk-calibrated evidence packs covering classification, purpose, control, evidence, outcomes, redress and cooperation.
Prohibited optimisation objectives	Make firms state what their AI system must not optimise for: manipulation, complaint abandonment, cancellation friction, vulnerable-user conversion, revenue through confusion or suppression of statutory rights.
Agentic permission architecture	For agentic AI, require granular, revocable, time-limited permissions; default approval before consequential actions; spending limits; counterparty limits; data-disclosure limits; and action logs.
Redress-by-design	Require quick reversal routes, evidence preservation, human escalation capable of altering outcomes and no-fault correction for unauthorised agentic action.
Regulator access to evidence	Give regulators access to audit trails, prompt libraries, A/B testing records, optimisation objectives, supplier documentation and incident records where consumer harm is suspected.
Safe harbours for low-risk AI	Create safe harbours for genuinely assistive, bounded, reversible, well-disclosed and well-governed AI, especially for SMEs.
Cross-sector triage	Develop common intake categories and referral protocols so consumers do not need to determine whether an AI-related harm falls under the CMA, ICO, Ofcom, or FCA before seeking help.

4. Recent developments that should inform Phase 2

Several developments since the Phase 1 submission sharpen the need for an operational framework.

First, UK regulators are already moving from abstract AI principles to sector-specific implementation. The CMA's 2026 guidance on AI agents makes clear that consumer law applies whether customers interact with a person or an AI agent, and that a business is responsible for what an AI agent does in the same way it is responsible for an employee, including where a third party designed or supplied the system. [4] That proposition should be foundational for the DRCF: agentic AI must not become a liability-laundering device.

Second, the ICO's Tech Futures work recognises that agentic AI combines generative AI with tools and new ways of interacting with the world, thereby increasing systems' ability to work with contextual information and to automate open-ended tasks. [6] This reinforces the distinction between answer systems and action systems. Once AI moves from informing to acting, the governance problem changes from accuracy alone to mandate, authority, evidence, reversibility and accountability.

Third, the FCA's Mills Review, published on 6 July 2026, frames AI in retail financial services as a shift toward AI-enabled, continuous and delegated consumer journeys. The FCA page also notes that its current approach relies on principles-based and outcomes-focused frameworks rather than on separate AI-specific rules, and is supported by AI Live Testing and the Supercharged Sandbox. [7][20] That is consistent with the approach

proposed here: outcomes-based duties should be operationalised through evidence, testing, and redress, rather than replaced by generic AI legislation.

Fourth, the Data (Use and Access) Act 2025 makes the UK framework for significant automated decision-making more permissive while retaining safeguards: people must receive information about significant decisions, be able to make representations and challenge them, and obtain human intervention. [10] In consumer AI, those safeguards must be made usable. A right to challenge a decision is weak if the consumer cannot identify AI involvement, obtain an action log, or reach a human with authority to change the outcome.

Fifth, DRCF and Ofcom research on consumer use of generative AI shows that consumer-facing AI is already changing how people access information. The DRCF's 2025 research examined consumer awareness, understanding, perceived benefits and harms, including deep dives into debt, financial advice, and AI-enabled web search. [19] Ofcom's "answer engines" paper describes the shift from search tools that locate pages to generative AI search tools that generate answers. [11] This matters because consumer protection tools built for webpages, rankings, and disclosures may not work when the consumer is presented with a single, fluent answer.

Sixth, the EU's Digital Fairness work is a useful comparator because it focuses on dark patterns, addictive design, unfair personalisation, price-marketing practices and digital contracts, with particular attention to vulnerable consumers and minors. [17] The UK need not copy the EU Digital Fairness Act. But the EU process confirms that deceptive design, personalised opacity and exploitative optimisation are now central consumer-law issues, not merely privacy issues or interface aesthetics.

Finally, Ofcom's July 2026 enforcement action against Virgin Media for making cancellation difficult is not an AI case, but it is an important warning. Ofcom identified systemic and repeated failings involving call-dropping, unnecessary transfers and repeated holds that prevented or delayed switching. [18] AI can make precisely this kind of "sludge" cheaper, more personalised, and harder to evidence. The DRCF should therefore treat cancellation, complaint, and redress friction as core AI governance metrics, not marginal customer service matters.

5. A Consumer AI Accountability Framework

The DRCF should develop a Consumer AI Accountability Framework built around seven elements: classification, purpose, control, evidence, outcomes, redress and cooperation. This framework should not replace existing law. It should translate existing law into AI-mediated consumer journeys.

5.1 Classification

The first question should be functional: what does the system do? A chatbot that translates a complaint into plain English should not be treated like an agent that cancels an energy contract or accepts a financial product. The DRCF should use the following ladder.

Tier	Examples	Expected safeguards
Tier 1: Assistive AI	Grammar, translation, accessibility, basic search support, simple comparison.	Clear labelling, basic reliability, privacy compliance, complaint route.
Tier 2: Advisory AI	Recommendations, advice-like outputs, complaint triage, product steering.	Domain boundaries, confidence calibration, source grounding, escalation.
Tier 3: Consequential AI	Credit, insurance, debt, utilities, telecoms, healthcare access, essential services, eligibility, dispute resolution.	Impact assessment, fairness testing, explainability sufficient for challenge, meaningful human review.
Tier 4: Agentic AI	Purchases, cancellations, negotiation, contractual acceptance, data disclosure, payment instructions, service changes.	Permission architecture, approval before action, limits, logs, receipts, revocation, reversal.

Tier 5: Systemic consumer AI	Search, ranking, recommender systems, advertising, pricing, automated negotiation, platform governance.	Systemic risk assessment, auditability, regulator access, cross-sector monitoring and competition analysis.

5.2 Purpose

Purpose should be treated as a governance control, not a wording exercise. A consumer AI purpose statement should specify the consumer-facing function, the commercial objective, the foreseeable rights-effects and the prohibited optimisation objectives. 'Improve customer experience' is too vague. 'Reduce support time while preserving complaint completion, escalation, cancellation and statutory rights' is testable. This matters because optimisation objectives can convert lawful automation into exploitative optimisation: a system rewarded for conversion, retention, delay reduction, complaint closure or cost avoidance may learn to suppress material information, personalise friction, steer consumers away from redress or exploit vulnerability events. [4][11][12] [22]

Purpose must also be relational. Consent or disclosure tells us whether a consumer clicked; it does not indicate whether the system changes the conditions under which consumers, non-users, or affected groups exercise their rights. Consumer AI may affect household members, authorised representatives, guarantors, dependants, co-travellers, or consumers who never see the model, because they are ranked, priced, routed, or delayed behind the interface. The DRCF should therefore require a rights-effects analysis: who is affected, what is altered, who benefits, who bears risk, who can contest and which alternative design would preserve the purpose with fewer rights costs. [22][25][26]

Recent case law illustrates the same discipline. Business architecture does not define necessity; ordinary commercial convention does not justify unnecessary data collection; and commercial interests can be legitimate only through lawfulness, necessity, reasonable expectations and balancing. The point for the DRCF is not to import EU doctrine wholesale into UK consumer AI policy, but to use the underlying method: firms should be required to evidence why the selected purpose is lawful, specified, consumer-supportive and no broader than the consumer objective requires. [29][30][31]

Purpose discipline also supports innovation. The UK pro-innovation agenda is best served by a framework that identifies lawful, beneficial AI uses and the safeguards that let them scale. Firms should record the less burdensome rights-preserving alternatives considered, and regulators should avoid treating bounded, reversible, and genuinely assistive AI as high risk merely because it uses AI. The test should be functional: what is the system for, what must it not optimise for, and what consumer outcome will show that the purpose is being met? [8][9][23]

5.3 Control

Control should not be collapsed into consent. Consumers need controls at points where control can perform legal and practical work: approval before action, spending limits, data disclosure limits, permission expiry, revocation, access to a human, and a record of what happened. Consent to data processing is not consent to AI-mediated assistance, and consent to AI-mediated assistance does not authorise an agent to purchase, cancel, accept terms, move funds, or disclose data. This distinction is essential for agentic AI because the risk is not merely misinformation; it is delegated action. [4][6][10][22]

Firm-side control is just as important. A consumer-facing firm should identify accountable owners for the AI journey, objective setting, deployment approval, release management, supplier changes, incident response, system pause, and redress. This fits the CMA's 2026 position that businesses remain responsible for AI agents used with customers, and it avoids treating the ordinary consumer as the systems integrator of last resort. [4][5]

Controls should be safe by default and proportionate to function. Assistive AI may require labelling, basic reliability and complaint routes. Consequential AI requires meaningful human escalation, fairness testing and reversibility. Agentic AI requires a mandate architecture: granular, revocable and time-limited permissions; contemporaneous confirmation before consequential action; and receipts showing what the system did and under what authority. The Data (Use and Access) Act 2025 points in the same direction for significant

automated decisions: wider use of automation is acceptable only where information, challenge and human intervention are real rather than formal. [10]

5.4 Evidence

Evidence is the operational core of accountability. The DRCF should resist evidence-light governance: generic AI labels, model cards or supplier assurances are insufficient where AI affects money, rights, eligibility, essential services, complaints or redress. Firms deploying consequential or agentic consumer AI should maintain an evidence pack covering purpose, data sources, system version, relevant instructions or prompts, permissions, testing, human oversight, supplier documentation, incident records, monitoring metrics and redress outcomes. This is consistent with NIST and ISO management-system approaches, but the evidence should be translated into consumer outcomes rather than technical documentation alone. [13] [14][28]

Evidence should be layered. Consumers need an intelligible record of what the AI did, when, under which permission, with what consequence, and how to reverse or challenge it. Regulators need fuller records: optimisation objectives, journey tests, A/B testing records, prompt libraries, model changes, supplier incidents, complaint metrics, vulnerability outcome gaps and escalation failures. Firms do not need to disclose trade secrets to every consumer. Still, they should not be able to defeat redress by saying that the relevant evidence is internal, proprietary or held by a supplier. [22][24][32]

The framework should also distinguish regulatory intelligence from evidence relied on to establish responsibility. Cross-regulatory cooperation may require sharing complaints, technical assessments, market signals or incident data. But material relied on to attribute harm should be traceable, lawfully obtained, disclosable where rights of defence require it and capable of being challenged. This matters because the same AI journey may raise consumer protection, data protection, competition, communications and financial services issues. [24]

5.5 Outcomes

Outcomes should be the measure of quality. Average accuracy is not enough: a system can be technically impressive while failing consumers in financial difficulty, older consumers, disabled consumers, low-literacy consumers, minority language users, children or consumers attempting to exercise cancellation or redress rights. The core consumer AI outcomes should be understanding, control, fair treatment and remedy. These map to the logic of the FCA Consumer Duty, but should be adapted cross-sectorally rather than transplanted mechanically. [27]

Outcomes should also be relational and group-sensitive. The question is not only whether one consumer received a correct answer; it is whether the system changes the choice environment for classes of consumers or affected non-users. Relevant measures include differential friction, vulnerability-outcome gaps, complaint abandonment, time to human escalation, reversal success, unauthorised action rates, source-grounding failure and whether consumers can identify the responsible firm. This approach draws on relational rights analysis: rights protection should track the effects of system design, not merely the presence of individual consent or disclosure. [22][25][26]

An outcomes model also serves the UK's innovation agenda. It gives firms room to choose technical means while requiring proof that the chosen means produce fair, contestable and reversible consumer outcomes. That is more innovation-friendly than prescriptive technology rules and more protective than principle-only supervision. The DRCF should therefore promote outcome testing before deployment, post-deployment monitoring after material changes and sector-specific tolerability thresholds for high-stakes uses. [8][9][13][20][23]

5.6 Redress

Redress should be designed into the AI journey, not added after failure. For low-stakes AI, ordinary complaint routes may suffice. For consequential and agentic AI, redress requires records, rapid escalation, interim protection, reversal, and a human with the authority to change the outcome. A right to challenge is weak if the consumer cannot know what the AI did, whether it acted within mandate, which firm controlled it or what evidence exists. [10][32]

Where a firm-controlled AI agent takes legally or economically significant action outside the consumer's mandate, the default should be no-fault correction for the consumer. The firm can allocate responsibility among developers, integrators, suppliers and insurers after correcting the consumer's position. This narrow form of strict or near-strict liability is more proportionate than strict liability for all AI harms and more realistic than requiring the consumer to prove a black-box failure from the outside. It is also consistent with the CMA's position that businesses remain responsible for AI agents used with customers. [4][5]

Relational harms may require structural remedies. If the harm arises from personalised cancellation friction, complaint abandonment, discriminatory routing, exploitation of vulnerabilities, ranking, search, or AI-generated authority effects, redress should include journey redesign, suppression of optimisation objectives, model or prompt changes, audit access, public correction, group-level remediation, and regulator coordination. Individual compensation is necessary but insufficient where the system architecture produced the harm. [18][22][24]

5.7 Cooperation

Consumer AI is cross-regulatory by design. A single system may simultaneously be a consumer interface, a data-processing operation, a financial-services support tool, a telecoms support journey, a recommender system, a platform risk vector, a fraud-control mechanism, and a competition issue. The DRCF should therefore treat cooperation as part of the accountability framework, not as a background administrative virtue. [1][21][24]

An adapted duty-to-cooperate model would help. The DRCF should create cooperation triggers where an AI system materially affects another regulator's mandate; require a short scoping note for significant cross-sector AI issues; distinguish regulatory intelligence from evidence relied on for enforcement; record material disagreements; and publish cross-regulatory scenario notes where firms need coherent guidance. The aim is not consensus at any price or a super-regulator. It is reciprocal visibility, lawful disagreement and fewer gaps in redress. [24]

Cooperation is also innovation infrastructure. Fragmented, duplicative or inconsistent requirements are not neutral: they raise compliance costs, slow lawful deployment and favour incumbents. A cooperative framework should therefore include shared terminology, incident taxonomies, safe harbours for low-risk AI, sandbox learning converted into public guidance and a single primary route for consumers where harm cuts across mandates. That would protect consumers while supporting the legal certainty, proportionality and responsible innovation commitments embedded in the UK's AI agenda. [8][9][21][23]

6. Responses to Phase 2 questions

Q17. What tools can industry offer to consumers or otherwise implement to manage AI risks?

Industry tools should not be limited to model cards, privacy notices or generic AI labels. Consumer AI is a socio-technical service. Governance must cover the consumer journey, interface, commercial objective, permission structure, supply chain, redress route and the model or system.

- Consumer AI impact assessments focused on consumer outcomes, not only data protection, cybersecurity or technical safety. These should ask who may be harmed, how harm might occur, whether harm is reversible, whether vulnerable consumers face worse outcomes and how consumers will obtain redress.
- Agentic AI permission architectures, including default approval before consequential action, granular permissions, spending limits, time limits, counterparty limits, data-disclosure limits, revocation and action pausing.
- AI action receipts showing what the AI did, when, on whose instruction, under which permission, using which data categories, with which counterparties and whether human oversight was involved.
- Pre-deployment journey testing using realistic scenarios involving children, consumers with low literacy, consumers in financial difficulty, disabled consumers and situational vulnerability events. Testing should examine confusion, overreliance, escalation failure and redress friction, not merely task completion.
- Post-deployment monitoring of complaints, reversals, unauthorised actions, hallucination harms, discriminatory outcomes, support failures, cancellation friction, complaint abandonment, escalation failures, model drift and supplier incidents.
- Supplier governance and standard contractual clauses requiring documentation, change notices, performance evidence, incident cooperation, audit support, logs and remediation assistance.
- Redress-by-design mechanisms, including quick reversal routes, human escalation with authority, complaint preservation, action pausing, interim protection and no-fault correction for unauthorised agentic actions.
- Prohibited optimisation objectives preventing systems from being rewarded for manipulation, opacity, churn prevention through friction, vulnerable-user conversion, complaint abandonment, revenue uplift through confusion or reduced exercise of statutory rights.

Consumers may benefit from knowing that such frameworks exist, but consumer-facing information about them must be simple and tied to practical rights. The useful disclosure is not "we have an AI governance framework"; it is "this system cannot spend money without approval", "you can see what it did", "you can reverse this action within X hours", and "this is who is 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?

The answer should be differentiated. Consumers should generally be able to opt out of AI when it materially affects advice, support, recommendations, complaint handling, price, eligibility, cancellation, redress, or other consequential outcomes, and when a non-AI route is reasonably feasible. A universal opt-out from all AI is neither realistic nor necessarily beneficial. AI may be embedded in fraud detection, cybersecurity, accessibility routing, service reliability or essential back-office functionality.

The DRCF should distinguish three concepts that are often conflated: consent to data processing, consent to AI-mediated assistance and authorisation for AI to act. A consumer may consent to data processing without authorising an AI system to enter a contract. A consumer may accept AI assistance without accepting AI decision-making. A consumer may authorise one purchase without authorising future purchases, data disclosures or changes to essential services.

Where data processing is relied on for consumer AI, the legal basis must be assessed under the relevant data protection regime and cannot be solved by a general AI notice. The Data (Use and Access) Act 2025 is relevant because it creates a more permissive framework for significant solely automated decisions while requiring safeguards including information, challenge and human intervention. [10] Those safeguards should be interpreted in consumer-facing contexts as requiring practical contestability, not merely formal rights.

Where a full AI opt-out is not appropriate, substitute protections should include purpose limitation, fairness testing, meaningful human review, outcome transparency, auditability, redress, and restrictions on the use of inferred vulnerabilities or sensitive traits for commercial optimisation. Consumers should not receive a degraded service merely because they decline unnecessary AI personalisation.

Q19. What tools, frameworks or standards can regulators leverage to safeguard consumer protection?

Regulators should leverage existing standards, but adapt them to consumer outcomes. NIST's AI Risk Management Framework and Generative AI Profile are useful because they emphasise governance, testing, risk mapping and incident response. [13] ISO/IEC 42001 is useful because it provides a management-system approach to AI governance, including policies, objectives, processes and continual improvement. [14] The OECD AI Principles provide a high-level international foundation for trustworthy and human-centred AI. [15]

However, general AI standards will not be enough. Consumer protection requires additional questions: does the system affect price, eligibility, cancellation, support, complaint routes, contractual terms, redress, vulnerability, or consumer choice architecture? Does it preserve autonomy and reversibility? Does it create evidential records? Does it allocate accountability in a way the consumer can use?

The DRCF should therefore develop a cross-sector Consumer AI Accountability Framework that maps general AI standards onto consumer outcomes. This would avoid two errors: pretending that existing AI standards automatically deliver consumer fairness, and creating bespoke rules that ignore internationally recognised governance practices.

Q20. What types or modes of guidance or other regulatory tools are most helpful to industry?

Industry needs guidance that is concrete enough to operationalise and proportionate enough not to deter the use of useful AI. The most helpful outputs would be:

- A shared DRCF consumer AI scenarios document mapping common use cases to likely regulatory hooks and expected safeguards. Examples should include AI customer support, complaint triage, telecoms switching, debt advice, AI shopping agents, personalised pricing, insurance eligibility, AI search, subscription management and AI-to-AI negotiation.
- Tiered checklists for SMEs and larger firms. Low-risk assistive AI should have a short checklist; consequential and agentic AI should require fuller evidence packs.
- Template AI action receipts and permission dashboards, tested with consumers.
- Standard clauses for AI procurement and supplier governance, including audit support, incident cooperation, change notices and evidence preservation.
- Regulatory sandboxes and live testing environments for beneficial AI, paired with clear red lines for exploitative optimisation, unauthorised agentic action, misleading AI presentation and AI systems that frustrate statutory rights.
- A public set of regulatory anti-patterns: AI consent pop-ups without real choice, trust marks without audit, human oversight that cannot alter outcomes, opt-outs that degrade service, and terms that disclaim foreseeable harms caused by firm-controlled AI.

The DRCF AI and Digital Hub is a useful model because it offered cross-regulatory advice to innovators from the CMA, Ofcom, ICO and FCA in one place, with the stated aim of helping products reach market responsibly, faster and with greater confidence. [21] That cross-regulatory model should be extended from advice into reusable scenario guidance and templates.

Q21. What is the current application of AI standards, and is there international best practice?

The current standards landscape is useful but incomplete. NIST AI RMF and ISO/IEC 42001 provide governance structures; OECD principles provide international normative alignment; the EU AI Act provides an example of risk-based legal classification; and the emerging EU Digital Fairness agenda focuses on digital consumer harms such as dark patterns, unfair personalisation and digital contracts. [13][14][15][16][17]

International best practice for the DRCF is not to copy any one model. The UK's strength is context-specific regulation through existing regulators. The weakness of that approach is fragmentation unless regulators share

vocabulary, scenarios, metrics and evidence expectations. The DRCF can therefore add distinctive value by creating interoperability between sectoral regimes: a common consumer AI grammar, not a single AI code.

A practical international lesson is that general risk management must be joined to specific consumer-law harms. A firm can have an AI management system and still deploy AI that personalises cancellation friction, obscures price, overstates authority, or makes redress impossible. The DRCF's framework should therefore require consumer-outcome evidence, not merely AI governance documentation.

Q22. Is informed and meaningful consent the right lever?

Consent is sometimes necessary but cannot be the central lever for consumer AI protection. Meaningful consent requires comprehension, voluntariness and genuine alternatives. These conditions are often weak where AI is embedded into essential services, customer support, search, recommendations, pricing, complaint handling or fraud prevention. Consent performs especially poorly where the system is adaptive, and the consumer cannot know in advance how it will behave.

The DRCF should resist the creation of "AI cookie banners". More consent prompts may increase fatigue without increasing protection. The more useful route is machine-readable permission and preference infrastructure: consumers should be able to set enduring limits on what agents may do, while firms remain responsible for respecting those limits and for designing systems that do not exploit predictable inattention.

Design constraints, default limits, confirmation requirements, action receipts, human escalation, cooling-off periods, redress rights, prohibited optimisation objectives and regulator access to evidence should supplement consent. The consumer should not be asked to consent to a risk that the firm itself has not properly identified, tested or governed.

Q23. Are there other levers?

Yes. The most important levers are structural and procedural rather than informational:

- Approval before consequential action: an agent should not spend money, accept terms, disclose sensitive data, cancel essential services or change contractual status without confirmation.
- Cooling-off and reversal: high-impact AI-mediated transactions should include practical reversal rights and action pausing.
- AI action receipts: consumers should receive records of consequential AI actions.
- Permission dashboards: consumers should be able to see, limit and revoke agent permissions.
- Human escalation capable of changing outcomes: "human in the loop" is insufficient if the human cannot intervene.
- Redress routing: consumers should have a single primary point of accountability with the consumer-facing firm.
- Prohibited optimisation objectives: firms should not optimise for complaint abandonment, cancellation friction, vulnerable-user conversion or revenue through confusion.
- No-fault correction for unauthorised agentic action: consumers should not bear interim loss while suppliers and deployers debate causation.
- Record-preservation duties: consequential AI systems should preserve the evidence needed to investigate harm.
- Vulnerability-use restrictions: detected vulnerability should be used to support the consumer, not to increase price, friction, pressure or commercial extraction.

Q24. Are there existing examples in other sectors or jurisdictions that utilise tools efficiently?

The FCA Consumer Duty is the most important domestic comparator because it shifts attention from formal disclosure to consumer outcomes, including understanding and support. The DRCF should adapt the logic, not transplant the rulebook wholesale. The relevant consumer AI outcomes should be understanding, control, fair treatment and remedy.

The FCA's Supercharged Sandbox is a useful innovation-support comparator because it provides secure environments, datasets and expert support for firms developing AI use cases. [20] The DRCF AI and Digital Hub is another model because it reduces cross-regulatory uncertainty for innovators. [21] These tools should be paired with safeguards: sandboxes should not be used for reputational laundering, and participation should not be treated as proof of compliance beyond the tested scope.

The EU AI Act is useful as a risk-based comparator, particularly because it distinguishes unacceptable, high-risk and lower-risk AI. [16] The Digital Fairness Act process is useful because it recognises that dark patterns, unfair personalisation, and digital contracts require attention under consumer law. [17] The UK can learn from both while retaining a more agile, sectoral approach.

Ofcom's enforcement of cancellations is also relevant as a non-AI example. It demonstrates that friction in cancellation and switching can cause direct consumer harm even without AI. [18] The DRCF should treat AI-enabled friction as a high-priority enforcement and monitoring problem, particularly where friction is personalised or dynamically allocated.

Q25. Can outcomes-based approaches analogous to the Consumer Duty support consumer protection?

Yes. An outcomes-based approach is the best fit for consumer AI, provided outcomes are measured and evidenced rather than merely asserted. A consumer AI version of the Consumer Duty should focus on four outcomes:

1. Understanding: consumers know when AI materially affects them and what it can and cannot do.

2. Control: consumers can approve, limit, revoke, opt out where appropriate, contest and escalate.

3. Fair treatment: the AI does not produce unjustified differential outcomes or exploit vulnerability, inattention or dependency.

4. Remedy: harm can be identified, investigated and corrected without imposing impossible evidential burdens on the consumer.

This approach fits the UK's pro-innovation model because it does not require prescriptive rules for every AI use case. It requires firms to evidence that the systems they deploy are designed, tested and monitored to produce fair consumer outcomes. The more consequential, opaque, autonomous, or difficult to reverse the system is, the stronger the evidence should be.

The FCA's Mills Review reinforces this direction. The FCA page describes the FCA's approach as relying on existing, principles-based and outcomes-focused frameworks, combined with proactive industry engagement and support through AI Live Testing and sandboxes. [7] This is precisely the model the DRCF should adapt cross-sectorally: principles plus evidence, not principles as slogans.

Q26. Who should be held accountable, and to what extent, for AI risks materialising?

Accountability should track control, capability and benefit. The primary accountable actor should normally be the consumer-facing firm that deploys AI into the consumer journey, presents it as part of its service, benefits from the interaction and controls the route to redress. That firm can allocate risk to suppliers through contract. Still, the consumer should not have to prove whether the fault lay with the model developer, cloud provider, API, data broker, prompt layer, integration layer or deployment context.

For agentic AI, accountability should be heightened because the system can act. Where a firm-controlled AI agent makes or causes an unauthorised transaction, accepts terms outside the consumer's mandate, discloses data outside permission, cancels or alters a service without authority, or fails to honour statutory rights, the consumer should receive rapid correction and should not bear interim loss while firms debate causation. This supports a limited strict or near-strict liability model for unauthorised agentic actions, without imposing strict liability for every AI error.

For other AI harms, a rebuttable presumption is more proportionate. Where the firm controlled the system, set the objective, benefited from the interaction, and held the evidence, the burden should shift to the firm to show that the AI system was properly designed, tested, monitored, and remediated. This is fairer than universal strict liability, but more realistic than requiring consumers to prove a black-box failure from the outside.

Developers and upstream providers should not be invisible. They should owe obligations to provide documentation, safety information, change notices, incident support and audit cooperation to deployers. But consumer redress should remain centred on the public-facing service provider.

Q27. Do tolerable risks exist for consumers, and how should they be operationalised?

Tolerable risks do exist, but they are contextual. A risk may be tolerable where the likely harm is low, the consumer benefit is clear, the system is sufficiently transparent for the context, the consumer remains in control, the harm is reversible, and redress is accessible. A risk becomes less tolerable where the system affects high-stakes interests, acts autonomously, personalises vulnerability, obscures responsibility, produces irreversible consequences or shifts loss onto the consumer.

Some risks should be presumptively intolerable unless exceptional safeguards are in place:

- Unauthorised financial transactions or payment instructions.
- Hidden acceptance of legal terms or contractual changes.
- Discriminatory access to essential services.
- Exploitative targeting of children or consumers in vulnerable circumstances.
- AI systems that obstruct cancellation, complaints, statutory rights or redress.
- Agentic systems that act outside the consumer's mandate.
- Back-office systems that use vulnerability detection for price increases, retention friction or complaint deprioritisation.
- AI systems that create material consumer effects without preserving evidence for challenge.

The DRCF should therefore distinguish tolerated risk from transferred risk. Consumers may tolerate some risk where they understand it, benefit from it, remain in control and can obtain redress. Consumers should not be treated as having tolerated risks that firms have hidden, externalised, buried in terms, converted into personalised opacity or made practically impossible to contest.

Q28. What metrics, thresholds or proxy indicators are feasible?

Metrics should combine technical performance, consumer journey outcomes, redress outcomes and governance indicators. The DRCF should avoid a single metric for all consumer AI. Thresholds should be sector-specific because a tolerable hallucination rate for entertainment is not tolerable for debt advice, insurance eligibility or telecoms cancellation.

Metric category	Examples
Transaction metrics	Unauthorised action rate; failed-confirmation rate; transaction reversal rate; financial-loss remediation time; chargeback or refund escalation.
Advice metrics	Consequential hallucination rate; source-grounding failure; abstention rate; escalation rate; correction rate; overconfidence indicators.
Fairness metrics	Disparate pricing, eligibility, support, complaint and cancellation outcomes; vulnerability-outcome gaps; differential friction by consumer group or inferred value.
Control metrics	Permission revocation rate; opt-out rate; override rate; approval-before-action compliance; number of actions taken without contemporaneous confirmation.
Redress metrics	Complaint volume; complaint abandonment; average time to human escalation; reversal time; evidence-production time; success rate of human escalation.
Journey metrics	Drop-off at key rights-exercising points; time to cancel; number of transfers; repeated prompts; friction added to complaint or cancellation flows.
Governance metrics	Model drift; recurring incident classes; unresolved audit findings; supplier incident response time; missing logs; overdue risk assessments; unresolved vulnerability findings.
Systemic metrics	Market-wide pricing effects; ranking concentration; correlated agent behaviour; dominant-agent steering; repeated cross-platform error patterns.

The DRCF should also encourage regulators to use synthetic users, mystery shopping and controlled testing environments to detect differentiated treatment and personalised friction. Many AI harms will not be visible from a single screenshot. They may arise from A/B testing, prompt chains, interaction history, inferred vulnerability, ranking logic or optimisation objectives. Regulators therefore need architecture-aware evidence powers.

7. A practical implementation model

The DRCF could implement this framework in stages.

Stage 1: Common vocabulary. Publish working definitions of assistive AI, advisory AI, consequential AI, agentic AI, systemic consumer AI, material consumer effect, vulnerability event, AI-mediated choice architecture, AI action receipt, meaningful human escalation and redress-by-design.

Stage 2: Scenario guidance. Publish cross-regulatory scenarios mapping consumer AI use cases to likely safeguards. This should include specific cases: AI customer support, complaint triage, AI shopping agents, telecoms switching, debt advice, financial guidance, AI search, subscription management, personalised pricing and AI-to-AI negotiation.

Stage 3: Evidence packs. Develop short, medium and enhanced evidence-pack templates. Low-risk AI should not face high-risk documentation burdens. High-impact and agentic AI should require permission design, evidence-based testing, supplier documentation, incident procedures, redress pathways, and named accountable owners.

Stage 4: Consumer-facing controls. Standardise core controls: approve before action; never spend without confirmation; do not disclose sensitive data; do not change essential services; maximum spend; permission expiry; revoke permissions; view action log; reverse action; and escalate to human.

Stage 5: Regulatory tools. Develop a shared incident taxonomy, regulator access protocols, synthetic-user testing methods, mystery-shopping powers and cross-regulatory triage for complaints that cut across mandates.

Stage 6: Safe harbours and red lines. Create safe harbours for genuinely assistive, bounded and reversible AI, especially for SMEs. Pair these with red lines for exploitative optimisation, unauthorised agentic action, AI systems that frustrate statutory rights, misleading AI presentation and use of vulnerability detection for commercial exploitation.

8. Safe harbours and anti-patterns

A proportionate framework should include safe harbours for lower-risk and well-governed AI. Safe harbours could apply where the system is assistive rather than agentic, does not materially affect price or eligibility, does not use sensitive or inferred vulnerability data for commercial optimisation, provides behaviourally meaningful disclosures, preserves human escalation, maintains appropriate records and offers accessible redress.

Safe harbours should not be based on labels chosen by firms. A firm should not be able to call a system "assistive" if it is optimised for conversion, retention, cross-selling or complaint abandonment. Nor should it be able to call a system "low risk" if it affects vulnerable consumers, essential services, regulated products or legal rights. A safe harbour should reward accountable design, not self-description.

The DRCF should also identify regulatory anti-patterns. These include:

- More terms and conditions instead of usable controls.
- AI trust marks without audit or incident consequences.
- Consent requests without genuine alternatives.
- AI opt-outs that degrade the service.
- Human oversight cannot alter outcomes.
- Audit trails that firms cannot produce when harm occurs.
- Guidance so abstract that only large firms can operationalise it.
- Supplier contracts that push evidential gaps onto consumers.
- AI disclosure labels that reveal the presence of AI but not its authority or consequences.
- Personalisation that changes friction, disclosure, support or redress asymmetrically.

9. Conclusion

Consumer-facing AI should be governed by its effects on consumers, markets, and rights. Generative AI creates risks of misplaced authority, hallucinated confidence and unclear responsibility. Agentic AI poses risks to action, authorisation, transaction and control. Back-office AI poses risks of invisibility, differential treatment and contestability. Systemic AI poses risks to market-wide choice, ranking, price and competition conditions. All can be useful. All can be harmful. The task is to regulate function, not mythology.

The DRCF is well-placed to lead this work because consumer AI risk rarely fits within a single regulatory silo. The same AI system may raise consumer protection, data protection, competition, financial services, communications, online safety and fraud issues. The question is not which regulator owns AI. The question is whether the regulatory system can develop a shared risk grammar that protects consumers, supports innovation and avoids fragmented, duplicative or inconsistent obligations.

The best answer is proportional accountability. Low-risk AI should not be smothered by high-risk compliance machinery. High-impact AI should not be shielded by low-risk language. Consumer protection and innovation are not opposing goals. Clear, risk-calibrated accountability can increase consumer trust, reduce regulatory uncertainty and create better conditions for AI services that are useful, fair and commercially sustainable.

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

Selected references

[1] 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.

[2] M.R. Leiser, "Submission to the Digital Regulation Cooperation Forum: Consumer Interest and AI, Phase 1 Response" (June 2026).

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

[4] 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.

[5] 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.

[6] Information Commissioner's Office, "ICO Tech Futures: Agentic AI" (2026), 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/.

[7] Financial Conduct Authority, "Review into the long-term impact of AI on retail financial services (The Mills Review)" (published 6 July 2026), https://www.fca.org.uk/publications/calls-input/review-long-term-impact-ai-retail-financial-services-mills-review.

[8] Department for Science, Innovation and Technology, AI Opportunities Action Plan (13 January 2025), https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan.

[9] Department for Science, Innovation and Technology, Government Response to the AI Opportunities Action Plan (13 January 2025), https://assets.publishing.service.gov.uk/media/678639913a9388161c5d2376/ai_opportunities_action_plan_government_repsonse.pdf.

[10] Department for Science, Innovation and Technology, "Data (Use and Access) Act 2025: data protection and privacy changes" (27 June 2025), https://www.gov.uk/guidance/data-use-and-access-act-2025-data-protection-and-privacy-changes.

[11] 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.

[12] Ofcom, "Exploring what impact the adoption of AI 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.

[13] National Institute of Standards and Technology, AI Risk Management Framework and Generative AI Profile, https://www.nist.gov/itl/ai-risk-management-framework.

[14] ISO/IEC 42001:2023, Information technology - Artificial intelligence - Management system, https://www.iso.org/standard/42001.

[15] OECD, Recommendation of the Council on Artificial Intelligence, updated 2024, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

[16] European Commission, "AI Act" policy page, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[17] European Commission, "Review of EU consumer law / Digital Fairness Act" (2024-2026), https://commission.europa.eu/law/law-topic/consumer-protection-law/review-eu-consumer-law_en.

[18] 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.

[19] Digital Regulation Cooperation Forum, "Understanding Consumer Use of Generative AI" (11 April 2025), https://www.drcf.org.uk/publications/papers/understanding-consumer-use-of-generative-ai.

[20] Financial Conduct Authority, "Supercharged Sandbox" (2026), https://www.fca.org.uk/firms/innovation/supercharged-sandbox.

[21] Digital Regulation Cooperation Forum, "AI and Digital Hub", https://www.drcf.org.uk/ai-and-digital-hub.

[22] M.R. Leiser, 'Against the Sovereign User: Fundamental Rights, Technology and the Case for Relational Rights' (SocArXiv / OSF preprint, 2026), https://osf.io/preprints/socarxiv/fd3e5_v1.

[23] M.R. Leiser, 'The Innovation Mandate: Making Europe's Digital Supervisory State Accountable for Rights, Growth and Scale' (policy paper, July 2026).

[24] M.R. Leiser, 'Recalibrating the EDPB and the EDPS Under a Duty to Cooperate: An Institutional Case for Cooperative Digital Administration' (policy paper, June 2026).

[25] Daniel J. Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880.

[26] Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford Law Books 2009).

[27] Financial Conduct Authority, 'Consumer Duty', https://www.fca.org.uk/firms/consumer-duty; Financial Conduct Authority, FG22/5 Final Non-Handbook Guidance for Firms on the Consumer Duty (July 2022), https://www.fca.org.uk/publication/finalised-guidance/fg22-5.pdf.

[28] ISO/IEC 42005:2025, Information technology - Artificial intelligence - AI system impact assessment; see ISO, 'Artificial intelligence: top standards', https://www.iso.org/sectors/it-technologies/ai.

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

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

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

[32] Case C-203/22 CK v Magistrat der Stadt Wien and Dun & Bradstreet Austria EU:C:2025:117.