Protecting
Digital Citizens

A policy playbook for policymakers and the digital service industry

AGENCY RESEARCH

TABLE OF CONTENTS

What is this playbook?

This playbook includes plays for policymakers and the digital service industry to protect citizens against complex online harms and ensure that technology is developed responsibly.

Who is this playbook for?

Our playbook is for policymakers interested in exploring new policy approaches to protecting citizens against complex online harms. It is also for organisational decision-makers within the digital service industry interested in building trust in digital products and safeguarding users.

Why use this playbook?

As digital technology becomes increasingly integral to all aspects of modern life, it is essential to address its capacity for complex harm and ensure its responsibly developed. This playbook equips policymakers and the digital services industry with targeted strategies, practical tools, and clear guidance to mitigate risks while fully harnessing digital technologies' innovative potential.

How to use this playbook?

This playbook begins by outlining the complex harms associated with digital technologies. It then presents recommended strategies for policymakers and the digital services industry to safeguard citizens. To illustrate the practical application of these measures, three use cases - disinformation; FemTech; and smart homes - are included.

Who developed this playbook?

This Playbook was authored by the members of the Legal and Ethical Regulation of Complex Online Harms work package forming part of the multi-disciplinary and multi-university AGENCY project.

The playbook has benefited from the insights from the wider AGENCY project and has been supported by the EPSRC under Grant EP/W032481/2.

Executive Summary

Citizens must be safeguarded from complex online harms and the unintended consequences of new digital technologies.

Yet advances in digital technologies pose unique challenges, introducing vulnerability concerns and broader safety issues. For instance, the spread of disinformation, privacy risks in FemTech, and the misuse of smart home technologies illustrate how novel innovations can inadvertently undermine public trust and individual well-being.

This playbook serves as a strategic guide for policymakers and industry leaders, providing a structured approach to mitigating online harms. It advocates for a balanced co-regulatory model that strengthens legal oversight while enabling industry self-regulation to embed ethical safeguards within digital products and services from inception.

In the sections that follow, we define the concept of complex online harms, examine their intersectional nature, and present a two-pronged approach:

1. **Recommends that policymakers** implement an **'Agency by Design' co-regulatory framework** that prioritises user autonomy, transparency, and ethical digital governance.
2. **Supports industry leadership in Corporate Digital Responsibility (CDR)** and **Responsible Research and Innovation (RRI)** integrating ethical considerations into the technology design, deployment, and governance.

By advocating for a co-regulatory ecosystem that aligns regulatory intervention with industry-led responsibility, this playbook provides a blueprint for protecting citizens and ensuring that digital technologies produce positive outcomes for society, the economy, and the environment.

4

Are users protected from online harms?

In adopting a co-regulatory approach, our **Agency by Design** approach places a premium on user empowerment, proactive governance, and industry-led **Corporate Digital Responsibility (CDR)** / **Responsible Research and Innovation (RRI)** initiatives to ensure that digital products and services are designed with ethical safeguards from the outset. This approach shifts the focus from reactive compliance to proactive, anticipatory governance that relies on companies to assume responsibility for customers as they are your customers and your responsibility and can be affected by complex online harms.

At their most simple complex online harms are intersectional harms that have complex technical and societal causes. They may encompass:

- **Disinformation and Manipulated Content**: The spread of misleading or deceptive information that can affect citizens' agency and trust in media. This issue is exacerbated by AI-generated deepfakes, synthetic media, and algorithmically amplified misinformation, which can distort reality and erode citizens' ability to make informed choices.
- **Data Privacy in FemTech Products**: FemTech applications, including menstrual tracking apps, fertility monitors, and digital health platforms, process highly sensitive personal data. Inadequate data protection measures can lead to privacy violations and potential misuse of user information.
- **Unintended Consequences of IoT Technology**: The proliferation of smart home devices has introduced unforeseen risks, such as intimate partner abuse facilitated through IoT technology. Abusers may exploit remote access controls, surveillance features, or automated functionalities to exert coercive control over victims.

Our research has shown that effectively mitigating complex online harms requires a dual strategy that aligns policy intervention with industry-led responsibility. To this end propose a two-pronged approach that:
Empowers policymakers to adopt an Agency by Design framework, ensuring that regulatory mechanisms proactively safeguard user autonomy, transparency, and ethical digital governance.

Mandates industry leadership in Corporate Digital Responsibility (CDR) and Responsible Research and Innovation (RRI) to embed ethical safeguards within digital products, services, and platforms from inception. By integrating these complementary approaches, we aim to establish a co-regulatory ecosystem where policymakers set clear, agency-enhancing standards, and the industry assumes active responsibility for designing and deploying digital technologies that prioritise user rights, security, and informed decision-making. The following sections detail each pillar of our proposals, providing plays for policymakers and industry to adopt.

What can policymakers do to protect users?

7 principles for system safety: Agency by Design

01 Incorporate diverse voices in system design

02 Provide transparent and intelligible information

03 Offer granular user control tools

04 User-defined safety and privacy settings

05 Collaborative agency and social resilience

06 Meaningful feedback and redress

07 Whole of life cycle support

Agency by Design in more detail

To protect users from complex online harm, we advocate for a co-regulatory approach grounded in the Agency by Design framework. Drawing on the provided sources and conversation history, here are seven policies for enhancing user empowerment and control over different technologies, which can be used in a technology-neutral way to devise/reflect international standards and best practices:

Incorporate Diverse Voices in System Design

Firms providing technologies to end users should actively involve diverse user groups in designing and developing their systems with a view to user safety. This includes individuals from marginalised or vulnerable groups, drawing insights from women, racialized communities, and people with disabilities, who are disproportionately affected by online harms. Their lived experiences and insights can help identify potential biases, blind spots, and gaps in knowledge within existing systems. This participatory design approach shifts users from passive consumers to active collaborators, fostering a sense of ownership and agency.

Provide Transparent and Intelligible Information

Technology providers should clearly and accessibly explain the functions of their systems, the terms upon which a service is provided, and the potential risks or issues arising from system misuse. Users should be able to understand how these systems function, how their data is used, and how to ensure device or app security. This transparency empowers users to make informed choices about their interactions and adjust their preferences accordingly, including which security setting to use, and how.

Offer Granular Control Tools to Users

Users should have access to a range of tools that allow fine-grained control over their security and safety preferences. This could include options to filter content based on keywords, topics, sources, and user accounts in the context of social media, or allow for different forms of conditional access for smart home technologies. Technology service providers should empower users to take responsibility for their own safety on terms most suitable for their circumstances.

Agency by Design (continued)

Enable User-Defined Safety and Privacy Settings

Users should have control over their privacy and safety settings, allowing them to manage their data sharing, visibility, interactions, and who is able to access their devices and on what terms. This could include limiting the collection and use of personal data, setting up different accounts for smart devices with different levels of access and control, with the possibility of 'emergency response' functions to either disable individual accounts or devices.

Promote Collaborative Agency and Social Resilience

Technology service providers should facilitate collaborative agency among users by providing tools and features that enable them to collectively address harmful content, security flaws, or behaviours. This could involve mechanisms for users to report potential safety concerns of an app collectively, support one another, and crowdsource possible solutions.

Facilitate Meaningful Feedback and Redress Mechanisms

Technology service providers should establish accessible and responsive channels for users to provide feedback on safety concerns, drawing from their collaborative discussions. They should also provide clear and transparent mechanisms for users to report, with clear, transparent, and timely responses to these reports. Effective feedback and redress mechanisms are crucial for ensuring user trust and enhancing online safety.

Whole of life cycle support

Technology service providers should maintain levels of support, including security updates, collaborative discussion and reporting mechanisms, and redress for identified safety concerns for the entirety of a product or service's life cycle. In particular, if a technology is being discontinued, specific policies to ensure safety of critical devices during the End of Life (EoE) period of a service should be adopted.

By implementing these policy recommendations, technology service providers can move towards a model of agency by design that genuinely empowers users and fosters a safer and more inclusive digital environment. However, this must be supported by the digital service industry, with policy makers providing oversight of compliance with these principles through the establishment of a co-regulatory system.

What can be done by industry?

The use cases provided across the portfolio of Agency projects highlight the blurring of human physical and digital borders within the 'new normal' of a predominantly digital society post-COVID-19. As a result, across demographics and communities, consumers now span a range of digital technologies and reside in a permanent 'online' status. Necessarily, a sharp focus has centred on technologies that now form part of, and shape our daily lives via digitisation, digitalisation, and digital transformation. Tempered with the increased adoption of technological use and solutions, and drawing on the benefits of convenience and frictionless transactions (social and economic), questions arise surrounding whether all members of society indeed benefit from the digital era.

In particular, who is responsible for ensuring that the design of technology, the subsequent gathering of data, and its reuse are managed in a responsible and authentic manner, producing positive outcomes for society, the economy, and the environment?

To this end, we present the combination of **Corporate Digital Responsibility (CDR) and Responsible Innovation (RI)** frameworks to act as a point of reference for organisations seeking a method to navigate responsible business practices in the digital era. CDR is a set of practices and behaviours that help an organisation use data and digital technologies in ways that are perceived as socially, economically, and environmentally responsible. This overlaps with RI, similarly, a framework for developing new technologies, products, or services with careful consideration of their ethical, social, and environmental impacts. Both frameworks involve anticipating potential risks and harm, engaging stakeholders in the development process, being transparent about methods and outcomes, and maintaining accountability throughout the innovation lifecycle and beyond CDR to ensure that advancements benefit society while minimising negative unintended consequences. Through fostering an environment in which CDR and RI are embedded into daily practices and actions, an empowering Agency by Design approach to system safety is more likely to be effectively implemented.

What is a CDR/RRI Framework?

To start evaluating where to begin a responsibility journey from an organisational perspective, what follows are some tools that could be used to facilitate an Agency by Design approach to system design.

CDR/RRI share fundamental principles illustrated in the table below, which can form an entry point for evaluating where your organisation ranks in terms of existing competencies. In addition, it highlights areas for improvement to protect customers' digital experiences, creating the conditions necessary to promote Agency by Design in the provision of technology-related services to end-users.

Principles	Actions
Purpose and trust	• Establish and adhere to Digital Responsibility Code • Define & Agree on Corporate Purpose aligned with goals of CDR • Implement Strong Digital Governance e.g., Digital Ethics Board/reporting
Fair & equitable access for all	• Innovative, accessible and inclusive products & services • Promote Justice, Equity, Diversity & Inclusion • Responsible Employment Rights
Promote societal wellbeing	• Implement strong privacy and responsible data practices • Promote digital maturity skills and digital wellbeing
Consider economic and societal impact	• Plan for sustainable & responsible automation • Transparency with stakeholders with verifiable 3rd party data • Share digital economy benefits with relevant stakeholders
Accelerate progress with impact economy	• Invest in sustainability/environmental/impact returns • Use verifiable environmental offset • Accelerate and innovate sustainable consumer behaviours
Create a sustainable planet to live	• Understand and work to meet UN Sustainable Development Goals • Consider policies to innovate and go beyond Carbon Negative
Reduce tech impact on climate and environment	• Implement environmental IT strategy • Shift to renewables • Mitigate and minimise impact, minimise use of offset

Evaluating CDR and RI

The evaluation process can take the form of holding 'discovery' meetings to raise understanding and awareness of the key tenets of the responsible framework. From these initial sessions a survey can be designed and conducted around the core responsible principles to gain a competency benchmark. Such surveys can be tailored to fit your organisational needs.

Next, using the benchmarks as a starting point, workshops can be held periodically with multi-unit team members to ensure that all stakeholders feel part of the process in representing their unit and add their 'voice' to responsible practice. This is based on research into organisational change. Change is more likely to succeed when the following is in place:

- The senior leaders can make the 'right' interventions.
- Initial conditions of the organisation (i.e., addressing the organisation's culture, this factor has been recognised/accepted by the senior leaders).
- Systematic and frequent testing.
- Fast feedback loops.

Workshops provide a platform to embed the processes 1-4, create feasible use cases, and members of the workshops can act as 'responsible champions' once back in their teams. This will help increase the likelihood of successful change, where a sense of ownership pervades across the organisational hierarchy, as opposed to being imposed from the C-Suite level downwards.

How to use a CDR/RI framework

The following sections provide guidelines on how to take the first steps, suggest several tools to help shape an organisation's journey to digital responsibility, that will align to strategic and operational objectives, which can be tailored depending on your sector. These examples are based on research conducted by several organisations.

1. Find your principled approach

Set out a responsible strategy taking a whole business and product/service end to end lifecycle approach. This demands consideration from design to decommissioning, including where third-party elements are procured. Focusing on CDR Principle 1: Purpose and Trust.

Make your principles matter

Assess the current Environmental Impact and Sustainability of own technology and supply chain providers (including cloud, banking, and pensions) and create a timeline to shift to preferred state over agreed timeline. Take a whole lifecycle approach to Algorithmic (and Risk) Impact Assessment, not just Data Protection Impact Assessment (DPIA), (who, how, what, when, why, mitigations, remediation). Build the outcomes into existing agile design/develop/deploy processes and become an integral aspect of your (responsible) risk management protocols and registers.

2. Get your house in order

Create a CDR Business Case, factoring in an increase in costs but also rewards in the short, medium, and long-term for accelerating progress in the impact economy. What is the impact in investing in responsible practice for you and your clients?

Take ownership of the outcomes, recognise responsibility through internal facing and, external client facing understanding of roles and responsibilities, and who is accountable in the good times and the bad times. Lay out risk and reward clearly. **Avoid siloed working** that could potentially be born out of rapid growth and pre-existent operating models. Continue to promote inter-team/inter-disciplinary working and review communication channels.

How to use a CDR/RI framework

End to end approach, walk through your current energy consumption and sustainability, data journey, physical and cyber security practices. Internally audit it and identify the gaps. Examine collaborative projects such as the Green Software Foundation (see: https://www.thoughtworks.com/en-gb/about-us/events/green-software-foundation-summit). In addition, align the responsible actions with (1) Principles with existing (2) Practices, (3) Policies, and (4) Procedures need to align with CDR strategy and end goals. Continue to promote inter-team/inter-disciplinary working and review communication channels.

CDR/Responsible Champions, as recommended earlier top down and bottom-up approaches to compliance that pervades culture. A C-suite level champion(s) in addition to organisation wide "CDR champions"/role models/"go-to" persons who actively seek CDR innovative approaches, and stimulate company-wide feedback.

Recruitment/people services review: engage in recruitment and talent acquisition, retention, and management which actively seeks responsible skill sets, growth, and inclusion in role with budget and resource to match. This will feed into continuous evaluation of CDR/RI competence, capability, and capacity levels within the workforce to prepare for future potential development in the digital space. If not, recognise skills gaps, and recruit where possible. The CDR/RI framework can be used to attract talent with aligned values. This can extend to training and "growing your own" responsible workforce through honing competences, capability, and capacity organically across all roles.

How to use a CDR/RI framework

3. Beware of the unknown unknowns

Consequence scanning and worst-case scenario planning, first order, second order and third order effects, scope of scenario, impacted/influenced, likelihood and severity. Understand and assess your unknown unknowns. Do not be afraid to do stakeholder identification and analysis with clients in the room. There are different tools for consequence scanning, one type can be found here: Product Development with Consequence Scanning - TechTransformed (https://www.tech-transformed.com).

Engage stakeholders, understand, and assess the impact to your clients, and the ultimate end users of YOUR products and services, whether labelled/white labelled. Get diverse and inclusive perspectives of all who could be impacted or influenced economically and socially. Develop relationships with charities, communities, or groups to engage underrepresented users.

4. Be digitally and socially inclusive

Linked to the prior point, **understand that it is not what you do**, but the way that you conduct business, it that matters. As demonstrated in our scenarios, social media will capture the real-world impacts of your products and services and will report immediately to a potential global audience. There will always be an end user. Consider:

- How does what you do impact end users/society?
- Is there a viable alternative for this product/service for someone who is not digitally engaged?
- What happens if someone does not have access to devices, data/broadband, bandwidth, or network availability (4G/5G)?
- What does this product/service do for and/to the future workforce and/or future education? Could it create skills wastage? How will that skills wastage be redeployed?

How to use a CDR/RI framework

5. Be prepared to communicate

Share your vision for CDR/RI with your workforce, align your culture with your modes of communication. Every role counts and the message should reach every person in their role, irrespective of their location and position in the organisation. Be transparent about your responsible efforts publicly, whether on your website, with your clients, supply chain, or within the local community.

Develop and promote a "Crisis Comms Plan" that demonstrates transparency to your core (what happened, why it happened, remediation actions taken, how you intend to stop it happening again, what is going to happen going forwards). Such actions demonstrate your awareness of the potential of online harm, and how to take remedial actions aligned with defining your commitment to responsible business practice.

6. Thrive on feedback

Measure, review, reflect, and repeat. Measure your successes and evaluate areas for improvement as potential RI projects.

Embed a "human in the loop".

Build CDR into your Data Governance Framework. Establish strong digital governance through an internal and external facing **Digital Ethics Advisory Board** (Multi-layered, Multi-faceted, multi-disciplinary, and Multicultural. Diversity of thought, expertise, experience, and protected characteristics is key). This aligns with the EU AI Act to oversee the implementation and continued responsible use of digital technologies. Even if you aren't based in the EU, adopting this approach as a best practice future proofs your company for any regulatory alignment with EU-based standards in the future.

Sectoral co-regulation case studies

To understand how these plays can be operationalised, we have developed three use cases focusing on **Disinformation, FemTech and Smart Homes**.

Disinformation

AI-generated disinformation presents a critical challenge for digital ecosystems. Social media platforms, which rely extensively on AI algorithms to curate and amplify user-generated content, are especially vulnerable to the rapid spread of false narratives. AI-powered recommendation systems, designed to maximise user engagement, can inadvertently exacerbate the reach of misleading information by prioritising highly engaging yet often inaccurate content.

Use Case: AI-Driven Amplification of Election Disinformation

To illustrate the risks posed by AI-driven disinformation, consider the following use case:

A large social media platform, ConnectSpace, has recently encountered a surge in disinformation campaigns targeting a forthcoming national election. The platform boasts millions of active users, making it a primary source of political news and discussion. However, malicious actors have begun disseminating false narratives regarding voter eligibility, polling dates, and alleged election fraud. These misleading posts circulate rapidly, and ConnectSpace struggles to contain the rapid spread of such content.

Co-regulatory plays

1. Improve Citizens' Digital Literacy

Improving media literacy standards is key to giving users the agency to navigate disinformation. Integrating media literacy into the national curriculum would be one key way of providing citizens with the analytical skills to identify disinformation. The government underscores its commitment to ensuring that young people develop robust critical thinking abilities. These competencies include detecting bias, assessing sources' credibility, and evaluating information's reliability. Such a curriculum-wide approach ensures that all citizens gain the analytical tools to navigate an increasingly complex digital landscape and recognise disinformation.

Sectoral co-regulation case studies

2. Promote Cross-Sector Collaboration

The government should establish robust frameworks to facilitate cooperation among digital platforms, regulatory authorities, and the broader academic community. This would promote the development of best practices and the ongoing exchange of expertise and up-to-date information concerning risks associated with disinformation. Fact checking, collaboration with platform trust and safety/content moderation teams, and deprioritisation of harmful content should be maintained as standard operating procedures, underscored with memoranda of understanding as reflected in the European Union's 'Strengthened Code of Practice on Disinformation'.

3. Promote Transparency

The government should consider adopting a legal framework similar to the European Union's Political Advertising Regulation (2024/900), which provides for transparency of political advertising, including the source, financing, and target audiences for such advertising, as well as prohibiting micro-targeting of individuals where this is not fully compliant with the principles of the GDPR.

4. Promote Agency by Design in Disinformation Reporting Mechanisms

The incorporation of Agency by Design principles by platform operators would help to address these issues. Incorporating diverse voices in the design of disinformation control mechanisms, including representatives of populations more likely to be targeted by disinformation, would allow for a more comprehensive understanding of the safety risks. Providing transparency regarding a platform's disinformation policies, as well as the tools available to users, helps to empower and provide agency to vulnerable user groups. Allowing for granular control on the basis of user-defined settings will allow users to curate the forms of online content visible to them, aiding in anti-disinformation efforts, when combined with allowing for the crowdsourcing of identified forms and techniques of disinformation. Should collective reporting and redress mechanisms be incorporated, platforms can provide meaningful responses by either deprioritising content, or providing users with additional tools to filter out electoral disinformation. Finally, by providing these tools during the platform life cycle, these policies and tools can be updated and adapted as situations change.

Sectoral co-regulation case studies

FemTech

Female-oriented technologies (FemTech) refers to digital technologies focused on women's health and well-being. These technologies, including menstrual tracking apps, fertility monitors, and wearable health devices, leverage data-driven insights to give users personalised health recommendations. However, the inherently sensitive nature of the data collected - ranging from menstrual cycles and fertility patterns to behavioural and biometric information presents complex privacy, security, and ethical risks.

Use Case: Privacy Risks in AI-Driven Menstrual Tracking Apps

To illustrate the risks posed by AI-driven tracking apps, consider the following use case:

MyOvu, a femtech application, offers users detailed insights into reproductive health by tracking menstrual cycles, fertility windows, and related health indicators. The app processes large volumes of personal data, including daily self-reported symptoms, biometric information from wearable devices, and behavioural patterns gleaned from users' browsing histories. To improve predictive accuracy and offer personalised recommendations, MyOvu recently integrated AI-driven analytics and partnered with third-party health research institutes. While this integration enhances the app's capabilities, it also raises significant privacy concerns. Users have expressed growing unease over collecting and sharing sensitive health information.

Co-regulatory plays

1. Recognise FemTech Products as Medical Devices

Our research indicates that a significant proportion of FemTech products do not classify themselves as medical devices despite providing guidance on menstrual cycles and related health indicators. Under current UK law, self-classification allows these products to circumvent the more stringent regulatory requirements applicable to medical devices. This regulatory gap raises concerns regarding such products' accuracy, safety, and efficacy, underscoring that policymakers should close this gap and regulate FemTech products appropriately.

Sectoral co-regulation case studies

2. Tackle non-compliance with the GDPR

We have demonstrated that many FemTech products do not comply with GDPR, and regulators need to do more to ensure that users' data is collected. This non-compliance poses significant risks to user privacy. Regulators must intensify oversight to ensure that all data collection and processing practices within the FemTech sector strictly conform to GDPR requirements, thus safeguarding users' legal rights and interests. ISO/IEC 27001:2022 should be adopted as standard for data security principles.

3. Promote Agency by Design in FemTech App Privacy Controls

Incorporate Agency by Design principles in app design. Incorporating the voices of underrepresented women, including from religious or ethnic minorities, as well as disabled users, as they may have distinct safety and privacy concerns not currently identified by app providers. Providing transparency regarding privacy policies by ensuring clear and intelligible graphics on what data is collected and who has access to it, as well as the tools available to modify privacy settings, helps to empower and provide agency to vulnerable user groups. Allowing for granular control on the basis of user-defined settings will allow users to determine which data that they are willing to have collected, and which should be maintained as secure on the device. Collective reporting and redress mechanisms made available through in-app systems and forums can ensure women are able to identify any particular security concerns with the software, and providers can provide meaningful responses by responding to those privacy concerns, or providing new user-defined settings for privacy control. Finally, security updates and privacy refinements should be provided for the entire product life cycle, with specific EoE policies devised.

Sectoral co-regulation case studies

Smart Homes

Our research has demonstrated that smart home devices and interconnected systems introduce substantial privacy and security risks, potentially exposing users to complex digital harms. A primary concern is the widespread lack of adequate information and awareness regarding these risks, harms, and vulnerabilities—particularly among household users. This deficiency impairs users' ability to exercise informed digital autonomy and agency, thereby increasing their susceptibility to privacy breaches, unauthorised surveillance, and other forms of technological exploitation.

The opacity surrounding data collection, processing, and storage practices is a critical challenge in the smart home ecosystem. Many users are unaware of the extent to which their devices continuously gather and process personal data. This lack of transparency, combined with insufficient user control over data retention and sharing, exacerbates the risks associated with smart home technologies. Consequently, malicious actors can exploit these vulnerabilities for nefarious purposes, ranging from unauthorised surveillance to coercive control.

Use Case: Misuse of IoT-Connected Smart Speakers in Intimate Partner Abuse

To illustrate the potential real-world consequences of these security and privacy issues, consider the following use case:

VoiceMate, an IoT-enabled smart speaker, is designed to enhance convenience through voice-activated functionalities such as music streaming, home automation, and hands-free assistance. Equipped with sensitive microphones and continuous connectivity to cloud-based services, VoiceMate passively listens for activation commands, capturing and processing audio inputs throughout the day. While this technology streamlines daily activities, it also presents significant risks when misused.

Recent reports highlight incidents where individuals have weaponised VoiceMate as a tool for intimate partner abuse. In some cases, abusers have exploited the device's always-on capabilities to covertly record conversations without the victim's knowledge. Additionally, they have accessed cloud-stored audio logs to monitor or blackmail victims, leveraging sensitive recordings as a means of control. Beyond passive surveillance, perpetrators have actively manipulated the device to issue verbal threats, create an atmosphere of intimidation, and exert psychological dominance over their victims.

Sectoral co-regulation case studies

Co-regulatory plays

1. Expand the Product Security and Telecommunications Infrastructure (PSTI) Act

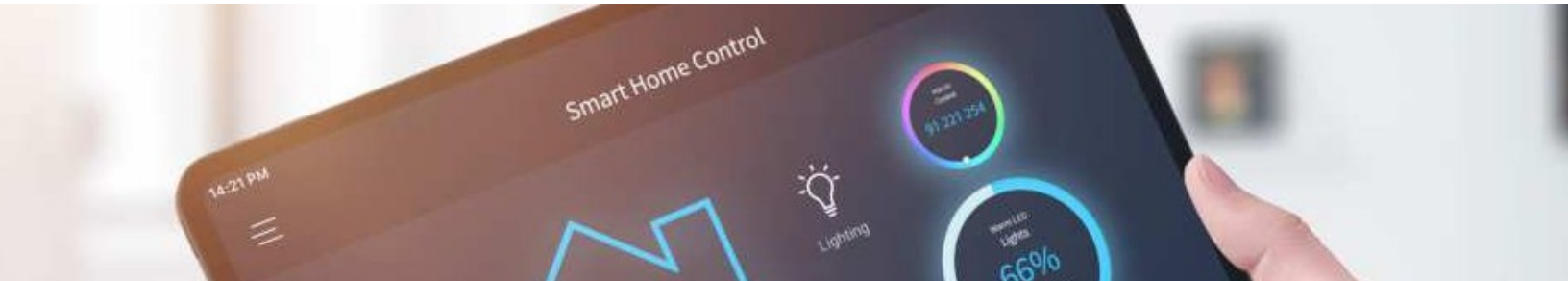
The PSTI Act establishes baseline security requirements, including enabling users to set passwords, providing clear channels for reporting security issues and communicating product updates. While these measures serve as a starting point, they are inadequate given the increasing sophistication of cybersecurity threats and unintended consequences such as IPA. To ensure comprehensive protection of users against complex harms such as IPA, the Act should mandate additional security measures, including:

- **Multi-Factor Authentication (MFA):** Requiring users to verify their identity through multiple authentication factors would significantly reduce the risk of unauthorised access.
- **Enhanced Access Controls and User Notifications:** Implementing stricter access control mechanisms and real-time notifications for suspicious activities would help users respond promptly to potential security breaches.
- **Robust Security Vulnerability Reporting Framework:** Establishing a standardised and transparent reporting framework would ensure that vulnerabilities are identified, disclosed, and mitigated efficiently.

Expanding these compulsory security features would strengthen the Act's ability to address evolving cybersecurity challenges and enhance consumer protection.

2. Adopt certification for secure enabled devices

The European Union Cyber Resilience Act (Regulation 2024/2847) provides for a certification regime for Internet-enabled devices that is complied with as a condition for market access, with a failure to abide by cybersecurity standards resulting in penalties ranging from fines to market approval withdrawal. Adopting a similar cyber-certification regime in the UK would allow for increased levels of trust and confidence amongst British consumers, as well as ensuring the whole of life-cycle security principles that are provided for in the Cyber Resilience Act.



Sectoral co-regulation case studies

3. *Adopt certification for secure enabled devicAdopt Agency by Design principles for Smart Home device technologieses*

Incorporate Agency by Design principles in the security systems for Smart Home devices, both at the hardware and software/app levels. Incorporating the voices of underrepresented groups, particularly the representatives of intimate partner violence support groups and charities focused on Violence Against Women and Girls, as these user groups may have distinct safety and privacy concerns not fully addressed by current security settings on Smart Home devices.

Provide transparency regarding security settings and who can access a Smart device in the home, as well as the tools available to modify privacy settings, helping to empower and provide agency over the technologies employed. Allowing for granular control on the basis of user-defined settings will allow users to determine who is able to access systems, and under what conditions. App providers should also implement a form of emergency response system, where users have the ability to restrict access to or disable devices where necessary, returning systems such as heating and lighting to manual analogue control.

Collective reporting and redress mechanisms made available through in-app systems and forums can ensure users are able to identify any particular security concerns with the hardware or software, and providers can provide meaningful responses by responding to those privacy concerns, or providing new user-defined settings for privacy control. Finally, security updates and privacy refinements should be provided for the entire product life cycle, with specific EoE policies devised.

Final Takeaways

- Organisations and people do not design technologies with harm in mind, rather, these are much-discussed "unintended consequences" caused by lack of data, access to groups that use the technology and so on.

- Agency by Design represents a set of core principles that can be implemented by industry and overseen by government as a means of empowering individual users as part of a co-regulatory system.

- CDR/RI and the tools cited in this section are meant to offer insights into how you could unpack, understand and act when faced with the scenarios examined in the Agency portfolio of projects. The frameworks and suggested tools are not meant as the "only" options, nor should be viewed as a panacea to ensure that in abiding by this guidance, online harms and unintended consequences will not occur.

- Rather, the frameworks assist in raising awareness and understanding whilst must be aligned to the acknowledged methods of design, risk and project management that are adopted in your organisation across the design, development, testing, deployment and continuing digital responsibility (e.g., witnessed in global legislation and regulation).

AGENCY

Assuring Citizen Agency in a World with Complex Online Harms

AGENCY is funded by UKRI/EPSRC in the programme: Protecting Citizens Online and is aligned with National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online (REPHRAIN)

AUTHORS

Dr Becca Owens (becca.owens@ncl.ac.uk)
Professor Benjamin Farrand (ben.farrand@ncl.ac.uk)
Professor Karen Elliott (k.elliott@bham.ac.uk)

AGENCY RESEARCH