MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

COVERSHEET

Minister	Hon Nicola Willis	**Portfolio**	Economic Development
	Hon Cameron Brewer		Commerce and Consumer Affairs
Title of Cabinet paper	Amendments to the Fair Trading Act 1986	**Date to be published**	23 April 2026

List of documents that have been proactively released		
Date	**Title**	**Author**
10 September 2025	Regulatory Impact Statement: Updating the penalties regime in the Fair Trading Act 1986	MBIE
22 May 2025	Regulatory Impact Statement on the safe harbour provision to support online service providers to disrupt online scams	MBIE

Information redacted **YES**

Any information redacted in this document is redacted in accordance with MBIE's policy on Proactive Release and is labelled with the reason for redaction. This may include information that would be redacted if this information was requested under Official Information Act 1982. Where this is the case, the reasons for withholding information are listed below. Where information has been withheld, no public interest has been identified that would outweigh the reasons for withholding it.

Some information has been withheld for the reasons of Confidentiality, Confidential advice to Government and Privacy of natural persons.

Regulatory Impact Statement on the safe harbour provision to support online service providers to disrupt online scams

Decision sought	Cabinet decision on introducing legislative limitations on civil liability (a legal "safe harbour provision") for online service providers for disrupting suspected online scams, provided certain conditions are met.
Agency responsible	Ministry of Business, Innovation and Employment
Proposing Minister	Minister of Commerce and Consumer Affairs
Date finalised	22 May 2025

The Minister is proposing to introduce a limitation on civil liability (a legal "safe harbour provision") for online service providers when they take action to disrupt suspected online scams. The safe harbour provision would apply only if specified conditions are met. These may include, for example, a good faith requirement, that the online service provider had reasonable grounds for believing the content is scam content, and a requirement for the online service provider to promptly reverse the action in the event of an error.

The Minister is also proposing to scope complementary, non-legislative designated expert entities (referred to as a "trusted flagger") to identify suspected scam activity and support proactive scam intervention by issuing disruption recommendations to online service providers. The preferred option therefore includes both a regulatory measure (a safe harbour provision) and a non-regulatory measure (scoping trusted flaggers). This regulatory impact analysis is primarily focussed on assessing the introduction of the safe harbour. Further detail regarding the trusted flaggers is still being worked through but it will not require legislative change, regulatory oversight or additional funding. If the implementation of trusted flaggers is delayed or not progressed for any reason this will not impact the introduction of the safe harbour provision.

These two proposals together are designed to reduce perceived legal risk and enable more confident, timely action by online service providers to prevent scams, without imposing new duties or regulatory burdens on online service providers. It is a targeted and proportionate approach that supports solutions led by the online service provider industry, while maintaining appropriate safeguards for consumers and businesses.

Outlined below is a summary of the type of entities that are intended to be captured by the term "online service provider":

Entity Type	Role	Actions They Can Take

Domain Name Registrar / Host (e.g. Domainz)	Manages registration of website domain names (e.g. www.scamwebsite.co.nz).	Suspend, cancel, or redirect a domain name associated with scam activity.
Website Hosting Provider (e.g. Bluehost)	Hosts the website content and underlying files on a server.	Take down or disable access to scam-related content or entire websites.
Telecommunications Provider / Internet Service Provider (ISP) (e.g. Spark and OneNZ)	Provides internet access and routing for users.	Block access to scam websites or domains at the network level (e.g. Domain Name System (DNS) or Internet Protocol (IP) blocking).
Digital Platform Provider (e.g. Google and Meta)	Distributes or links to content via search, social media, or advertising.	Remove or demote scam-related posts, advertisements, pages, or user accounts.

Summary: Problem definition and options

What is the policy problem?

Online service providers are key players in detecting and disrupting online scams. However, they face potential legal risk when taking proactive steps to detect and disrupt scams. We have used the term "disruption" in this analysis to refer to actions including blocking, removing, sinkholing[1], or otherwise restricting access to suspected scam content. Online service providers have told us that the potential exposure to liability under contract or tort law if legitimate activity is inadvertently affected is stopping them from confidently taking more proactive measures to disrupt scams.

This potential liability risk has created a risk-averse operating environment, where online service providers may delay or avoid taking timely action – even when they have strong indicators of scam activity. As a result, known scams could remain active for longer than necessary, increasing the risk of consumer harm and eroding trust in digital systems.

What is the policy objective?

The proposed change aims to reduce concern among online service providers that they could be legally liable if they make errors, and support their confident, timely disruption of suspected online scam activity. The objective is to create an enabling environment for voluntary, industry-led responses, where providers can act quickly and proportionately to disrupt scams without fear of liability. Success will be measured by:

- increased proactive scam disruption (e.g. website sink holing)
- feedback from online service providers of increased legal clarity and confidence to disrupt scams, and
- reduction in scam-related harm to consumers over time.

What policy options have been considered, including any alternatives to regulation?

1. **Status quo (no regulatory change)**
   Online service providers would continue to bear the risk of civil liability if they disrupt suspected scam content in error. Given the risk averse nature of the online service provider industry, we anticipate online service providers to continue to take a cautious approach and not disrupt suspected scams if this could expose them to legal risk.

2. **Establish a 'trusted flagger' mechanism (an expert entity whose reports are prioritised) for scam identification**
   A government-supported operational entity would be established to help identify scam activity and provide recommendations to online service providers on scam content. This would improve confidence and willingness to act but would not remove

[1] Sinkholing is a cybersecurity technique used to quietly take control of scam or malicious websites. Instead of shutting the website down, internet traffic is redirected to a safe server run by a trusted organisation. This stops the scam from working and can also help gather information about who is being targeted.

liability or create a duty for online service providers to act on the entity's recommendations. This is a non-regulatory option.

3. **Legislative limitation on liability (a "safe harbour provision")**
   Government would introduce a legislative limitation on civil liability for online service providers taking reasonable steps to disrupt scams by establishing a new legislative defence. The provision would be subject to conditions to protect the interests of legitimate businesses using online services and prevent misuse.

4. **A positive duty to act**
   Government would introduce a legislative requirement for online service providers to take reasonable, proportionate steps to disrupt scams. This would increase accountability by imposing enforceable obligations. It would require regulatory oversight.

What consultation has been undertaken?

We have been engaging with industry since early 2024 on strategic and coordination issues related to addressing scams. Since late 2024, our engagement has focused on potential regulatory hurdles and solutions, including the issue of liability for online service providers who disrupt scams. Our targeted consultation with industry and consumer groups includes:

a) Government Agencies and Private Sector Scams Data workshop in November 2024
b) A workshop in December 2024 to understand problems relating to disrupting scam activity.
c) A workshop in March 2025 to test a wider set of coordination and regulatory proposals to address scams.

A summary of the workshops can be found at **Annex 1**.

Most attendees supported taking a targeted regulatory approach, and for government to provide a legislative safe harbour provision to address liability risks for disrupting suspected scam websites.

We have not undertaken any public consultation due to the Minister's expressed concern for moving ahead with this proposal quickly. Should regulatory changes be progressed, stakeholders will have an opportunity to provide feedback during the select committee stage.

There are currently no plans to publish an exposure draft.

Is the preferred option in the Cabinet paper the same as preferred option in the RIS?
Yes

Summary of Minister's preferred option: option 2 and option 3: clarifying expert entities to serve as 'trusted flaggers' and establishing a legislative limitation on civil liability

Costs (Core information)

The Minister's preferred approach is to combine option 2 and option 3, clarifying expert entities to serve as "trusted flaggers" and establishing a legislative limitation on civil liability (a safe harbour provision). Only the legislative limitations on civil liability have regulatory implications.

The approach would impose minor compliance costs on online service providers. These costs would relate primarily to one-off adjustments to internal processes for assessing and meeting safe harbour provision conditions. Courts may occasionally be involved in determining whether conditions for the safe harbour provision are met, but this is expected to be infrequent.

Modest resourcing may be required to establish or support existing entities to perform the trusted flagger role, but this is expected to be manageable within baselines.
Legitimate businesses that use online services may experience some disruption if online service providers mistakenly block their content. We consider the probability and impact of this risk to below, given that there will be clear procedural safeguards to minimise the risk, and a process for online service providers to reverse the disruption should they have made an error.

There are no significant direct costs for government, as implementation relies on existing legal and regulatory mechanisms. Minor resources may be required for awareness-raising, but no new operational systems or public funding streams are proposed.

Benefits (Core information)

The approach would reduce online service providers' concerns that they would be legally liable if they mistakenly disrupt a legitimate business, which online service providers report has prevented them from confidently disrupting suspected scam websites. The trusted flagger model would enhance online service providers' decision-making credibility and support faster, coordinated scam content removal.

Consumers would benefit from earlier intervention in scam incidents, which would reduce the harm that scams cause to consumers and build consumers' trust in digital systems. These benefits are non-monetised but potentially high and supported by international precedent and stakeholder submissions.

The proposals could bring confidence and eventual financial reward to legitimate businesses that operate online as trust in the system increases.

Balance of benefits and costs (Core information)

We assess that the benefits of the preferred approach outweigh the costs. Implementation is low-cost and builds on existing civil remedies. The approach would reduce barriers that have prevented online service providers from disrupting scams. The approach also supports

broader efforts to strengthen trust and safety in digital services. The approach is aligned with international frameworks, including Australia's, and is expected to improve system responsiveness, which will reduce consumer losses to scams over time.

Implementation

The option to limit potential civil liabilities would be implemented through a legislative amendment. This will function as a defence to civil claims in tort or contract where an online service provider has acted in accordance with defined safe harbour provision conditions. No dedicated funding is required. This option is intended to support earlier and more confident disruption of scams by online service providers, without requiring a comprehensive legislative framework or a new regulator.

The trusted flagger option is a non-legislative mechanism: Expert entities that can identify likely scam content and provide coordinated disruption recommendations to online service providers (a trusted flagger model). This will be implemented operationally via an existing government agency, or the new anti-scam alliance under development. This option would support scam identification and website disruption recommendations, improving consistency and confidence in provider decision-making. The trusted flagger roles may require minor operational resources but can be delivered within existing baselines. The implementation of the legislative safe harbour is not dependent on the trusted flagger and can proceed independently.

Limitations and Constraints on Analysis

We have not conducted full public consultation on this policy issue, nor on the full range of options, due to time constraints and Ministerial direction. Had consultation occurred, it would have provided further evidence on the nature and scale of the problem, stakeholder views on liability risks, and the potential impacts on legitimate businesses whose online activity may be mistakenly disrupted.

To partially mitigate this, we reviewed select committee submissions on Australia's Scam Prevention Framework Bill, which also included a legislative safe harbour provision to address industry concerns about disrupting scams. This enabled us to understand views of businesses that may be affected by such a provision, and of law firms.

The overwhelming majority of submissions on Australia's Scam Prevention Framework Bill from the private sector, consumer representative groups, and a small business representative group supported the inclusion of a safe harbour provision. One of the concerns raised by submissions was the potential negative effects on legitimate small businesses who could have their website mistakenly disrupted under the guise it was a scam, and potentially suffer loss of sales and impacts to brand reputation[2].

This risk could be mitigated through:

1. the establishment of a 'trusted flagger' regime to support online service providers to make decisions on disrupting suspected scam content, and

[2] Submissions – Parliament of Australia. Submission 26: Australian Small Business and Family Enterprise Ombudsman at page 4.

2. including a condition in the safe harbour that requires online service providers to promptly restore services where they have been mistakenly disrupted.

A legislative safe harbour provision for disrupting digital content has been implemented in some other contexts, for example the Harmful Digital Communications Act 2015. However, a safe harbour provision for disrupting scams in New Zealand would be novel. Australia's Scams Prevention Framework only came into force in February 2025, so it is too early to assess how it is working in practice or what its real-world impacts may be. We are therefore unable to understand the real-world impacts of such a provision.

I have read the Regulatory Impact Statement and I am satisfied that, given the available evidence, it represents a reasonable view of the likely costs, benefits and impact of the preferred option.

Privacy of natural persons

Responsible Manager(s) signature:
Glen Hildreth
Manager Consumer Policy
22 May 2025

Quality Assurance Statement

Reviewing Agency: Ministry of Business, Innovation and Employment	**QA rating:** Meets

Panel Comment:

The Panel consider that the information and impact analysis summarised in the RIS meets the Quality Assurance criteria. The RIS is clear, concise, complete and convincing. While a full public consultation has not been carried out, there has been a targeted consultation with key stakeholders, and the analysis incorporates results from consultation on a similar Australian proposal.

Section 1: Diagnosing the policy problem

1. Scam disruption is a fast-evolving policy issue, shaped by rapid technological change, fragmented regulatory responsibility, and increasing consumer vulnerability. Scam-related harm is growing in New Zealand. Scam losses are estimated to be between just under $200 million and $2 billion annually.[3]

2. Scams are increasingly complex and fast-moving, exploiting digital infrastructure like telecommunications networks, social media, messaging apps, and online banking platforms. These scams often rely on multiple touchpoints across different service providers, making coordinated, timely disruption essential.

3. Scam disruption in New Zealand is currently industry-led, without an overarching duty-based legal framework. This light-touch approach has delivered positives to date in the banking sector – where industry has worked closely with government and introduced voluntary measures to provide better protections for consumers. Online service providers, including telecommunication companies, domain name hosts and digital platforms, also have a key role in scam disruption as they can restrict access to or remove scam content.

4. However, online service providers must navigate potential legal risk under contract and tort law if they take steps to disrupt suspected scam content that may later prove legitimate. This creates significant hesitancy to act pre-emptively, particularly in fast-moving scenarios.

5. This concern around liability has been raised by online service providers in Australia during the development of Australia's Scams Prevention Framework. This was addressed by including a legislative defence, called a "safe harbour" into the Bill that would protect the relevant entities from civil liability for any errors made for disrupting suspected scam intelligence. This was designed to remove any barriers that might prevent industry from acting swiftly to disrupt suspected scams, and therefore subjecting more consumers to the same scam. The Australian safe harbour provision has several conditions that must be met to ensure that the interests of legitimate businesses are protected and that the provision is not misused.

6. Under the status quo, Government expects industry to play an increasing role in scam disruption. The New Zealand Anti-Scam Alliance (the Alliance), launched by Hon Simpson on 10 July 2025, is a cross-sector initiative to improve New Zealand's ability to prevent, detect and disrupt scams. It brings together government agencies, banks, telecommunications providers, digital platforms, and consumer groups to coordinate efforts and enhance consumer protection. It aims to improve how scam-related intelligence is shared and acted upon across sectors, update industry codes of practice, raise public awareness, and support businesses and consumers with tools to prevent and respond to scams. MBIE is the coordinating agency for the Alliance.

[3] Data sources: the nearly $200 million figure is based on reported data from Payments New Zealand, released by MBIE as part of the annual Fraud Awareness week campaign in 2024. The $2 billion figure is from the 2024 *State of Scams in New Zealand* report by Netsafe.

7. However, in the absence of legal protection, online service providers may limit their intervention to low-risk, clear-cut scams, or delay action while seeking legal assurance. This is likely to constrain the pace and consistency of disruption responses which may reduce the effectiveness of voluntary action over time.

8. We have received feedback from the National Cyber Security Centre (NCSC) and the New Zealand Telecommunications Forum that cyber incident disruption tools (that may include tools to disrupt scams) can significantly reduce consumer harm. The NCSC's 2024 Cyber Threat Report estimated $38.8 million worth of harm was prevented by these tools. Confidential advice to Government

What is the policy problem or opportunity?

9. Scam disruption is an increasingly urgent policy issue, but providers' ability to act quickly and confidently is constrained by potential liability when disrupting suspected scams. The current legal framework does not create an enabling environment for providers acting in good faith to protect consumers from scams. As a result, voluntary action is often delayed or inconsistent, increasing the risk of harm to consumers from scams.

10. It is difficult to assess the scale of this problem as we cannot determine the quantity of scams that would be disrupted if the safe harbour did exist. Instead, we are relying on anecdotal evidence from online service providers that there are more proactive measures they would be able to take if they had more certainty that they would not be exposed to legal liability for good faith errors. We are also relying on the submissions from industry in Australia that commented on the need for a safe harbour in the development of the Scams Prevention Framework.

What objectives are sought in relation to the policy problem?

11. The primary objective is to provide greater clarity and confidence for providers to take timely, reasonable and proportionate action to disrupt suspected scams without fear of prosecution.

What consultation has been undertaken?

12. Industry engagement has been ongoing since early 2024 on strategic and coordination issues. Since late 2024, engagement has begun with industry on potential regulatory hurdles and solutions. This engagement has included a range of issues related to scam prevention, including this liability concern raised by industry.

13. Targeted consultation was undertaken in November through an agency-led workshop, and in December 2024 and March 2025 where the Minister convened government, industry and consumer groups to understand problems relating to stopping scam activity (December) and to test a wider set of coordination and regulatory proposals to

address scams (March). More information on the attendees and topics of discussion is set out in Annex 1.

14. No formal public consultation has been undertaken due to the Minister's expressed concern that consultation would incur delays to implementation. In the absence of formal consultation feedback, we reviewed submissions from the Australian select committee's consideration of the Scams Prevention Framework Bill, which included a similar safe harbour provision.

Section 2: Assessing options to address the policy problem

What criteria will be used to compare options to the status quo?

15. **Effectiveness** – how well does the option support more confident, proactive scam disruption by providers?

16. **Proportionality** – is the option appropriately targeted to the problem, avoiding unnecessary regulatory burden while still achieving the objective?

17. **Practicality** – how feasible is the option to implement, considering the alignment with existing industry structures, cost and time to implement/operationalise?

18. These criteria will be equally weighted.

What scope will options be considered within?

19. This RIS considers options that address the specific problem of perceived legal risk faced by online service providers when taking voluntary action to disrupt scam websites. The scope is deliberately narrow, reflecting Ministerial direction to focus on a safe harbour provision rather than a broader regulatory framework.

20. The analysis includes both legislative and non-legislative options that could enable or encourage proactive scam disruption, provided they do not significantly expand government enforcement powers. Options that would shift to a prescriptive or centralised model, such as a mandatory takedown regime, were not considered. These have been ruled out due to concerns about proportionality, implementation complexity, and misalignment with New Zealand's existing regulatory approach in the digital and communication sector which is more supportive of risk-based interventions and industry-led responses. The Government's current focus on reducing compliance burden and enabling private sector innovation also supports a more targeted approach.

21. To shape the options, we considered relevant international models (particularly Australia's Scams Prevention Framework and trusted flagger initiatives in Europe), and New Zealand's Harmful Digital Communications Act 2015 (**HDCA**). Full replication of international models was deemed inappropriate given differences in legal systems, institution responsibilities and levels of government intervention. The preferred approach is targeted, enabling, and industry-led, which is consistent with feedback from stakeholders and the Government's emphasis on light-touch, responsive regulation.

22. As a result, the options assessed in this impact analysis are bounded by:
    - Ministerial preference for a safe harbour provision
    - Industry and agency feedback on practicality and proportionality
    - The need to support voluntary action while avoiding significant compliance or administrative burden.

What options are being considered?

Option One – Status quo

23. Under this option, no amendments would be made to existing legislation or operational settings. Online service providers would continue to operate within the current legal framework, which lacks explicit provisions addressing potential liability for mistaken actions intended to proactively disrupt scams. This would continue the current process of online service providers only proactively disrupting scams within their current risk appetite. This may result in ongoing hesitation to act and potentially deterring proactive efforts to protect consumers.

Option Two – Trusted flagger model

24. This option would establish a framework where designated experts (e.g. a government agency or independent entity) identifies, collects and collates scam reports and recommend to online service providers that they disrupt suspected scam activity. Online service providers typically grant trusted flaggers of harmful online content priority status, which means the online service providers review their recommendations without undue delay.

25. While this option does not remove legal liability from providers when taking action to disrupt scams, it can support more confident and timely decision-making by giving online service providers a credible basis for action. We have heard anecdotally from engagement that having an approved entity flagging online content is useful for disrupting scam content.

26. The trusted flagger mechanism could be implemented through a voluntary arrangement between entities - as is already the case with Netsafe and the New Zealand Police in the online harm space.

Option Three – Legislative safe harbour provision

27. This option would amend the Fair Trading Act 1986 to establish a statutory defence (a "safe harbour") that online service providers can rely on if they are subject to a civil liability claim because they have mistakenly disrupted online activity by a legitimate business when intending to disrupt a scam; and the business suffered harm as a result. This type of legal defence is incorporated into the Australian Scams Prevention Framework Act. The Australian legislation has a requirement for entities to meet certain conditions in order to rely on the safe harbour. A similar limit on civil liability is also included in New Zealand's Harmful Digital Communications Act 2015 (HDCA). In that context, the defence applies to online content hosts who act on approved takedown requests. The defence offers legal protection where online content hosts act in good faith and follow the statutory process.

28. To ensure that access to justice is not unreasonably interfered with, the safe harbour provision under this option would apply only when specific conditions are met, such as acting in good faith and within defined parameters. This approach aims to provide legal clarity to encourage proactive, industry-led scam prevention measures by reducing the risk of liability in the event of a mistake.

Option Four – Positive duty to act

29. This option proposes introducing a positive statutory duty for online service providers to take reasonable and proportionate action to prevent scams circulating on their networks. The duty would require providers to establish systems and processes to identify, assess and where appropriate, disrupt suspected scam activity. This model would be broadly based on Australia's new Scams Prevention Framework, which is the first of its kind worldwide and yet to be assessed for its effectiveness but offers a potential precedent for this type of regulatory approach.

30. This model aims to clarify what action is expected of industry and provide protection where those actions are taken appropriately. However, it would likely require a regulator to monitor and enforce the duty.

Stakeholder feedback on the options

31. As outlined above, MBIE undertook targeted consultation on this proposal. Feedback indicated support for the introduction of a legislative safe harbour provision to support industry to take proactive steps to disrupt scam activity. There was no negative feedback from any stakeholders present. Later feedback from telecommunications sector participants indicated preference for a positive duty to accompany the safe harbour provision. This was to align better with overseas approaches such as those taken in Australia. The telecommunications sector considers that a positive duty would be more persuasive in encouraging industry initiatives. Due to time constraints for both formulating policy and implementation, we consider that a positive duty should not be pursued.

32. The telecommunications industry body and one of its members also supported a trusted flagger-type model as an operational solution, to sit alongside the legislative safe harbour provision.

33. Feedback was positive from businesses and groups, including law firms, that represented their interests during the Australian select committee for the Scams Prevention Framework Bill. Most submitters thought that the introduction of such a provision is a necessary and useful tool to encourage industry-led action. Many submitters on Australia's Bill either have separate New Zealand legal entities of the same name or operate in both jurisdictions (eg banking sector participants, digital platforms). The combination of direct engagement and reviewing Australia's consultation has illustrated support for introducing a legislative safe harbour provision in New Zealand. Most of these submitters highlighted the 'window' of time that the safe harbour provision should apply – for example, 28 days. This would be considered in drafting.

How do the options compare to the status quo?

	Option One – Status quo	Option Two – Trusted flagger model	Option Three - Safe harbour provision	Option Four– Positive duty to act
Effectiveness	0 Maintains perceived legal risk. Providers likely to remain hesitant to act proactively.	+ Improves coordination but legal risk for providers persists. Confidence may increase incrementally.	++ Directly addresses perceived legal risk.	0 Creates enforceable obligations. Unclear if this will reduce liability concerns.
Proportionality	0 No new burdens. No change in disruption capability or protection.	+ Flexible and non-regulatory. Accountability remains unclear.	++ Targeted and clear. Enables voluntary action without imposing duties.	+ Could impose burdens inconsistent with light-touch regulation.
Practicality	0 No changes required. Relies on existing frameworks.	- Requires new operational model and likely funding.	- Requires legislation but no operational change.	-- Requires legislation and regulatory enforcement capability.
Overall assessment	0 Does not resolve the problem. Maintains status quo of risk aversion.	+ Partial improvement through coordination, but core issue remains.	+++ Enables confident, timely disruption.	- Could improve outcomes but higher cost and complexity.

Key for qualitative judgements:

++ much better than the status quo

+ better than the status quo

0 about the same as the status quo

- worse than the status quo

- - much worse than the status quo

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

34. The preferred approach combines a regulatory tool (safe harbour) and a non-regulatory support mechanism (trusted flagger) because together they offer complementary benefits: reducing legal risk to enable faster action, while also supporting more accurate and coordinated disruption through expert input, without constraining providers' discretion.

35. This combination of these two options is the preferred approach because it most directly addresses the core policy problem: perceived legal risk, which is currently a barrier to timely and proactive scam disruption by online service providers. This uncertainty creates hesitation to disrupt scams, particularly where actions may involve content removal that could later be legally challenged. A trusted flagger regime could support this through greater coordination and as an optional source to support verification of the suspected scam website.

36. The proposed conditional legal protection from civil liability would be available where specific criteria are met. The criteria could include (based on the Australian legislative model), where the action to disrupt the suspected scam website is:
    - taken in good faith
    - the online service provider has a reasonable belief that the content is a scam
    - taken within a certain time and
    - reversed promptly if found to be in error.

37. Combining Options 2 and 3 enables confident, timely action by industry, without imposing new positive duties or broad regulatory obligations on the public. This approach is targeted and risk-based, reflecting the Government's commitment to a light-touch, enabling regulatory approach that supports industry-led solutions and protects consumers.

38. The primary beneficiaries of the proposal are online service providers, who gain greater legal protection in the event they make an error. Consumers also benefit indirectly through improved scam disruption and reduced potential financial harm. No significant adverse distributional impacts are expected, as the safe harbour provision does not impose new obligations on other sectors or groups. There is some risk that legitimate businesses or individuals would be affected if content is mistakenly disrupted. This risk is mitigated by conditions attached to the safe harbour provision. This helps balance the interests of scam prevention with the protection of lawful commercial activity and free expression.

39. We do not anticipate significant impact on business competition in the online services sector, as the safe harbour provision applies broadly to all providers engaging in scam disruption and does not create market advantages for any one online service provider.

40. New Zealand's approach to intermediary liability under the Harmful Digital Communications Act 2015 provides a useful precedent. That framework introduced a legal safe harbour provision for online content hosts who follow a defined complaints-handling process, enabling action to limit harmful content while avoiding undue liability.

41. A similar approach to scam disruption would reduce legal risk for online service providers acting in good faith, while maintaining appropriate safeguards for consumers and businesses.

42. The safe harbour provision is:
    - Highly effective because it removes a key barrier to voluntary disruption efforts
    - Proportionate because the protection is conditional and the online service provider must be able to reverse the disruption if an error is found, and
    - Practical because it does not require a new regulatory body, powers or processes, and would operate alongside existing regulatory structures.

43. The trusted flagger is an optional support mechanism and enables online service providers to both act independently upon their own volition (for example, a telecommunications provider receives internal reports and disrupts a potential scam webpage). Or alternatively, for trusted flaggers to raise an issue with the online service provider for them to act on.

44. This option does not constrain provider discretion. Online service providers may act independently where confident the safe harbour provision conditions are met or seek a trusted flagger's input in cases of uncertainty or where coordination is required.

45. This assessment assumes online service providers will use the safe harbour provision as intended — i.e. in response to scam threats rather than for unrelated content moderation purposes. Benefits are primarily non-monetised but expected to be high, including improved scam disruption rates, reduced financial harm to consumers, and increased industry confidence to act. The benefit-cost ratio is expected to improve over time as scam tactics evolve and reliance on rapid disruption responses grows, increasing the value of a responsive and enabling regulatory environment.

Is the Minister's preferred option in the Cabinet paper the same as the agency's preferred option in the RIS?

46. Yes. The Minister supports a safe harbour provision and a trusted flagger model (options 2 and 3).

What are the marginal costs and benefits of the preferred option in the Cabinet paper?

Affected groups	Comment	Impact	Evidence Certainty
Additional costs of the preferred option compared to taking no action			
Online service providers	Minor compliance effort to meet safe harbour provision conditions. One-off adjustment with no ongoing burden.	Low to moderate (non-monetised)	Medium – Based on stakeholder feedback and precedent from the HDCA.
Courts (via litigation)	Occasional role in determining if conditions of immunity were met in legal disputes.	Low (non-monetised)	Low – our analysis is based on expected limited volume of cases and alignment with Australian and HDCA precedent.
Legitimate businesses (affected third parties)	Risk of revenue loss or disruption if legitimate services are mistakenly disrupted. The frequency of this occurring is expected to be low. The provider must reverse the disruption promptly if an error is found.	Low to moderate (non-monetised)	Low – impact is contingent on error. The conditions will mitigate this risk.
Trusted flagger agencies	Some resourcing required to support the trusted flagger function. Cost depends on volume, agency model, and use of existing capacity. Depending on agency/ organisation, it could support existing scams remit.	Low (non-monetised)	Medium – Depends on design choices and operating model.
Total monetised costs	N/A – monetised costings unavailable.		
Non-monetised costs	Low – potential lost revenue for affected legitimate businesses and compliance cost for online service providers to self-assess actions against safe harbour provision conditions.		

	Low to moderate – compliance effort, possible court use, limited resourcing of trusted flagger role.		
Additional benefits of the preferred option compared to taking no action			
Online service providers	Increased legal protection enables more proactive scam disruption without fear of liability. Trusted flagger gives online service providers an additional data point to inform online disruption activity.	Potentially high (non-monetised)	Medium – Based on stakeholder engagement and safe harbour provision precedent.
Consumers	Faster scam disruption reduces financial harm and builds trust in digital systems. Real-world experience from the National Cyber Security Centre's cyber disruption tools shows that blocking scam domains leads to faster disruption with minimal false positives. This builds public trust in digital infrastructure.	High (non-monetised); broad public impact	Medium to high – Supported by scam loss data and consumer surveys.
Courts (via litigation)	Nil – some benefit in government providing the defence to support the court's decision-making, but marginal benefit only.	Nil (non-monetarised)	Low – as a similar safe harbour provision in Australia has not yet been implemented.
Legitimate businesses with online activity	Clearer position that government supports timely action by regulated groups to disrupt suspected scam websites. More action by online service providers to disrupt scam websites. Could also improve trust and confidence of customers transacting with the business online.	Low-medium (non-monetised).	Medium – Based on submissions from affected businesses and Australia's Scam Prevention Framework Bill.
Total monetised benefits	Not quantified but likely to reduce financial losses for consumers over time		
Non-monetised benefits	High – more confidence for online service providers to disrupt online scams and increased public trust in digital services.		

Section 3: Delivering an option

How will the proposal be implemented?

47. The proposed safe harbour provision would be implemented through amendments to existing legislation, such as the Fair Trading Act 1986. A trusted flagger mechanism could be developed and operationalised through the Alliance. It would not require new statutory powers or enforcement capability. Two key pillars of the Alliance are 1. Collaboration: focused on increasing data sharing and intelligence and coordination and 2. Disruption: focused on tackling scammers at scale by uniting national efforts across sectors. We consider the Alliance is well placed to develop and operationalise trusted flagger(s).

48. While the safe harbour is the primary regulatory intervention, we have chosen to pair it with a non-legislative trusted flagger mechanism to provide operational support. This combination reflects feedback from stakeholders and industry that legal protection alone may not be sufficient to support timely and confident disruption decisions, and the fact that a broader duty-based framework is out of scope for this work. The trusted flagger mechanism is not a condition for using the safe harbour but offers a flexible tool to aid decision-making – particularly where scam intelligence is complex or contested.

49. The effectiveness of the safe harbour provision (Option 3) does not rely on the trusted flagger (Option 2) being in place. Therefore, if there are delays or risks associated with implementing the trusted flagger mechanism, the legislative safe harbour can proceed independently.

Responsibility for implementation

50. The liability limitations would provide a conditional statutory defence, meaning regulated groups are responsible for assessing whether their actions meet the legal criteria for protection. Therefore, there is no need for a new enforcement agency, or an extension of an existing agency's enforcement or regulatory powers. If a dispute arises, courts will determine whether those conditions are met, and the provider would have the onus of proving the conditions under the safe harbour provision. MBIE may play a limited role in issuing non-binding guidance to support interpretation and sector readiness, including through its coordinating role in the Alliance to develop and operationalise trusted flagger(s).

51. Providers may act independently under the safe harbour provision without input from the trusted flagger. The flagger is intended to improve coordination and decision confidence, to provide further intelligence about whether a website or online content is a scam.

Timing and commencement

52. Confidential advice to Government The safe harbour provision will not have retrospective effect, meaning it cannot be relied upon by industry for any action taken before amending legislation comes into force.

53. As above, we propose to work with Alliance members to develop and operationalise the trusted flagger roles. We anticipate that the trusted flagger roles could be operationalised by early 2026, to coincide with the implementation of the Alliance work.

Funding

54. No additional operational funding is required to implement the safe harbour provision. The proposal does not involve proactive monitoring or administration by regulators. Any cost to government would be limited to policy development and legislative drafting, which is minor and already resourced.

55. An existing entity or the operational scam alliance under development could pick up this role. Should existing entities take on trusted flagger roles, this would need to be funded within baselines, though it is possible that some work would need to be stopped to pick up this role.

Supporting awareness and compliance

56. Clear guidance materials and a targeted communications effort will be needed to ensure affected industry bodies and consumer groups understand the purpose, scope and limits of the safe harbour provision.

Risks and mitigation

57. The key implementation risk is that providers may misinterpret or overextend the safe harbour provision, leading to unnecessary service disruption or consumer complaints. This risk is mitigated by clearly defined statutory conditions (e.g. good faith, proportionality, reversibility), sector consultation, and supporting guidance.

58. There is also a risk that providers may over-rely on trusted flaggers and delay action while awaiting confirmation. This will be mitigated by guidance clarifying that the safe harbour provision permits proactive, independent action, and that the flagger is a supporting mechanism.

59. Existing systems already enable rapid reversals of erroneous takedowns — usually within hours — which helps mitigate the risk of wrongful disruption. The safe harbour would sit alongside these procedural safeguards.

Further work

60. Following passage of legislative amendments, MBIE will support implementation through stakeholder engagement, provision of guidance, and monitoring any early application issues. Further changes to regulatory settings on scams are not anticipated at this stage.

How will the proposal be monitored, evaluated, and reviewed?

61. The proposed safe harbour provision would be a conditional statutory defence from civil liability. It would not require ongoing regulatory oversight or administration. Instead, the safe harbour provision would apply only where providers meet clearly defined conditions (such as acting in good faith, proportionality, and within a defined timeframe).

62. There is no proactive enforcement role for regulators. If a dispute arises, it would be for the courts to assess—after the fact—whether the conditions of the safe harbour provision were met. In such cases, the onus would rest on the provider to demonstrate compliance with the statutory criteria. This approach is consistent with the safe harbour provision under the Harmful Digital Communications Act 2015.

63. Impact will be monitored in the following ways:
    - Industry feedback on whether the safe harbour provision is increasing confidence to take proactive scam disruption measures (qualitative)
    - Observable shifts in industry responsiveness - e.g. through voluntary reporting or illustrative case studies (qualitative)
    - The frequency and nature of disputes or legal challenges invoking the safe harbour provision (qualitative)
    - Stakeholder feedback on the use, clarity, or limitations of the trusted flagger model (qualitative)
    - Number of scam reports trusted flaggers provide to industry (quantitative).

64. MBIE will monitor the implementation of both the safe harbour provision and the trusted flagger model, supported by feedback from providers and industry associations. Early insights will be used to assess whether additional operational support, clarification, or adjustment is needed — including any impacts on the efficiency of scam response. This will align with broader stewardship responsibilities under the Fair Trading Act.

Annex 1: Summary of Industry engagement

Three key engagements have been taken with government, industry and non-government organisations since late 2024. The workshops discussed key ongoing scam issues that industry are experiencing including:

- the lack of data sharing that is preventing the swift disruption of scam material
- the need for a central point of contact across government to ensure efficient communication of anti-scam activities and initiatives
- better industry-led commitments and voluntary codes, particularly from online service providers to ensure all participants are contributing equally to anti-scam activities, and
- barriers both perceived and real that are preventing industry from disrupting scams.

Some participants also directly highlighted the need for a comprehensive legislative intervention, similar to Australia's Scam Prevention Framework Act 2025.

The workshops include:

1. *Government Agencies and Private Sector Scams Data workshop in November 2024*

   Workshop was hosted by the Financial Markets Authority (FMA). Attendees were:

   - Banking and payments sector: BNZ, GetVerified Limited, Payments NZ.
   - Telecommunications sector: 2Degrees, Spark NZ, Telecommunications Forum.
   - Digital sector: Google New Zealand, Domain Name Commission, Meta, Microsoft.
   - Consumer representatives: Netsafe.
   - Government: FMA, MBIE, Commerce Commission, Department of Internal Affairs (DIA), Inland Revenue (IRD), Police, National Cyber Security Centre, Serious Fraud Office.

2. *A workshop in December 2024 to understand problems relating to stopping scam activity*.

   Attendees were:

   - Banking sector: Kiwibank, ASB, ANZ, Westpac, New Zealand Banking Association.
   - Telecommunications sector: Spark NZ, One New Zealand, 2 Degrees, Telecommunications Forum.
     Digital sector: Meta, Apple, Google New Zealand.
   - Consumer representatives: Consumer New Zealand, Netsafe, Banking Ombudsman.
   - Government: MBIE, Police, DIA, IRD, National Cyber Security Centre, FMA, Commerce Commission.

3. *A workshop in March 2025 to test a wider set of coordination and regulatory proposals to address scams*

   Attendees were:

- Banking sector: Kiwibank, ASB, ANZ, Westpac, New Zealand Banking Association.
- Digital sector: Meta, Apple, Google New Zealand.
- Telecommunications sector: Spark NZ, One New Zealand, 2 Degrees, Telecommunications Forum.
- Consumer representatives: Consumer New Zealand, Netsafe, Banking Ombudsman.
- Government: MBIE, Police, DIA, IRD, National Cyber Security Centre, FMA, Commerce Commission.