Impact Assessment

Title:	Technology Testing Defence
Type of measure:	Primary Legislation
Department or agency:	Home Office
IA number:	HO IA 1022
Type of Impact Assessment	Enactment
RPC reference number:	N/A
Contact for enquiries:	CrimeandPolicingBillTeam@homeoffice.gov.uk
Date:	25/02/2026

1. Summary of proposal

1. Technology, particularly Artificial intelligence (AI), is increasingly being used to generate and distribute illegal and harmful child sexual abuse material (CSAM), extreme pornography (EP), and non-consensual intimate image (NCII) abuse. The rapid advancement and widespread availability of AI technologies—without adequate guardrails—has accelerated the volume of AI-generated abuse circulating online. Testing AI models for their capability to produce such content is essential to ensure comprehensive safety standards and effective guardrails. Under the current legal framework, companies and public safety bodies face significant liability risks if they attempt to test for illegal content. This prevents developers and safety researchers from identifying failure modes, validating safeguards, and addressing risks at source. Industry have approached government for support with the legal barriers around capability testing of AI models.
2. In the Crime and Policing Bill, the government is introducing a power to create a statutory defence which will establish a clear legal framework to allow authorised persons to test AI models, to determine their capability of creating CSAM, NCII and EP (hereafter also referred to collectively as 'illegal content'). This power does not itself

introduce the defence but allows one to be created at a later stage via secondary legislation, subject to Parliamentary oversight. Under future regulations brought forward using this power, only persons authorised by the Secretary of State for the Home Department (Home Secretary) to undertake technology testing activities will be able to rely on the statutory defence, and this authorisation will be subject to stringent conditions. Allowing controlled testing will enable authorised entities to understand models' capabilities, identify weaknesses, design out harmful outputs, and create effective safeguards to prevent criminal misuse.

2. Strategic case for proposed regulation

3. Developers and safety organisations need to test technology, particularly AI models, for harmful outputs to understand risk, identify vulnerabilities, and build safeguards, but under UK law, creating or possessing CSAM, including synthetic images, is a strict liability offence.[1] Currently, if developers were to test for CSAM, they risk criminal prosecution if harmful material is generated during testing. Similar restrictions apply to EP and NCII under section 63 of the Criminal Justice and Immigration Act 2008 and related statutes.[2] Establishing a statutory defence would remove these legal barriers for authorised persons conducting testing operating subject to strict conditions, enabling comprehensive and proactive testing before product release.

4. There are documented cases of AI-generated CSAM, EP and NCII, with reports from law enforcement and child protection agencies highlighting the scale and speed of proliferation.

- **CSAM:** In 2024, the National Centre for Missing and Exploited Children's CyberTipline saw a 1,325 per cent increase in reports involving Generative AI from 4,700 reports in 2023 to 67,000 reports in 2024[3]**.**
- The Internet Watch Foundation (IWF) also published data showing that reports of AI-generated CSAM more than doubled from 199 on 1 January to 31 October 2024 to 426 in the same period in 2025.[4]
- They also reported that the severity of CSAM material intensified over the past year with Category A[5] content (the most severe images) rising from 2,621 to 3,086 items, now accounting for 56 per cent of all illegal material compared to 41 per cent in 2024.

---

1 Protection of Children Act 1978, (section 1) – GOV.UK: https://www.legislation.gov.uk/ukpga/1978/37/section/1; Sexual Offences Act 2003, (sections 62-64) – GOV.UK: https://www.legislation.gov.uk/ukpga/2003/42/part/1/crossheading/preparatory-offences

2 Criminal Justice and Immigration Act 2008 (sections 63 to 64) – GOV.UK: https://www.legislation.gov.uk/ukpga/2008/4/contents; Voyeurism (Offences) Act 2019 (Section 67A) – GOV.UK: https://www.gov.uk/government/publications/implementation-of-the-voyeurism-offences-act-2019, Domestic Abuse Act 2021 (section 69) – GOV.UK: https://www.legislation.gov.uk/ukpga/2021/17/section/69; Online Safety Act 2023 (sections 170 to 174) – GOV.UK: https://www.legislation.gov.uk/ukpga/2023/50/part/8/chapter/2

3 2024 CyberTipline Report - CyberTipline Data: https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata

4 IWF, 2025: IWF backs new rules as AI imagery grows more extreme: https://www.iwf.org.uk/news-media/news/ai-imagery-getting-more-extreme-as-iwf-welcomes-new-rules-allowing-thorough-testing-of-ai-tools/

5 Indecent imagery categories set by the UK sentencing council are as follows:
- Category A: Images depicting penetrative sexual activity; images involving sexual activity with an animal; or sadism
- Category B: Images depicting non-penetrative sexual activity
- Category C: Other indecent images not falling within categories A or B.

- There have been reports of the AI tool Grok being used to create AI-generated sexualised images of women and children[6] – the IWF has confirmed it has identified Category C images of girls aged between 11 and 13 that appear to have been created using the tool.
- **EP:** It has been reported that 98 per cent of deepfake videos online are pornographic and that 99 per cent of those targeted are women or girls[7]. The National Police Chiefs' Council reports that AI enabled Violence Against Women and Girls (VAWG) is likely to be the fastest growing threat to women and girls in the next 12 to 24 months, in particular that image-based offences will become the predominant type of offence experienced and reported by women and girls.[8]
- **NCII:** this abuse is a rapidly escalating issue that primarily effects women and girls. In 2023, 71 per cent of reports received by the Revenge Porn Helpline (RPH) were made by women, and it has been estimated that 3.18 million NCII images are being created, per year[9]. They have also reported cases grew from 521 in 2015 to 22,264 in 2024.[10] In this research, reported complaints regarding synthetic sexual content (AI generated imagery) were low (46), however, RPH state that this represents a significant increase of 119 per cent compared to the previous year (21), indicating a concerning increasing trend in production of synthetic content. AI will only exacerbate the current trend and allow NCII abuse to grow in scale rapidly.

5. The consequences of failing to introduce a lawful framework for testing AI systems are significant, with far-reaching implications for public safety, victims, and law enforcement.
   - **Unchecked deployment of AI models**: Developers currently struggle to assess whether their safeguards are effective in preventing the generation of illegal content. This leaves critical vulnerabilities undetected and allows AI models to be deployed with harmful capabilities that have not been evaluated, which can be exploited by malicious actors.
   - **Escalating victimisation and gendered harm**: Women and children, who are already disproportionately affected, will face increasing levels of AI-enabled abuse. Without intervention, victims will continue to experience severe psychological trauma, reputational harm, and heightened vulnerability to blackmail and coercion using fabricated or manipulated images. This fuels a growing wave of technology-facilitated VAWG, now identified as one of the fastest-emerging threats.
   - **Operational and enforcement challenges and deterioration of public safety**: Law enforcement agencies are already struggling to differentiate real images from increasingly sophisticated synthetic abuse imagery. The National Crime Agency has been clear that the volume of online child sexual exploitation and abuse is

---

6 BBC, 2026: IWF finds sexual imagery of children which 'appears to have been' made by Grok - BBC News: https://www.bbc.co.uk/news/articles/cvg1mzlryxeo

7 Security Hero, 2023: https://www.securityhero.io/state-of-deepfakes/#advancements-in-deepfake-techonology, https://www.legislation.gov.uk/ukpga/2008/4/part/5/crossheading/pornography-etc. See section 63, which outlines definitions of 'Extreme Pornography'.

8 NPCC, 2024: vawg-strategic-threat-and-risk-assessment-underpinning-and-informing-the-2024-vawg-statement.pdf

9 NWGfL, 2025: The Scale of Non-Consensual Intimate Image (NCII) Abuse: A Data-Driven Global Analysis: https://swgfl.org.uk/magazine/the-scale-of-non-consensual-intimate-image-ncii-abuse-a-data-driven-global-analysis/

10 Revenge Porn Helpline, 2023: revengepornhelpline.org.uk/assets/documents/revenge-porn-helpline-report-2023.pdf?_=1714738699

extremely high and growing[11]. With each report potentially representing a child in need of safeguarding, case demand outstripping capacity results in risks that children will remain in harmful situations as law enforcement's ability to triage is influenced by the AI influx. As AI models become more capable, without a lawful route to conduct testing, this challenge will intensify and present a risk to public safety.

3. SMART objectives for intervention

6. Passage of the Crime and Policing Bill will create a power for the Home Secretary to allow authorised individuals to conduct technology testing activities for illegal content. The technology testing defence will become operational following the introduction of secondary legislation. Successful roll out of the technology testing defence will limit the creation and therefore exposure to AI generated illegal content for the general public and reduce the ability for criminals to use AI to produce harmful content. The technology testing defence will play an important part in the government's aim to halve VAWG in the next decade[12]. This intervention is intended to create a clear, practical, and protected testing process which in turn will collectively strengthen the UK's ability to manage AI-related harms.

7. The testing defence is relevant to addressing the increasing capability of technology, particularly AI, to produce harmful imagery and content. It contributes to wider efforts to reduce online harms, support law enforcement, and strengthen the UK's AI safety and security framework. While it will not be possible to isolate or directly attribute wider societal impacts to this measure alone, enabling earlier identification and mitigation of risks has the potential to reduce harm and support public confidence in the responsible development of AI.

8. At first, there is an expectation that the most immediate uptake in applications will occur from industry front runners who have the infrastructure and experience to conduct testing of AI models within the rigorous parameters designed. Home Office will closely review each application before granting authorisation to test. Testing will require strict adherence to the conditions set out in the secondary legislation, and the supporting guidance that the Home Office will publish. Uptake of the testing defence may be seen over time through increased testing activity through expansion of authorised testers within entities, or via an increase in applications for authorisation over time from a broader range of entities, such as small-to-medium enterprises, and third parties such as charities supporting entities with their testing efforts.

9. Over time, the measure is expected to contribute to a reduction in the generation of AI-produced harmful material, including reduced production of CSAM, EP and NCII outputs. The testing defence will provide greater assurance that entities testing harmful outputs are doing so transparently and according to strict conditions, further negating risks to the general public.

---

[11] Technological Tipping Point Reached in Fight Against Child Sexual Abuse - National Crime Agency: https://www.nationalcrimeagency.gov.uk/technological-tipping-point-reached-in-fight-against-child-sexual-abuse

[12] Freedom from violence and abuse: a cross-government strategy to build a safer society for women and girls – Gov.uk: https://www.gov.uk/government/publications/freedom-from-violence-and-abuse-a-cross-government-strategy/freedom-from-violence-and-abuse-a-cross-government-strategy-to-build-a-safer-society-for-women-and-girls-accessible

4. Description of proposed intervention options and explanation of the logical change process whereby this achieves SMART objectives

10. As outlined in part two of this impact assessment, evidence shows there is an uptake in AI generated CSAM, NCII and EP. This technology is rapidly developing, meaning this issue will only be exacerbated without intervention. From engagement via group and individual discussions with entities in the technology industry, the Home Office understands they are keen to have the powers to test their AI models to ensure they do not produce harmful outputs. To enable this, these entities need a statutory defence to ensure they do not incriminate themselves whilst testing for illegal content under existing laws regarding creation and possession.

11. The preferred option introduces a power, within the Crime and Policing Bill, to create defences at a later stage via secondary legislation, for authorised persons carrying out technology testing activities. Authorisation will be subject to strict conditions such as secure handling and deletion of harmful material, staff vetting, and transparency reporting. Authorisation will be granted on a yearly basis. Throughout the approved testing authorisation period, the Home Office will maintain close oversight requiring strict, detailed timely reporting of issues to ensure security protocols are adhered to and breaches can be addressed.

12. The measure addresses the legal risk developers and safety organisations face from testing AI models for harmful CSAM, EP, NCII capabilities. By enabling controlled testing, the intervention ensures vulnerabilities can be identified and mitigated before deployment, reducing the risk of criminal misuse, limiting the creation of resulting illegal content and supporting public trust in AI governance.

13. This applies a tried-and-tested principle of controlled exemptions used in other domains (for example, forensic testing, cybersecurity red-teaming). It mirrors approaches in international frameworks such as the EU AI Act[13] and U.S. AI Center for AI Standards and Innovation guidance[14], which emphasise proactive testing under strict conditions.

[13] Regulation - EU - 2024/1689 - EN - EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[14] Practices for Automated Benchmark Evaluations of Language Models: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.800-2.ipd.pdf

Table 1: Theory of change – Technology testing defence

Inputs	Activities	Outputs	Outcomes	Long-term outcomes	Impacts
Public sector FTE	Introduce a power to create a statutory defence.	Impact assessment in the public domain, introduction of the power.	Parliament now could create a technology testing defence through the power.	Enhance the UK's position as a trusted market for safe AI development.	N/A
Public sector FTE	Production of legal framework and guidance documents.	Entities familiarise themselves with the testing defence.	Entities have a better understanding of the testing defence and can decide if they wish to apply.	Increased testing activity from a broader range of entities, such as small-to-medium enterprises, and third parties such as charities supporting entities with their testing efforts.	Broader understanding across the AI sector of the legal provisions required.
Public sector FTE and technology company FTE.	Compliance checks are completed as part of the application process.	Authorisations for the technology testing defence. Confirmations that authorised entities are compliant.	Ensure that entities who intend to use the defence are doing so within parameters of the authorisation.	Reduced likelihood of a breach of the conditions of the defence.	The Home Office build a picture of the type of entities applying for the defence and ability to comply with the conditions.
Public sector FTE and technology company FTE.	Spot checks are undertaken for authorised entities.	Confirmation of ongoing compliance.	Increased assurance in testing regime.	Greater assurance in testing regime.	Improved public trust in technology governance.
Technology company FTE, Home Office FTE	Entities apply for the AI testing defence. Applications are reviewed by the Home Office.	Number of entities that are granted authorisation increases.	Increased testing of AI models to identify harmful content, Entities benefit from reduced legal risk with testing their models.	Harmful outputs are identified and designed out, and safety features are enhanced.	Enhance the UK's position as a trusted market for safe AI development. Improved industry self-regulation.

5. Summary of long-list and alternatives

14. This section details the long-list and alternatives. It includes a description, assessment of scope for SaMBA and medium-sized businesses, and the reasons for rejection ahead of the short-list stage. Option 0 and Option 1 (the preferred option) are not included to avoid duplication with Section 6.

15. The long list of options has been assessed against the following critical success factors (CSFs), in line with Green Book guidance:

CSF 1: **Objective and strategic fit**: how well the option meets the SMART objectives, provides a statutory defence for technology testing and aligns with VAWG Strategy commitments[15].

CSF 2: **Value for money**: how well the option optimises social value in terms of potential costs, benefits and risks.

CSF 3: **Supplier capacity and capability**: how well potential suppliers can feasibly provide the inputs that are required for the option.

CSF 4: **Affordability**: how well the option can be financed from available government funds.

CSF 5: **Achievability**: how well the option is likely to be delivered given an organisation's ability to respond to the changes required.

Table 2: Comparison of long-list options

Option	CSF assessment (meets CSFs / reasons option not taken forward)	SaMBA and Medium-Sized Business Scope
Option 0 – 'Do nothing' (Business as usual)	Retain current legislation without introducing a statutory defence for authorised testing. **CSF 1 – Objective and strategic fit:** Fails to address the legal risk preventing compliance testing, leaving AI models vulnerable to misuse and escalating harm. **CSF 2 – Value for money:** No improvement in alignment with government objectives on reducing VAWG and improving national security, therefore failing to meet CSFs 1 and 2.	There will be no impact on small and micro or medium size businesses.

[15] Freedom from violence and abuse a cross-government strategy, gov.uk. https://www.gov.uk/government/publications/freedom-from-violence-and-abuse-a-cross-government-strategy

Option 1 – Introduce a power to create a technology testing defence	Introduce a power within the Crime and Policing Bill to create a statutory defence to allow authorised persons to test technology, such as AI models, for CSAM, EP, and NCII content. This authorisation would be subject to conditions to ensure testing is carried out safely, this would include secure handling and deletion of harmful material, staff vetting, and transparency reporting. The authorisation would be time limited and subject to review before being renewed. The defence would allow authorised testers to identify vulnerabilities, verify safeguards, and design out harmful capabilities in multi-modal AI models (for example, text-to-image, image-to-image, text-to-video). It would also provide legal certainty, enable proactive risk management and support law enforcement by reducing the volume of harmful imagery and improving victim identification. **This is the government's preferred option.** **Reasons for Selection** **CSF 1 – Objective and strategic fit:** Directly addresses legal barriers to testing: addresses the legal risk arising from under child protection and obscenity laws that currently prevents transparent testing of AI models. **CSF 2 – Value for money:** Aligns with government objectives on reducing VAWG, creating a safer environment for AI development, and making AI models safer for public use and therefore provides value for money. **CSF 3 – Supplier capacity and capability:** This option is opt-in which avoids imposing costs on companies that lack the technical capacity to meet requirements. Small and micro businesses are likely to lack the resources to engage and low uptake is expected amongst medium-sized businesses.	Small and micro businesses will be in scope but are unlikely to be directly impacted because the technology testing defence is opt-in and only applies to organisations that apply for authorisation and can meet strict eligibility and safety requirements, including secure data storage and compliance reporting. Those without sufficient in-house capability may be able to participate by partnering with larger or third-party organisations to meet the required standards[16]. In the early phase of the scheme, it is expected that initial uptake will be limited and the number of authorised entities estimated to be around one to five. Engagement with subject matter experts will continue to inform the development of secondary legislation and the impact on small and medium size businesses. Medium-sized businesses are more likely to be in scope than SaMBAs, but still face the same high bar for authorisation to meet the required eligibility standards. Many such firms have the necessary infrastructure to comply

[16] Engagement with larger entities highlighted that these entities may struggle with resourcing testing work due to the dispersed nature of technology companies (spanning multiple jurisdictions). It follows that the same issues will be present for SMEs, although this market has not been engaged directly.

	CSF 4 – Affordability: This option includes cost recovery for administrative activities through the charging of administrative fees paid by applicants. **CSF 5 – Achievability:** Testing under the statutory defence remains voluntary and targeted and the initial uptake is expected to be small.	with the prescribed conditions, and those that do not may be able to partner with accredited larger or third-party organisations to test their AI model safeguards on their behalf. The measure does not create blanket exemptions by firm size; instead, it limits burdens by being voluntary and targeted, avoiding imposing requirements on firms that do not want (or are unable) to undertake testing. Where businesses cannot meet requirements alone, the main mitigation is partnering with larger or third-party organisations to conduct testing on their behalf. The rationale for not having size-based exemptions is that the activity involves handling highly harmful/illegal content, so safeguards (and eligibility) need to be informed by risk and capability, not company size.
Option 2 – Non-statutory guidance	Issue government guidance on how to mitigate illegal content production using measures covered under existing law. This would involve publishing best practice principles for safe testing, including recommendations on operational safeguards, secure handling of harmful outputs, and reporting obligations. The guidance would aim to support the technology industry in identifying vulnerabilities and improving safeguards without legislative change, but would be	Guidance would apply universally, but small and micro businesses lack capacity and resource for comprehensive testing, limiting effectiveness. Medium-sized businesses could adopt guidance but without legal certainty, businesses may be unlikely to conduct testing.

	limited given the lack of statutory protection under this option. **CSF 1 – Objective and strategic fit:** The option does not provide a statutory defence for this testing; therefore, businesses would not be able to conduct lawful testing, preventing the implementation of safeguarding improvements. **CSF 2 – Value for money:** Ineffective in reducing harm or improving trust, as voluntary compliance cannot guarantee consistent standards. This option does not enable safeguarding improvements and therefore does not provide value for money. **CSF 3 – Supplier capacity and capability:** This option is applicable universally, but small and micro businesses are likely to lack the resources to engage and low uptake is expected amongst medium-sized businesses.	
Option 3 – Industry self-regulation	Rely on AI developers formally creating internal compliance frameworks for testing and safeguarding against harmful outputs. This approach would encourage companies to set their own standards for risk assessment, testing protocols, and mitigation measures, supported by voluntary codes of conduct. **CSF 1 – Objective and strategic fit:** Not pursued as it does not provide a statutory defence for this testing, meaning that businesses would not be able to conduct meaningful or lawful testing. Self-regulation could lead to inconsistent standards across industry, potentially leading to unsafe testing practices. **CSF 2 – Value for money:** Market incentives are insufficient; evidence from the OSA 2023 shows voluntary measures fail to address systemic harms. This option does not enable safeguarding improvements and therefore does not deliver value for money. **CSF 3 – Supplier capacity and capability:** Small and micro business are likely to lack	Small and micro businesses are unlikely to implement costly compliance frameworks due to resource constraints. Medium-sized businesses may attempt partial adoption but lack enforcement mechanisms, reducing overall effectiveness.

	the resources to engage, and partial adoption is expected amongst medium sized businesses.	
Option 4 – Broader legislative reform	Undertake a comprehensive overhaul of child protection and obscenity laws to create exemptions for authorised AI testing. This would involve amending multiple statutes (for example, Protection of Children Act 1978, Sexual Offences Act 2003, Criminal Justice and Immigration Act 2008) to clarify liability and introduce new provisions for controlled testing environments. **CSF 1 – Objective and strategic fit:** This approach may align with strategic fit and objectives as it offers entities some protection for legal testing, but risks conflicting with operational law enforcement **CSF 2 – Value for money:** Existing legislation is effective, and a comprehensive overhaul would be resource intensive for government and businesses, representing poor value for money. **CSF 3 – Supplier capacity and capability:** This option requires significant compliance burdens for businesses, including small and micro businesses as all developers would be in scope. **CSF 4 – Affordability:** A comprehensive overhaul would be resource intensive for government. **CSF 5 – Achievability:** Slower to implement than a targeted statutory defence. This option would incur a large compliance burden reducing achievability.	Would bring all developers into scope, including small and micro businesses, imposing significant compliance burdens. Evidence suggests this would be disproportionate and could stifle innovation among smaller firms.
Option 5 – Develop a Memorandum of Understanding (MoU) between the Home Office, law enforcement,	The MoU would set out operational safeguards for testing technology models. The MoU would clarify expectations and provide assurance that, where agreed safeguards are followed, enforcement action would not normally be pursued. This would build trust and operational clarity between government, industry, and law enforcement. **CSF 1 – Objective and strategic fit:** The Home Office have received legal advice that	Small and Micro Businesses: Excluded due to resource constraints and inability to meet MoU requirements. Medium-Sized Businesses: Included if they could demonstrate capability to comply with safeguards.

and authorised AI testing entities	an MoU does not create the same legal certainly as a legislation would therefore failing to create the conditions for testing and not aligning with strategic fit and objectives. **CSF 2 – Value for money:** This option would be resource intensive for government and therefore does not deliver value for money. **CSF 3 – Supplier capacity and capability:** Small and micro businesses are expected to be excluded due to resource constraints, and few medium sized businesses are expected to demonstrate ability to comply with the established safeguards.	

Table 3: Rating of options against the CSFs

	Option 0 - (Do nothing)	Option 1 – Introduce a power to create a technology testing defence (preferred)	Option 2 - Non-statutory guidance	Option 3 - Industry self-regulation	Option 4 - Broader legislative reform	Option 5 - Develop MoU
CSF 1: Objectives and strategic fit	Red	Green	Red	Red	Amber	Red
CSF 2: Value for money	Red	Green	Red	Red	Red	Red
CSF 3: Supplier capability and capacity	Green	Amber	Amber	Red	Red	Red
CSF 4: Affordability	Green	Green	Green	Green	Red	Green
CSF 5: Achievability	Green	Green	Green	Green	Red	Green

RAG key: Green = fit for purpose (meets the critical success factor). **Amber** = partially fit for purpose (partially meets the critical success factor; issues/risks to address). **Red** = not fit for purpose (does not meet the critical success factor).

6. Description of shortlisted policy options carried forward

16. Following the consideration of the options at the long list stage, the options shortlisted for assessment in this IA are:

- **Option 0:** Business as usual ('Do nothing')
- **Option 1:** Introduce a power to create a technology testing defence. **This is the government's preferred option.**

7. Regulatory scorecard for preferred option

Part A: Overall and stakeholder impacts

(1) Overall impacts on total welfare		Directional rating
Description of overall expected impact	The technology testing defence reduces the legal risk for authorised entities to test and design out harmful capabilities. The regulation should benefit households as the measure should reduce the production of CSAM/EP/NCII which carry societal harms. The net impact on authorised entities is likely to be positive. Entities which apply for the defence will face increased costs, but this may be outweighed by the non-monetised benefit of reduced legal risk. Additionally, the technology testing defence uplifts industry standards. The overall impact is likely to be positive as the illegal content which the technology testing defence aims to reduce carries high unit costs. Improved safeguarding implies a reduction in total societal harm associated with this content.	**Positive** **Based on all impacts (incl. non-monetised)**
Monetised impacts	Overall Net Present Social Value (NPSV) is -£1.24 million, with a range of -£0.35 million to -£3.05 million. Over half of this estimate driven by costs to the Home Office for processing the applications from entities to become authorised.	**Negative** **Based on likely £NPSV**
Non-monetised impacts	Non-monetised costs include set up costs to the Home Office and entities. The low-probability, high-risk scenario of a breach in the conditions of the testing defence is a non-monetised cost to both the public and private sector. It was not possible to monetise any benefits due to an absence of quantifiable impact driven by data limitations. The benefits listed below are expected to have significant impact which justifies the directional rating of "positive":	**Positive**

	• Reduction in legal risk for authorised entities. • Reduction in AI-generated harmful content. • Improved safety, compliance and oversight of the AI industry. • Enhanced public trust in AI and government governance frameworks.	

Any significant or adverse distributional impacts?	No significant or adverse distributional impacts have been identified.	**Neutral**

(2) Expected impacts on businesses		
Description of overall business impact	Interested entities will face costs associated with familiarisation, fees, and compliance. Authorised entities will benefit from reduced legal risk which is non-monetised. Although it can be assumed that the reduced legal risk outweighs the cost associated with applications, it has not been established whether this benefit outweighs all of the costs to business (including familiarisation and fees). Therefore, the overall impact remains uncertain.	**Uncertain**
Monetised impacts	Business NPV is -£0.37 million, with a range of -£0.09 million to -£1.01 million. This includes the value of fees paid as the fees represent a financial cost to business.	**Negative** **Based on likely business £NPV**
Non-monetised impacts	Non-monetised costs include highly variable costs such as the creation of a physical environment for testing and insurance costs, as well as costs associated with the low-probability scenario of a breach of the testing conditions (such as legal challenges and reputational damage). Non-monetised benefits include the reduced legal risk when developing and testing technology.	**Uncertain**
Any significant or adverse distributional impacts?	The technology sector will be impacted. Since entities would be choosing to opt-in to de-risk their activity, this is not considered to be significant or adverse. No significant or adverse regional impacts have been identified.	**Neutral**

(3) Expected impacts on households		
Description of overall household impact	Affected households will benefit from improved safeguards. Higher barriers to the production of CSAM/EP/NCII imply a reduction in the societal harms associated with this material.	**Positive**
Monetised impacts	It has not been possible to generate monetised impacts to households due to an absence of evidence on the real-world impact and a absence of data to indicate scale.	**Uncertain**

Non-monetised impacts	Non-monetised benefits include a reduction of the societal harms associated with CSAM/EP/NCII, particularly the emotional harm associated with AI-enabled abuse.	**Positive**
Any significant or adverse distributional impacts?	Victims and potential victims of CSAM, EP, NCII are expected to benefit from better safeguarding for technology enabled abuse.	**Positive**

Part B: Impacts on wider government priorities

Category	Description of impact	Directional rating
Business environment: **Does the measure impact on the ease of doing business in the UK?**	The intervention is expected to strengthen the UK's attractiveness as a destination for technology investment by providing a clear and lawful route for responsible model testing and reduced legal uncertainty.	**Supports**
International Considerations: **Does the measure support international trade and investment?**	The regulation is likely to enhance the UK's position as a trusted market for safe technology development. This may support international investment, but overall attractiveness depends on the approaches taken by international partners.	**Uncertain**
Natural capital and Decarbonisation: **Does the measure support commitments to improve the environment and decarbonise?**	There is no expected environmental impact as part of this intervention.	**Neutral**

8. Monitoring and evaluation of preferred option

17. The Home Office intends to carry out a post-implementation review (PIR) 12-months after secondary legislation comes into force. This timeline has been chosen because:

- The fast-moving nature of AI capabilities and the associated risks, which evolve significantly within a year rather than over longer cycles.
- The need to quickly assess whether the statutory defence is operating safely and as intended.
- The critical importance of detecting early indicators of non-compliance, emerging harms, or operational challenges before they become systemic.
- The evidence that AI-enabled harms are developing rapidly and existing data sources remain incomplete, making early reassessment essential.

- A 12-month PIR ensures the system remains responsive, proportionate, and aligned with technological realities.

18. The PIR will evaluate the process against the objectives included in this impact assessment. This will consist of an assessment of whether the original objectives have been met and will be addressed by setting up an authorisation unit within the Home Office, who will manage applications and compliance performance. This includes assessing the proportion of entities meeting the conditions and the number of entities who are authorised, the quality and timeliness of evaluation reports, and identification of breaches and supplementary actions. The Home Office will conduct the 12-month PIR, evaluating the overall success of the technology defence against a range of measures, and assessing what steps should be taken, if any, to improve the scheme. The assessment measures will be refined and confirmed in due course.

19. Alongside the PIR, the authorisation unit will closely monitor the process and gather data from authorised entities throughout implementation of the testing defence to ensure entities adhere to the conditions, and appropriate action is taken to address any breaches. Monitoring will rely on multiple mechanisms, such as:

    - Spot checks and independent audits to collect data and assess compliance against prescribed conditions. This would be triggered where risk factors are identified, or concerns emerge regarding appropriate use of authorisation and testing safety.
    - Annual accreditation cycles and compliance reviews for those entities seeking to renew their authorised status, assessing whether these entities meet the prescribed conditions.
    - Mandatory evaluation reports, completed after each testing period.

20. Success of the technology testing defence will be assessed against the objectives listed in section three of this impact assessment. Successful implementation would include:

    - High interest and uptake in authorisation (five entities or more authorised) to conduct the testing defence from technology industry leaders.
    - Excellent adherence from authorised entities to the conditions and safeguards that will be set out in regulations, without breaches.

21. The main external factors affecting the success of the intervention include:

    - Pace of technology development, such as generative AI, which may rapidly create new exploitation vectors not covered by the initial framework.
    - Baseline data limitations, including under-recording of AI-enabled criminal activity and lack of traceability.
    - Rate of public adoption of AI tools, which may both increase risks or improve public understanding.
    - Emergence of new forms of criminality enabled by AI, which may necessitate policy or legislative updates.
    - Industry capability and investment in safety testing infrastructure and compliance functions.

22. Early review may be required if there are jumps in AI technological advancement and rapid adoption of AI tools that alters risk levels. Misuse or failure to comply with the conditions of testing could also trigger an early review.

9. Minimising administrative and compliance costs for preferred option

23. The preferred option has been designed to keep administrative and compliance burdens as low as possible for authorised entities, while still maintaining effective safeguards. Internal documents set out several measures that directly minimise costs:

    a. Clear, concise legislation and guidance.

        1) The legislation and accompanying guidance will be drafted to be as clear and concise as possible, reducing familiarisation time and removing unnecessary complexity.

    b. Supplementary supporting guidance is expected to be proportionate, keeping reading time manageable for industry.

    c. Use of templates and standardised reporting

        1) The preferred option includes the use of templates, standardised forms and structured reporting requirements, allowing entities to comply without designing their own systems from scratch.

        2) This significantly reduces the burden of form-filling, annual reporting, and compliance documentation.

    d. Proportional and risk-based compliance requirements

        1) The technology testing defence is opt-in and is specific to certain harms. There are fewer affected entities compared to broader safeguarding measures such as the Online Safety Act 2023 (OSA 2023).

        2) This avoids imposing costs on companies that lack the technical capacity to meet requirements.

    e. Minimal additional administrative fees

        1) Administrative fees will be limited to cost-recovery only, with fees waived for law enforcement bodies.

        2) This keeps upfront application and renewal costs low.

    f. Limited scope and voluntary participation

        1) Testing under the statutory defence remains voluntary and targeted. Only organisations that actively seek authorisation incur compliance obligations.

        2) Initial uptake is expected to be small, overall compliance costs across the sector remain low.

10. Main assumptions / sensitivities and economic / analytical risks

24. A main analytical risk stems from uncertainty as the technology testing defence is novel and takes effect through secondary legislation. Cost estimates to both the Home Office and entities rely on assumptions developed with policy experts. For example, the length of documentation (and the time taken to fill this out and review it) remains under development. Additionally, costs will scale with the volume of applications which is estimated over the 10-year appraisal period based on limited market engagement. To mitigate these risks ranges have been developed around uncertain variables.

25. It is assumed that there will be 100 per cent compliance with the conditions of the technology testing defence as there will be strong deterrents including potential prosecution and rigorous safeguarding measures. There is a low-probability, high-impact risk of the conditions of the technology testing defence being either accidentally or maliciously breached. This would have a reputational impact for the Home Office and for the affected entities, this may also result in legal fees, and could have an impact on the justice system if a criminal fine is sought following a serious breach. This has not been monetised as part of this IA.

26. It is assumed that the technology testing defence leads to a reduction in illegal content resulting in reduced societal harms. Despite engagement with industry partners, this mechanism is novel and as such not yet proven so there is a risk that the technology testing defence is ineffective. In the unlikely case that the defence has very limited impact upon the generation of illegal content, the impact upon overall welfare could be negative as stakeholders including the government and entities would incur costs, and the policy will not achieve its objectives.

Declaration

Department:

Home Office

Contact details for enquiries:

CrimeandPolicingBillTeam@homeoffice.gov.uk

…

Minister responsible:

Minister Kaur

I have read the Impact Assessment and I am satisfied that, given the available evidence, it represents a reasonable view of the likely costs, benefits and impact of the leading options.

Signed: 

Date: 13/10/2026

Summary: Analysis and evidence

Price base year: 2026/27

PV base year: 2026/27

Table 4: Appraisal Summary Table for Technology Testing Defence

		Option 0: Business as usual (baseline)	**Option 1: Introduce a power to create a technology testing defence**
Costs (£m)	Low	NA – no change in relation to baseline	£0.35 million
	High	NA – no change in relation to baseline	£3.05 million
	Best	**NA – no change in relation to baseline**	**£1.24 million**
			Setup costs include familiarisation costs and training costs. Familiarisation costs fall to the private sector. Training costs fall to the public sector and private sector. All other costs are ongoing and include application costs and costs associated with the monitoring and oversight of authorised entities which fall to the public sector and private sector.

Benefits (£m)	Low	-	-
	High	-	-
	Best	-	-
			No benefits of this policy have been monetised.
Net present social value (£m)	Low	NA – no change in relation to baseline	-£0.35 million
	High	NA – no change in relation to baseline	-£3.05 million
	Best	**NA – no change in relation to baseline**	**-£1.24 million**

	Option 0: Business as usual (baseline)	**Option 1: Introduce a power to create a technology testing defence**
Public sector financial costs (with brief description, including ranges)	NA – no change in relation to baseline	Public sector financial costs are estimated to range between **£0.25 million and £2.04 million**. Public sector financial costs include the cost of processing applications, staff training costs and costs of monitoring authorised entities.
Significant un-quantified benefits and costs (description, with scale where possible)	NA – no change in relation to baseline	Un-quantified benefits include a reduction in technologically generated harmful content, improved safety, compliance and oversight of the technology sector, and enhanced public trust in technology and government governance frameworks.
Key risks (and risk costs, and optimism bias, where relevant)	NA – no change in relation to baseline	It is assumed that there will be 100 per cent compliance with the regulation. There is a low-probability, high-impact risk of the conditions of the technology testing defence being breached. This would have a reputational impact for the Home Office and for the affected entities, may incur legal fees, and could have an impact on the Criminal Justice System (CJS).
Results of sensitivity analysis	NA – no change in relation to baseline	Ranged variations in critical variables (such as the time taken for the Home Office to process applications and the time taken for businesses to produce required paperwork) have been used to capture uncertainties around the technology testing defence.

Evidence Base

A. Problem under consideration, with business as usual, and rationale for intervention

Strategic Objective

1. Successful roll out of the technology testing defence will reduce exposure of AI-generated illegal content to the public and reduce criminals using AI to produce harmful content. The technology testing defence is aligned with the government's commitment to halve VAWG in the next decade. This intervention is intended to deliver a clear, practical, and protective testing process which in turn will collectively strengthen the UK's ability to manage AI related harms.

Background

2. There is currently no explicit statutory defence that allows authorised testing of AI models which may generate illegal or harmful content. Testing to deliberately generate illegal content is illegal in the UK regardless of intent. As a result, there are potential legal barriers for technology companies and safety companies wanting to test whether AI models can produce illegal content. This legal risk may prevent developers, researchers, independent evaluators, and public safety organisations from lawfully probing how and when AI systems generate harmful outputs.

3. This constraint creates a systemic safety gap: legitimate actors cannot identify harmful failure modes or validate whether safeguards function effectively before deployment. AI models may be released with unknown vulnerabilities, enabling malicious users to exploit weaknesses that lawful testers are currently prohibited from uncovering.

4. Existing legislation continues to criminalise all handling of illegal content, with no exemption for controlled testing environments. Currently, safety-critical testing cannot occur, risks cannot be properly mitigated, and the market cannot self-correct because of legal uncertainty over possible evaluation routes. As AI capabilities expand, the frequency and sophistication of harmful outputs is expected to increase, compounding the risks inherent in the business as usual (BAU) position.

5. Government intervention is necessary because the market cannot resolve the legal barriers that prevent safe and responsible AI testing. Without a statutory mechanism allowing authorised testing:

   - Harmful content risks remain unidentified until after deployment.
   - Safeguards cannot be validated or benchmarked.
   - Developers and safety assessors face legal risk that deters thorough testing.
   - Malicious actors may exploit vulnerabilities that responsible actors are legally unable to detect

Groups affected

6. The policy impacts several groups with the most affected being:

   a) AI developers and researchers — would gain a lawful route to test harmful outputs; only those meeting stringent requirements would be eligible.

b) Law enforcement and child protection agencies — benefit from clearer frameworks and reduced circulation of harmful imagery.

c) Victims of CSAM, EP, NCII (primarily women and children) — expected to see reduced abuse and better safeguarding.

7. Other affected groups include:

d) Medium sized AI companies — may be eligible for authorisation if they can meet compliance standards.

e) Small and micro businesses — this group could be in scope, however, these may be limited by resource constraints and the inability to meet necessary strict security and governance requirements for this specific type of testing. The testing defence is opt-in and authorisation based (rather than a blanket measure), limiting the impacts to these stakeholders.

f) Third sector organisations – groups such as the Internet Watch Foundation may gain authorisation to test on behalf of companies.

Consultation

8. Initial consultation has been undertaken within government and with the public, but a documented formal consultation has not yet been conducted. The findings are presented below.

Within government

9. The Home Office has worked closely to ensure cross-government discussion and collaboration among:

a. Home Office policy teams

b. Law enforcement stakeholders

c. The Artificial Intelligence Safety Institute (AISI)

d. Other relevant government departments and devolved administrations engaged in online safety, criminal justice, and digital regulation.

10. Input was sought to ensure the proposed power aligns with existing legislative frameworks (for example, OSA 2023) and operational demands.

Public Consultation

11. The Home Office has started conversations with a range of external stakeholders, including AI developers, safety researchers, industry bodies, victim support organisations, and experts in online harms and VAWG. These early engagements have helped shape the department's understanding of the operational, ethical and technical considerations needed for a safe testing framework.

12. The Home Office will continue to engage closely with these groups throughout the development of the operational framework, ensuring their expertise and perspectives inform the design of safeguards, authorisation requirements, compliance expectations and oversight mechanisms. Ongoing discussion will support a framework that is practical, proportionate, and responsive to emerging risks.

Policy objective

13. The main policy objective is to enable lawful, responsible, and effective safety testing of AI models so that developers, independent evaluators, and public safety bodies can identify and mitigate harmful outputs before deployment. The intervention aims to close the current statutory gap that prevents lawful handling of illegal content (such as CSAM, EP, and NCII) even in controlled testing environments.

14. Specifically, the intervention seeks to:

- Enable authorised AI testing to identify harmful failure modes
    - Provide a statutory mechanism allowing accredited or authorised testers to probe AI models for harmful outputs and assess model safeguards, removing the current legal barrier that prevents thorough evaluation.
- Improve the safety and reliability of AI models before deployment
    - Ensure developers can validate whether safeguards function effectively, identify vulnerabilities, and make necessary improvements prior to release, reducing the risk of malicious exploitation.
- Address a clear market and regulatory failure
    - Intervention is required because the market cannot self-correct; developers cannot lawfully test their systems under the BAU scenario, which prevents them from ensuring safety.
- Reduce public safety risks associated with untested AI models
    - By enabling lawful safety testing, the intervention reduces the likelihood of harmful material being generated, disseminated, or exploited by malicious actors. The ability to test safeguards is a critical precondition for reducing system-level risks.

B. Description of options considered

15. This section outlines the shortlisted options, including the BAU and the government's preferred option. In line with Green Book methodology, a longlist of options was reviewed before refining to a shortlist for detailed appraisal.

Option 0: 'Do nothing' (Business as Usual)

16. Under this option, current legislation continues to criminalise any creation, possession, or distribution of illegal content, even when generated solely for the purposes of technology testing. Developers and safety evaluators cannot lawfully test technology models to identify harmful failure modes or to validate safeguards. This results in persistent safety gaps: vulnerabilities remain unknown prior to deployment and may be exploited by malicious actors. This option fails to address the root cause of the problem and is not considered viable from a public safety perspective.

17. **Option 0** fails to address the root cause of the problem and is not viable from a public safety perspective but is taken forward as a benchmark

Option 1: Introduce a power to create a technology testing defence

18. This option introduces a statutory defence that permits authorised individuals and organisations to lawfully generate, access, and analyse illegal content strictly for the purposes of technology testing. This would allow:

    - Identification of harmful failure modes,
    - Validation of safeguard effectiveness, and
    - More comprehensive pre-deployment risk assessment.

19. The absence of a lawful testing route is the central regulatory failure, preventing responsible technology assurance. A statutory defence addresses this market failure by creating a controlled, accountable mechanism for safe testing.

20. **Option 1** addresses the regulatory failure directly and is the preferred option.

NPSV: monetised and non-monetised costs and benefits of each shortlisted option

21. This section details the appraisal for the preferred option. The section begins by detailing monetised costs to entities and monetised costs to the public sector. A narrative section outlines the non-monetised costs to both entities and the public sector which is followed by further detail around benefits to both entities and the public sector (all benefits are non-monetised). Lastly, the section notes that the technology testing defence is unlikely to have an impact on small and micro-businesses.

22. The Net Present Social Value (NPSV) is estimated over a 10-year appraisal period beginning in 2026/27 using a 3.5 per cent discount rate in line with HM Treasury Green Book guidance[17]. All figures are presented in 2026/27 prices. In line with HM Treasury Green Book guidance, fees (including the application fee and Disclosure and Barring Service (DBS) checks) are not included within the NPSV as they represent an economic transfer. However, these fees are included in the Business Net Present Value section as they represent a real financial cost to entities.

23. The Net Present Social Value (NPSV) of the preferred option is estimated to be **-£1.24 million, with a range of -£0.35 million to -£3.05 million.** The direct impact on business is estimated to be -£0.37 million, with a range of -£0.1 million to -£0.96 million. Non-monetised benefits (such as an increase in safety) may offset or outweigh the costs of the technology testing defence. These benefits have not been monetised due to an absence of quantifiable impact driven by data limitations. As such, the NPSV only reflects the costs of the technology testing defence.

24. In cases where data is not available, the best estimate from policy experts based on current operational understanding of the defence is used and this is noted throughout. Variables that are estimated in this way include the expected length of time spent processing documentation and the anticipated volume of applications. To mitigate this risk, ranges have been developed to capture uncertainty around these variables.

[17] Due to discounting over a 10-year appraisal period the totals costs will be less than the multiplication of their parts.

Monetised costs to entities

Table 4: Summary table of monetised costs to entities, 10-year total, 2026/27 PV, £ millions

	Low	Central	High
Familiarisation costs	0.00	0.02	0.1
Application costs	0.03	0.11	0.37
Costs to authorised entities	0.08	0.23	0.49
Total costs to business	**0.10**	**0.36**	**0.96**

Source: Home Office internal estimates.

25. The total costs to entities are estimated between £0.10 million and £0.96 million, with a central estimate of £0.36 million. Notably, this section includes the fees paid for submission of the technology testing defence application and for DBS checks as these fees represent a real financial cost to entities. The following section details the methodology for each of the cost lines.

Familiarisation costs

26. Familiarisation costs capture the opportunity cost of the time spent reading and comprehending the new guidance. Familiarisation costs are estimated as:

Hours taken to read the guidance x hourly wage x volume of readers

27. There are two familiarisation costs to business. Firstly, an initial familiarisation cost as entities determine if they are in scope of the technology testing defence. Secondly, a secondary familiarisation cost is incurred by entities who intend on applying for the defence and have not been granted authorisation previously. It is assumed that entities that have been granted authorisation previously will not need to re-familiarise with the guidance. This approach mimics the familiarisation approach proposed for the OSA 2023[18].

28. For both the initial and secondary familiarisation, it is estimated that length of the guidance will be between 10 to 15 pages, with a central estimate of 12.5 pages. The length of the guidance is based on the breadth of the measure and regulations set out in initial policy drafting. The central reading speed estimate is 300 words per minute with a range of 200 to 700 words per minute[19]. A comprehension multiplier between 85 per cent and 60 per cent[20], with a central estimate of 80 per cent, is applied. The central estimate for reading speed and comprehension assumes that the person reading the guidance is a "good reader" which is appropriate given the expected literacy and technical proficiency of employees within the technology sector.

[18] Online Safety Act 2023 Impact Assessment (paragraph 131 details familiarisation costs): https://assets.publishing.service.gov.uk/media/6716222b9242eecc6c849b09/Online_Safety_act_enactment_impact_assessment.pdf#page=42

[19] Words per minute estimates for screen reading, Readingsoft: http://www.readingsoft.com/

[20] Comprehension rates, Readingsoft: http://www.readingsoft.com/

29. The opportunity cost is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[21], which has been uplifted by 22 per cent[22] to account for non-wage labour costs.

30. For initial familiarisation costs, the number of entities expected to read the guidance is between 30 (the number of entities expected to apply[23] for the technology testing defence in the first year of the scheme) and 700 (the number of businesses in scope for secondary familiarisation of the OSA 2023 which is a larger, tangential, regulation). The central estimate of 365 entities is based on the midpoint of this range. It is assumed that one person in each entity would read the guidance to determine if the entity is in scope. This IA estimates initial familiarisation costs to be between £100 and £22,300, with a central estimate of £4,400. The upper estimate is significantly larger than the lower and central estimates to capture the possibility that many entities will read the guidance. It is likely that this number is an overestimate as the measure is only for generative tools which is more niche than broader safety measures such as the OSA 2023.

31. For secondary familiarisation costs, it is assumed that entities will read the guidance if they intend to apply for the technology testing defence and have not been granted authorisation previously. It is reasonable to assume that entities that have been granted authorisation in the past will not need to re-read the guidance as the application process will be standardised. Over the 10-year appraisal period, it is estimated that between 28 and 270 entities will require secondary familiarisation, with a central estimate of 129 entities. For each entity, it is assumed that each employee who would be involved in testing (5 to 10, with a central estimate of 7.5 employees[24]) would be required to familiarise themselves. This IA estimates secondary familiarisation costs to be between £400 and £77,900, with a central estimate of £10,800. Similarly to initial familiarisation costs, the upper estimate captures the possibility of a high level of interest driving a large volume of applications.

32. Following the methodology outlined above, total familiarisation costs are estimated between £542 and £100,182, with a central estimate of £15,140.

Application costs: Application form

33. Submission of the technology testing defence application incurs a fee payable to the Home Office and two opportunity costs; the opportunity cost of time taken filling in the application form and the opportunity cost of time taken meeting with legal advisors to ensure that the entity meets the requirements for authorisation.

34. For the fee and both opportunity costs, the volume of applications over the 10-year appraisal period is estimated between 140 and 500, with a central estimate of 300 applications. Large entities are expected to apply who are likely to have a high level of

---

[21] GONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14 Good reader defined by ReadingSoft as having a comprehension rate of 80% and a reading speed of 300 WPM on a screen. https://readingsoft.com/
ONS: - https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

[22] 22 per cent uplift based on ONS (2020) Index of labour Costs per Hour, seasonally adjusted.: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

[23] NB: 30 entities are expected to apply based on limited engagement with industry, but not all will be authorised by the Home Office. However, it is reasonable to assume that all entities that apply will incur familiarisation costs.

[24] Number of employees in a testing team based on current operational understanding from policy experts.

technical and legal expertise. Therefore, it is expected that all entities that apply will get the application form 'right first time' and will not require repeat applications.

35. The fee per application is estimated[25] using existing fees from other government licences. The fee is estimated between £44 and £97, based on the value of fees for a driving licence and a passport respectively. The central estimate, £70 per application, is the midpoint of these two values. Multiplying the volume of expected applications by the fee cost, this IA estimates the fee costs between £5,100 and £41,600, with a central estimate of £18,100.

36. Both opportunity costs are calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[26], which have been uplifted by 22 per cent[27] to account for non-wage labour costs. The volume of applications over the 10-year appraisal period are estimated between 140 and 500, with a central estimate of 300 applicants.

37. These opportunity costs are calculated as:

Hourly wage x volume of applications x hours taken

38. To fill in the application form, it is assumed that it takes one person between 0.5 and 1 hour with a central estimate of 0.75 hours based on the expected length and standardised nature of the application form. This IA estimates the opportunity cost of filling in the application form between £1,800 and £13,100, with a central estimate of £5,900.

39. It is assumed all employees who would be involved in testing (5 to 10, with a central estimate of 7.5 employees) would spend between one hour and half a day (3.7 hours), with a central estimate of 2.4 hours[28], meeting with legal advisors to determine if their entity can become authorised. No legal fees are estimated as the large entities expected to apply are likely to have in-house legal teams. No familiarisation costs are estimated for these legal teams as navigating the legislation surrounding this matter is likely to fall within the core duties of in-house legal teams.

40. This IA estimates the opportunity cost of would-be testers meeting with their legal teams to be between £1,300 and £9,700, with a central estimate of £4,600.

41. Following the methodologies outlined above, the total cost associated with the application form is between £8,200 and £64,500, with a central estimate of £28,700.

Application costs: Compliance visits

42. Entities may be subject to a compliance visit from Home Office inspectors before authorisation is granted. There is an opportunity cost of time as one individual from the would-be testing team is required to accompany the Home Office inspectors during the compliance visit. This opportunity cost is calculated as:

---

25 The value of the fee has not been finalised therefore this serves as the best estimate. The fee attempts to recoup application processing costs from the Home Office side which are estimated to be £60 per application, roughly in line with the central estimate of a £70 fee.

26 ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

27 22 per cent uplift based on ONS (2020) Index of labour costs per hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

28 The time taken meeting legal advisers is based on the current operational understanding of policy experts.

Hourly wage x hours taken for compliance visit x volume of compliance visits

43. The opportunity cost is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[29], which has been uplifted by 22 per cent[30] to account for non-wage labour costs. Based on current understanding of the visits from policy experts, the time taken for a compliance visit is estimated between three and four hours, with a central estimate of 3.5 hours.

44. A range is used to capture uncertainty around the volume of compliance visits. In the first and second year of the appraisal period, all entities will require a compliance check as the Home Office seeks to establish its understanding of the sector. Similarly, across the entire appraisal period, all entities will require a compliance check at the point of their first application for safety and compliance reasons. Only 50 per cent of repeat applications (excluding applications in the first two years of the scheme) will require a compliance check in the central estimate. In the high estimate, all repeat applicants will require an additional compliance check, and in the low estimate no repeat applicants will require an additional compliance check.

45. This IA estimates the opportunity cost of compliance checks between £900 and £30,500, with a central estimate of £9,800.

Application costs: Disclosure and Barring Service checks

46. Employees who will be testers are required to hold a valid DBS check that is less than three-years old which introduces two costs to entities; the fees paid for DBS checks and the opportunity cost of time spent filling in the DBS check form.

47. For both the fee and the opportunity cost, it is assumed that none of the testing team (5 to 10, with a central estimate of 7.5 employees) holds existing DBS checks. The volume of required DBS checks is the sum of the number of applications from entities without previous authorisation and the number of applications requiring repeat DBS checks (every three years). It is assumed that entities applying without previous authorisation will 'refresh' the DBS checks of their testers to maximise chances of authorised. Although this may be an overestimate, it is partially offset as the estimates do not account for staff turnover over the 10-year appraisal period. The volume of required DBS checks is estimated between 500 and 4,200, with a central estimate of 1,800 DBS checks required.

48. The fee per standard DBS check is £22 which informs the low and central estimates, and the fee for an enhanced DBS check (£51) informs the high estimate[31]. Multiplying the fee by the expected volume of entities and testers requiring DBS checks over the 10-year appraisal period, this IA estimates the costs associated with fees for DBS checks between £9,000 and £183,300, with and central estimate of £34,200.

49. The opportunity cost of time whilst filling out the DBS check form is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the

[29] ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

[30] 22 per cent uplift based on ONS (2020) Index of labour costs per hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

[31] GOV UK: https://www.gov.uk/government/news/dbs-fees-are-changing-in-december

ONS[32], which has been uplifted by 22 per cent[33] to account for non-wage costs. It is assumed that it takes one hour to fill in the application for a DBS check. The opportunity cost is calculated as:

Expected volume of DBS checks x hourly wage x hours taken

50. This IA estimates the opportunity cost of applying for DBS checks over a 10-year appraisal period to be between £14,500 and £107,800, with a central estimate of £48,500.

Costs to authorised entities: Accreditation report and mandatory evaluation report

51. Authorised entities must submit an accreditation report and a mandatory evaluation report (hereafter also referred to as 'paperwork') to the Home Office annually. This is to allow the Home Office to gain an operational understanding of the impact of the technology testing defence. The time taken to produce these reports results in opportunity cost of time which is the largest monetised cost component associated with the AI testing defence. This opportunity cost is calculated as:

Hourly wage x hours taken to produce paperwork x number of authorised entities

52. The opportunity cost is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[34], which has been uplifted by 22 per cent[35] to account for non-wage labour costs. Based on current understanding from policy experts, the paperwork should take 42 hours to complete. A range of 21 to 63 hours has been developed to capture uncertainty around the length of the paperwork and the time taken to complete it. Multiplying this by the volume of entities the Home Office expects to be authorised[36], this IA estimates the opportunity cost of the accreditation report and mandatory evaluation report to be between £69,100 and £425,900, with a central estimate of £211,100.

[32] ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

[33] 22 per cent uplift based on ONS (2020) Index of labour Costs per Hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

[34] ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

[35] 22 per cent uplift based on ONS (2020) Index of labour Costs per Hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

[36] Across the 10-year appraisal period between 130 and 270 entities are expected to be authorised, with a central estimate of 200 entities.

Costs to authorised entities: Spot checks

53. Authorised entities will be subject to Home Office spot checks to ensure compliance with the conditions of the technology testing defence. For the duration of the spot check one employee must accompany the Home Office inspectors thereby incurring an opportunity cost of time:

Hourly wage x hours taken for spot check x volume of spot checks

54. The opportunity cost is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[37], which has been uplifted by 22 per cent[38] to account for non-wage labour costs. Based on current operational understanding from policy experts, spot checks are planned to take between three to four hours, with a central estimate of 3.5 hours. A range has been developed to capture uncertainty around the volume of spot checks; the low estimate assumes each entity will receive a spot check every three years, the central estimate assumes each entity will receive a spot check every two years, and the high estimate assumes each entity will receive a spot check every year.

55. This IA estimates the opportunity cost of spot checks to be between £3,300 and £27,000, with a central estimate of £8,800.

Costs to authorised entities: Dissemination meeting

56. It is expected that authorised testers would seek to reduce risk by fully understanding the terms and conditions of the technology testing defence through a dissemination meeting. This incurs an opportunity cost of time which is calculated as:

Hourly wage x hours taken at dissemination meeting x volume of newly authorised entities

57. The opportunity cost is calculated with the annual median hourly wage of Quality and Regulatory Professionals, published by the ONS[39], which has been uplifted by 22 per cent to account for non-wage labour costs. Based on the length of the guidance, policy experts have advised that a dissemination meeting should take between one and half a day (3.7 hours), with a central estimate of 2.4 hours, and the entire testing team will attend (5 to 10 employees, with a central estimate of 7.5 employees). This is multiplied by the entities authorised for the first time across the 10-year appraisal period (20 to 40 entities, with a central estimate of 30 entities).

58. This IA estimates the opportunity cost of attending a dissemination meeting between £2,700 and £40,500, with a central estimate of £14,400.

---

37 ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

38 22 per cent uplift based on ONS (2020) Index of labour Costs per Hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

39 ONS: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

Monetised costs to the public sector

Table 5: Summary table of monetised costs to the public sector, 10-year total, 2026/27 PV, £ millions

	Low	Central	High
Training costs	0.00	0.01	0.01
Application costs	0.23	0.81	1.84
Monitoring and oversight costs	0.02	0.06	0.18
Total costs	**0.25**	**0.87**	**2.04**

Source: Home Office internal estimates.

59. The total costs to the public sector are estimated between £0.25 million and £2.04 million, with a central estimate of £0.87 million. The following section details the methodology underpinning each cost line.

Training costs

60. Training costs capture the opportunity cost of time whilst Home Office train themselves to manage the authorisation process. Training costs are calculated as:

Average hourly wage x hours taken to complete training x volume of staff completing training

61. To calculate the unit cost, the wage rate of one Senior Executive Officer (SEO) and one Grade 7 (G7) who are expected to complete the training are modelled. Based on the likely scheme design as advised by policy experts, the training is expected to take between one and two weeks, with a central estimate of 1.5 weeks. The training is a one-off cost. In the central and low estimates, the training is applicable to one SEO and one G7. In the high scenario training costs are multiplied by the two SEOs and two G7s which would be necessary to meet higher capacity demands.

62. This IA estimates the training cost between £3,600 and £14,600, with a central estimate of £5,400.

Application costs: Application form

63. Based on the likely scheme design as advised by policy experts, application forms will be processed by one SEO and one G7. The decision reached will be reviewed by one Deputy Director (DD). The SEO and G7 will require training to provide the detailed review required, but it is assumed that the DD will not require specific training to review overarching decisions. The cost associated with this FTE:

Hourly wage at grade x hours taken to process application at grade x volume of applications

64. Policy experts have advised it is likely to take between 21 and 42 hours of SEO time to process each application, with a central estimate of 32 hours. An additional one to two hours, with a central estimate of 1.5 hours is expected to be needed for administrative tasks[40]. The estimated opportunity cost of the time spent on administrative tasks is between £40 and £80, with a central estimate of £60.

[40] The cost of the time spent on administrative activities is expected to be recouped by the administrative fees that entities pay to the Home Office as part of their applications.

65. Based on current operational understanding from policy experts, it is expected that between 15 and 30 hours of G7 time will be required to process each application, with a central estimate of 22 processing hours.

66. Additionally, between two to four hours, with a central estimate of three hours, is allocated for a DD to process the application form.

67. Combining the time cost for SEOs, G7s, and DDs, and multiplying by the anticipated volume of applications, this IA estimates the total cost of processing application forms to be between £224,000 and £1,667,900, with a central estimate of £750,600.

Application costs: Compliance visits

68. Before authorisation is granted entities may be subject to a compliance visit from Home Office inspectors. Based on current design of the scheme, this compliance visit shall be conducted by one SEO or G7 from the Home Office, one member of AISI and one member of law enforcement at a rank of inspector or superintendent. The cost associated with this FTE is calculated as:

Average hourly wage x hours taken for compliance visit x volume of compliance visits x volume of people conducting compliance visit

69. Internal Home Office data[41] is used to calculate the wage rate of civil servants. The time taken for a compliance visit is estimated between three and four hours, with a central estimate of 3.5 hours.

70. A range is used to capture uncertainty around the volume of compliance visits. In the first and second year of the appraisal period, all entities will require a compliance check as the Home Office seeks to establish its understanding of the sector. Similarly, across the entire appraisal period, all entities will require a compliance check at the point of their first application for safety and compliance reasons. Only 50 per cent of repeat applications (excluding applications in the first two years of the scheme) will require a compliance check in the central estimate. In the high estimate, all repeat applicants will require an additional compliance check, and in the low estimate no repeat applicants will require an additional compliance check.

71. This IA estimates the FTE cost of compliance visits between £5,100 and £172,200, with a central estimate of £55,000.

Monitoring and oversight costs: Review of annual accreditation report and mandatory evaluation report

72. Home Office staff will be required to review the annual accreditation report and a mandatory evaluation report produced by authorised entities to gain an operational understanding of the impact of the technology testing defence. The cost associated with this FTE is estimated as:

Hourly wage rate at grade x average hours taken to review reports at grade x volume of reports x volume of staff reviewing reports

73. Internal Home Office data[42] is used to calculate the wage rate of civil servants; one SEO, one G7, and one DD are expected to review the reports. Based on current

[41] The wage rates are modelled using an internal Home Office staff cost model for staff across different grades and police ranks and includes non-wage labour costs.

[42] The wage rates are modelled using an internal Home Office staff cost model for staff across different grades and police ranks and includes non-wage labour costs.

understanding of the design of the scheme from policy experts, it is estimated that the paperwork will be between 5 to 10 pages, with a central estimate of 7.5 pages. The central reading speed estimate is 300 words per minute with a range of 200 to 700 words per minute[43]. A comprehension multiplier between 85 per cent and 60 per cent[44], with a central estimate of 80 per cent, is applied. This is multiplied by the volume of entities which are expected to be authorised[45].

74. This IA estimates the cost associated with report reviewing between £1,100 and £28,500, with a central estimate of £7,200.

Monitoring and oversight costs: Spot checks

75. Authorised entities will be subject to Home Office spot checks to ensure compliance with the conditions of the technology testing defence.

Hourly wage rate at grade x hours taken for spot check at grade x volume of spot checks x volume of people conducting spot checks

76. Internal Home Office data[46] is used to calculate the wage rate of civil servants and equivalent law enforcement ranks; spot checks will be conducted by one SEO or G7 from the Home Office, one G7 from AISI and one member of law enforcement at a rank of inspector or superintendent. Based on the current design of the scheme, policy experts have advised that spot checks are planned to take between three to four hours, with a central estimate of 3.5 hours.

77. A range has been developed to capture uncertainty around the volume of spot checks; the low estimate assumes each entity will receive a spot check every three years, the central estimate assumes each entity will receive a spot check every two years, and the high estimate assumes each entity will receive a spot check every year. Spot checks are estimated to cost between £18,600 and £152,700, with a central estimate of £49,700.

Non-monetised costs

78. Non-monetised costs include the cost of creating the required conditions for testing including a secure physical environment and audited computer access, administrative costs including ensuring General Data Protection Regulation compliance and insurance costs, and potential costs for the low-probability scenario that the conditions of the technology testing defence were breached.

79. The costs associated with creating the required conditions for testing and administrative costs are non-monetised due to data limitations. In both cases, a sense of scale cannot be established without further market engagement as costs will vary significantly depending on existing provisions and systems. On the one hand, the creation of a physical environment may be costly. On the other hand, the large firms expected to apply for the technology testing defence may already have the facilities and capabilities to meet this requirement in place.

---

43 Words per minute estimates for screen reading, Readingsoft: http://www.readingsoft.com/

44 Comprehension rates, Readingsoft: http://www.readingsoft.com/

45 Across the 10-year appraisal period between 130 and 270 entities are expected to be authorised, with a central estimate of 200 entities.

46 The wage rates are modelled using an internal Home Office staff cost model for staff across different grades and police ranks and includes non-wage labour costs.

80. In the event of a breach significant costs may be incurred. The likelihood of this is low as there will be strong deterrents including the threat of prosecution and rigorous safeguarding measures. Additionally, there will be a small number of authorised entities, which reduces the risk further.

81. In the event of a breach there will be reputational damage to the Home Office and the offending entity. Although the power to create an offence is intended to deter non-compliance with conditions and safeguards, it is possible that criminal proceedings will be required which will have an impact on the CJS. It is expected that any CJS impact is small as there will be few authorised entities and a high rate of compliance.

82. It has not been possible to estimate the travel costs for civil servants and law enforcement to travel to authorised entities to conduct compliance visits and spot checks. These costs would capture the opportunity cost of time whilst traveling and the cost of transport to travel to and from the site visit. This is due to uncertainty around the base location of civil servants and law enforcement and the site location of the entities. Costs could, therefore, vary significantly.

Monetised Benefits

83. It has not been possible to monetise any benefits associated with technology testing defence due to data limitations and uncertainty around how directly the measure will translate into reduction of real-world harms. Benefits associated with the introduction of the technology testing defence are covered in the non-monetised benefits sections.

Non-monetised benefits

84. Non-monetised benefits are expected to include the following:

- Reduced legal risk to entities.
- Reduction in technologically generated harmful content.
- Improved safety, compliance and oversight of the technology industry.
- Enhanced public trust in technology and government governance frameworks.

85. The AI testing defence allows authorised entities to safely and lawfully test AI models for their potential to generate illegal or harmful content, which reduces the legal risk for the entities. This could lead to stronger industry standards for internal safeguards, whistleblowing processes, and operational controls. Testing that internal safeguards are operating successfully could lead to reduction in AI-generated harmful content leading to reduced societal harms associated with CSAM, EP and NCII.

86. This intervention aims to provide a clear legal framework for AI testing involving high-risk harms (CSAM, EP and NCII) and reduce ambiguity for technology testers and the public. This could lead to enhanced public trust in AI and government governance frameworks and increased confidence that AI systems that are deployed in the UK meet high safety standards, are tested responsibly, and that harmful capabilities are identified before they cause real-world harm.

Impact on small and micro-businesses

87. It is unlikely that small and micro businesses will apply for the AI testing defence which means there will be little or no impact upon these businesses because the defence is optional and tightly controlled, that is only organisations that apply for authorisation and can meet stringent security/governance requirements are likely to use it, and

small/micro businesses are unlikely to apply or qualify. It is possible that small and micro businesses could choose to apply for the testing defence in which case they must adhere to strict eligibility and safety requirements, including secure data storage and compliance reporting. Those without sufficient in-house capability may be able to participate by partnering with larger organisations to meet the required standards[47].

C. Costs and benefits to business calculations

Business Net Present Value (BNPV)

88. This IA estimates the monetised opportunity costs associated with the AI testing defence and pass-through costs in the form of fees (for the AI testing defence and DBS checks). All benefits are non-monetised, with the main benefit being the reduced legal risk and ability to test AI models. The BNPV is negative[48] and estimated between -£0.14 million and -£0.72 million, with a central estimate of -£0.32 million.

89. The policy is an optional defence and only interested entities will apply. This reduces the likelihood of adverse distributional impacts. It is anticipated that only the AI sector will be affected, and it is expected that large businesses will be most likely to apply for the defence. On the other hand, this could mean that small and micro-businesses are less likely to benefit from the policy.

Costs to business

90. There are two types of cost to business; opportunity costs of time spent on various activities around the AI testing defence and fees which represent a pass-through cost. Section E details the methodology and estimates for each of these costs.

91. In summary, the opportunity costs include familiarisation costs, time taken to fill out the application form, time taken during compliance visits, time taken to produce paperwork, time taken during spot checks, and time taken for a dissemination meeting. This IA estimates the total opportunity cost between £0.1 million and £0.7 million, with a central estimate of £0.3 million.

92. The fees include DBS checks and AI testing defence fees (paid to the Home Office). This IA estimates the total pass-through cost between £19,600 and £149,300, with a central estimate of £66,600.

Benefits to business

93. All benefits to business are non-monetised. The main benefit of the technology testing defence is the reduced legal risk associated with testing the ability to generate illegal or harmful content, particularly CSAM, EP, and NCII. Secondarily, enhanced public trust in AI companies may benefit the reputation of businesses within this sector.

D. Costs and benefits to households' calculations

94. The intervention is expected to improve the safety of AI models which have been tested to verify that safeguards work. Functioning safeguards could lead to a reduction in the

[47] The Crime and Policing bill wording allows for third party testing, but the practicalities are still being finalised. There is no direct evidence of this type of testing yet.

[48] The benefits to business offset and may outweigh the costs but are non-monetised, so the costs to business drive a negative BNPV.

production of AI-enabled illegal content which implies a reduction in the societal harms associated with this material. It has not been possible to monetise the benefits to households of this improvement in safety due to an absence of data and evidence on the scale of the impacts.

95. Non-monetised benefits include a reduction of the societal harms associated with illegal content, particularly the emotional harm associated with AI-enabled abuse. Victims of CSAM, EP, NCII, who are primarily women and children, are expected to see reduced abuse and better safeguarding.

E. Business environment

96. The intervention is expected to strengthen the UK's attractiveness as a destination for AI investment by providing a clear and lawful route for responsible model testing. This reduces legal uncertainty for developers and assurance providers, thereby supporting long-term confidence in the UK as a safe and innovation-enabling market.

97. By removing the current legal barriers to conducting essential safety testing, the intervention improves incentives for innovation: firms will be able to test models more thoroughly, invest in safer product development, and bring higher-quality AI systems to market. This supports wider industry behaviour that prioritises safety-by-design rather than risk-avoidance caused by legal ambiguity.

98. The intervention is also expected to positively affect competitive dynamics. Clear regulatory permissions, combined with defined eligibility and safety requirements, enable small and micro businesses to participate where they meet the necessary standards, including through partnerships with larger or third-party organisations where in-house capability is limited. This helps reduce barriers to entry, supports market diversity, and mitigates the risk of market concentration around larger firms with greater legal risk-management capacity.

99. Overall, the policy is expected to enable responsible innovation, reduce legal and operational barriers, and improve business confidence, supporting a competitive and attractive UK AI ecosystem.

F. Trade implications

100. The proposed intervention is not expected to create significant negative impacts on international trade or investment and is more likely to enhance the UK's position as a trusted market for safe AI development by reducing legal uncertainty for firms operating globally. By establishing a lawful route for authorised AI safety testing, the measure supports investor confidence and aligns with emerging international expectations around responsible AI assurance.

101. There are no anticipated discriminatory requirements placed on foreign businesses compared with UK based firms. The measure does not introduce market access restrictions or barriers that would inhibit firms exporting AI models or assurance services, nor does it create additional regulatory burdens that differ materially for overseas suppliers.

102. If international partners take different approaches to AI safety testing regulation, there may be some divergence; however, such divergence is unlikely to create material obstacles for trade. The proposal instead provides clarity and predictability in the UK

regulatory environment, which may strengthen the UK's competitiveness as a location for AI model development and evaluation.

103. Overall, the intervention is expected to support trade and investment, reinforce the UK's reputation for safe and responsible AI governance, and avoid creating new technical barriers to trade.

G. Environment: Natural capital impact and decarbonisation

104. There is no expected environmental impact as part of this intervention.

H. Other wider impacts (consider the impacts of your proposals)

105. The intervention may support wider national security and systemic risk reduction by improving the safety of AI systems. Equalities considerations will continue to be monitored, though no disproportionate impacts on protected or vulnerable groups have been identified to date. No additional major public sector impacts have been identified beyond those already captured elsewhere.

I. Risks and assumptions

106. Broadly, risks and assumptions are driven by three limitations; analytical uncertainty around the assumptions used to estimate costs, Home Office delivery of secondary legislation, and broad uncertainty around how the market will react.

Analytical uncertainty

107. Throughout the cost modelling, opportunity costs of time to entities are estimated using the annual median hourly wage of Quality and Regulatory Professionals[49] (with a 22 per cent[50] uplift applied to capture non-wage labour costs). Excluding the uplift, this is £24.49 per hour. This wage is treated as the best estimate as the category best captures the activities of the would-be testers but may be an underestimate as rapid growth in the AI sector could be driving higher wages.

108. There is a risk in the assumption that there will be 100 per cent compliance with the defence meaning no costs associated with a breach of the conditions has been quantified.

109. Additionally, due to data limitations, some costs and all benefits are non-monetised which is detailed in Section E. The absence of monetised benefits means that the assessments of value including Net Present Social Value (NPSV) and Business Net Present Value (BNPV) do not reflect the potential benefits of the policy (or the magnitude of these benefits).

110. The rapid growth of the AI sector increases the risk that technology tools may evolve or scale faster than enforcement and regulatory responses. Current volume estimates may underestimate growth within the sector which would have a knock-on impact for all costs. Through a strategic lens, technology may be adapted to evade the new offence which would reduce the benefits associated with improved safety. As a result, the costs,

---

[49] ONS-: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/occupation4digitsoc2010ashetable14

[50] 22 per cent uplift based on ONS (2020) Index of labour Costs per Hour, seasonally adjusted: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/indexoflabourcostsperhourilchseasonallyadjusted

benefits and deterrent impact of the offence remain sensitive to technological change and behavioural responses in a fast-moving market.

Uncertainty in delivery

111. There are uncertainties from the Home Office side until secondary legislation is laid. This affects both the costs to the public sector and the costs to entities.

112. There is uncertainty around the length of the documents which will accompany the AI testing defence. The administrative processes remain under development. The length of the guidance, application form, accreditation and mandatory evaluation form are all based on existing best-estimates and may be subject to change. This will affect familiarisation costs and opportunity costs to entities[51].

113. Given the novel nature of this defence, there is also uncertainty around the time needed for compliance checks and spot checks which affects the opportunity cost of time faced by entities and the FTE cost to public bodies.

114. An important risk is the low-probability scenario of the rules of testing defence being breached. Given the severity of harms associated with AI-generated illegal content, and the strict conditions attached to authorisation, we expect full compliance to all requirements from authorised persons. Any breach of conditions could result in loss of authorisation, criminal liability, and significant legal risk to the organisation which is deemed to be an unacceptable risk. In the case of the breach there would be significant costs including societal harms associated with the production of illegal content material, reputational damage to the Home Office and the affected entity, a potential CJS impact, and potential legal challenges to both the Home Office and the affected entity.

Market uncertainty

115. The number of entities which intend to apply for authorisation for the technology testing defence remains unknown, which will affect the familiarisation costs and Home Office application costs. A range has been developed to capture this uncertainty.

116. Additionally, the number of people in a testing team (5 to 10, with a central estimate of 7.5) is based on policy experts current understanding but has not been tested operationally. This affects the cost of fees paid for DBS checks and the opportunity cost of time taken during a dissemination meeting.

[51] The size of the fee paid to the Home Office is also uncertain, but this has a smaller impact than the other assumptions and is not considered a key risk.

Annex

Mandatory specific impact test - Statutory Equalities Duties	Complete
Statutory Equalities Duties The power to create a statutory defence will neither directly or indirectly discriminate against a person or persons with protected characteristics or protected groups. The Home Office do not consider this provision will have any significant impacts on the specific considerations for limbs (b) and (c) of section149(1) of the Equality Act 2010: the policy is neutral in this regard. Further assessment against the public sector equality duty will be conducted ahead of making regulations on appropriate safeguards to identify and address any disproportionate impacts to protected groups **The SRO has agreed these summary findings.**	**Yes**