Campaign for AI Safety

# Consultation: UK Product Safety Review

**30 November 2023**

This submission is made by the Campaign for AI Safety in response to consultation on the UK Product Safety Review. We trust this submission is of assistance.

The Campaign for AI Safety is a not-for-profit association established in Australia with members worldwide. We are concerned about the dangers AI poses to people and advocate for a stop on the advancement of certain AI capabilities. We also advocate for regulation that promotes and mandates ethical AI. We are not affiliated with any political group. Please visit campaignforaisafety.org for more information.

**Question 23.To inform consideration of whether the civil product liability regime remains fit for purpose, can you provide any examples where the current product liability regime:**
**a) is unclear because of technological developments (e.g., lack of clarity about who is responsible for safety of an AI/smart product or when software is updated); or**
**b) doesn't enable consumers to seek fair redress; or**
**c) doesn't provide businesses with clarity and confidence to develop new products?**

There is a wide range of AI technologies and applications. We are most concerned with AI technology that has a high generality of capabilities and applications (e.g. models at or above the level of OpenAI GPT-3 or GPT-4 series of models). We refer to the UK Government's recently released 'Safety and Security Risks of Generative Artificial Intelligence to 2025' report[1] for a detailed coverage of the emergence of dangerous capabilities which could have large-scale disastrous consequences if highly advanced AI systems are misused or go wrong.

## Examples

The above societal risks outlined above aside, AI applications are causing harm. In 2018, a pedestrian walking a bicycle was killed by an Uber self-driving vehicle for which the [2]. Due to a lack of testing and previous data, the AI system did not make the correct conclusion that the pedestrian was a human before it was too late to avoid the collusion. In 2022, a local community group raised concerns about the accuracy of Toronto's new AI-enabled tool for measuring beach water quality when it found conflicting results using traditional methods[3]. We have also heard of cases where AI-enabled decision making led to longer criminal sentences for minority groups due to biases in risk assessments in the AI system[4] and a chabot used by an eating disorders helpline charity that provided harmful advice to people[5]. The OPSS might find the AI Incident Database[6] useful for real-life examples of harm caused by AI consumer products.

## Gaps identified in the current product liability regime

In terms of a) protecting the rights of those harmed by AI systems and b) encouraging AI developers to make their products safer, the UK product liability regime is not fit for purpose for the following reasons:

- It is **challenging for victims to meet the burden of proof** and other legal conditions under the current regime to establish liability and gain compensation for damage caused by unexplainable and self-learning AI applications

- There is a **lack of clarity about liability concepts and definitions** in the context of AI which creates uncertainty for parties in the AI supply chain about their

[1] Safety and Security Risks of Generative Artificial Intelligence to 2025 - GOV.UK
[2] Uber's self-driving operator charged over fatal crash
[3] Toronto's new tool for testing beach water quality under fire
[4] "Machine Bias", Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
[5] "US eating disorder helpline takes down AI chatbot over harmful advice", Lauren Aratani, The Guardian
[6] AI Incident Database is a crowdsourced database of incident reports of AI systems causing safety, fairness, quality and other problems.

responsibilities and delays or discourages consumers from seeking redress (e.g. ‘product’, ‘producer’, ‘defect’, etc in the Consumer Protection Act 1987)[7].

- It is **not right** that consumers of AI do not have the same level of protection as in cases not involving AI systems.

## Recommendations to address identified gaps

We propose the following adjustments to the product liability regime to sufficiently address the harms caused by AI systems.

1. **AI developers have a presumed duty to end users** and non-contracting third parties, for which AI developers can rebut if the harm caused is too remote to have been foreseeable. This then provides to the general public (i.e. non-contracting third parties) better legal protection in the event an AI application causes them harm.
    a. To go further, there should be a rebuttable presumption against an AI developer, of a causal link between a failed duty of care and harm caused by the AI system once a breach of a duty of care is established. This overcomes the lack of explainability regrading the workings of AI models and lack of transparency[8].
2. A **risk-based approach** so only significant harm is targeted which should include economic loss, personal injury and death, and significant immaterial harms such as impact on mental health, and discrimination. This would balance consumer safety while fostering innovation, and not result in a deluge of frivolous lawsuits against AI labs.
3. **Any terms that exclude liability within user-agreements** for AI-related software and applications, or which excludes an individual’s right to participate in class actions, are to be deemed **unfair and voided**.
    a. Further to this, we encourage the OPSS to investigate whether unfair contract terms within AI-user agreements are precluding access to justice (e.g. terms that exclude liability) for end users, consumers and small businesses. This may or may not already be covered by the unfair trading regulations.

Additionally, we support all the recommendations detailed in the Future of Life Institute’s position paper on AI liability[9] and put forward to the OPSS for consideration.

It may be that for some AI/smart products, the current framework is applicable and sufficient. If that is the case, there must be real or stronger enforcement of the existing laws that already apply such as through increasing the OPSS’ powers and imposing larger penalties.

---

[7] Study on the Impact of Artificial Intelligence on Product Safety - GOV.UK
[8] Shine Lawyers submission to Inquiry into Supporting Responsible AI in Australia
[9] FLI Position Paper on AI Liability

## Recommendations to address AI risks in consumer products

### Standards and data sheets

We support the recommendation in a 2017 report (commissioned by the UK Government into how to grow the UK AI industry) to develop standards for explaining decisions, processes, and services enabled by AI, to improve transparency and accountability[10]. There has been some progress since such as BS 8611 Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems[11].

To provide more transparency about how AI systems work and minimise usage in ill-suited contexts, we recommend **requiring AI developers to provide documentation or 'model cards' about its released models** that detail:

- performance characteristics
- training data (comprehensive references, not vague descriptions)
- the context in which models are intended to be used
- performance evaluation procedures.

We find the set of sections proposed in detail for model cards by Margaret Mitchell et al.[12] to be a useful guide for developing standards. We feel this need is especially important for complex, general purpose AI systems where the risk of harm is highest. We agree with the examples of 'high risk' AI set out in the draft EU AI Act[13].

We believe the guidance set by standards will guide innovation towards safer products as has historically been the case with product safety legislation in other industries[14].

### Disclosure of AI-generated content

We support mandating disclosure of any AI-generated content[15] published online, in textbooks, or in mass media, e.g. labels at the bottom parts of images. An exception should be made for computer-generated imagery in feature films and clearly recognisable animation as this would affect viewing pleasure. We believe that the obligation should primarily fall on the publishers of content (including online platforms and their users).

For audiovisual content, AI companies can be required to embed watermarks and provide means to detect or check AI-generated content. But because such technologies are likely to be circumventable[16] and will work poorly for AI-generated text[17], they should not be relied upon to be the primary enforcement mechanism.

---

[10] Growing the artificial intelligence industry in the uk - gov.uk see page 69
[11] BS 8611 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems
[12] "Model Cards for Model Reporting", Margaret Mitchell, et al. (14 Jan 2019): Figure 1, Summary of model card sections and suggested prompts for each (Page 3).
[13] "Proposal: Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts", European Commission (2021).
[14] https://media.nesta.org.uk/documents/the_impact_of_regulation_on_innovation.pdf
[15] Big tech pledges to watermark AI content to boost safety - Verdict
[16] "AI watermark remover tool lets users remove watermarks with a single click", Australian Photography (30 January 2023).
[17] "We pitted ChatGPT against tools for detecting AI-written text, and the results are troubling", Armin Alimardani, Emma A. Jane (20 February 2023).

## Enforcement and compliance

The supply of complex and highest risk AI systems and applications are dominated by companies with deep financial resources (Microsoft, Google, Meta, Anthropic). Their systems are already being used by large parts of the economy and the impacts are spread across many people and livelihoods.

To provide adequate incentive to these large companies and avoid penalties being seen as a cost of business, we propose **tough penalties for non-compliance** (such as percentage-of-worldwide-turnover fines and criminal penalties against corporations, their employees and directors, similar to sanctions under EU GDPR[18]).

As AI evolves rapidly and is being used across the economy, we are concerned the harms will become long-term, structural, and spread across many people. We have seen this unfold with digital platforms and digital markets run by a handful of the most powerful tech firms.

We therefore urge the UK Government to **make the necessary reforms now and not to leave AI liability to be handled through court systems** as it will take many years to develop a body of case law and statutory law. Or to rely on self-regulatory or voluntary measures that conflict with the firm's business and profit goals.

[18] Examples of fines issued for breaches of the *EU GDPR* legislation: GDPR Enforcement Tracker, CMS Hasche Sigle Partnerschaft von Rechtsanwälten und Steuerberatern mbB.