Information Commissioner's Office response to the consultation series on generative AI

1.0.1 | 18 December 2024

Information Commissioner's Office response to the consultation series on generative AI

Background

As part of the ICO's work on artificial intelligence (AI) regulation, we have been quick to respond to emerging developments in generative AI, engaging with generative AI developers, adopters and affected stakeholders. In April 2023, we set out questions that developers and deployers needed to ask[1]. In January 2024, we launched our five-part generative AI consultation series[2], which this consultation response summarises.

The series set out to address regulatory uncertainties about how specific aspects of the UK General Data Protection Regulations (UK GDPR) and the Data Protection Act (DPA) 2018 apply to the development and use of generative AI. It did that by setting out our initial analysis of these areas, along with the positions we wanted to consult on.

What follows is a summary of the key themes that emerged from the responses to the consultation. This summary is not intended to be a comprehensive record of all the views expressed, nor a response to all individual points raised by respondents.

In addition to summarising feedback, this consultation response sets out our analysis on how specific areas of data protection law apply to generative AI systems. This response does not cover the entirety of our regulatory expectations (these are covered in more detail in our core guidance on AI and data protection)[3]. This response does provide clear views on the application of data protection law to specific issues. In due course, we will be updating our existing guidance to reflect the positions detailed in this response.

The response also flags areas where we think further work is needed to develop and inform our thinking. We also recognise that the upcoming data reform legislation[4] may have an impact on the positions set out in this paper.

1 Generative AI: eight questions that developers and users need to ask

2 ICO consultation series on generative AI and data protection

3 Artificial intelligence

4 See the Data (Use and Access) (DUA) Bill

Executive summary

Our consultation series on generative AI and data protection covered five key areas:

- The lawful basis for web scraping to train generative AI models.[5]
- Purpose limitation in the generative AI lifecycle.[6]
- Accuracy of training data and model outputs.[7]
- Engineering individual rights into generative AI models.[8]
- Allocating controllership across the generative AI supply chain.[9]

In total, we received 192 responses from organisations and 22 from members of the public. The majority of the responses came from the creative industries, trade or membership bodies, the technology sector (including 'big tech' firms) and law firms. In addition, we also held roundtable sessions with civil society, creative industries and technology firms.

Many of the responses were highly detailed. They often combined evidence (eg technical explanations of generative AI), analysis (eg interpretations of the application of data protection law) and arguments (eg the negative or positive effects of generative AI on a particular stakeholder).

We are grateful to those who responded to the consultation. We are also grateful to those who were willing to discuss these issues in further detail with us. In particular, thank you to the British Screen Forum, the Ada Lovelace Institute and TechUK for facilitating roundtable discussions with the creative sector, civil society and the technology sector respectively. We also thank colleagues at the French data protection authority, the CNIL, for sharing their insights on this issue.

Summary of positions after consultation

The following points set out our positions after reviewing the consultation responses:

We retained our position on purpose limitation,[10] accuracy[11] and controllership.[12]

We updated our position on the legitimate interests lawful basis for web scraping to train generative AI models.[13]

- We heard that data collection methods other than web scraping exist, which could potentially support the development of generative AI. For example, where publishers collect personal data directly from people and license this data in a transparent way. It is for developers to demonstrate the necessity of web scraping to develop generative AI. We will continue to engage with developers and generative AI researchers on the extent to which they can develop generative AI models without using web-scraped data.
- Web scraping is a large-scale processing activity that often occurs without people being aware of it. This sort of invisible processing poses particular risks to people's rights and freedoms. For example, if someone doesn't know their data has been processed, they can't exercise their information rights. We received minimal evidence on the availability of mitigation measures to address this risk. This means that, in practice, generative AI developers may struggle to demonstrate how their processing meets the requirements of the legitimate interests balancing test. As a first step, we expect generative AI developers to significantly improve their approach to transparency. For example, they could consider what measures they can provide to protect people's rights, freedoms and interests. This could involve providing accessible and specific information that enables people and publishers to understand what personal data the developer has collected. We also expect them to test and review these measures.
- We received evidence that some developers are using licences and Terms of Use (ToU) to ensure deployers are using their models in a compliant way. However, to provide this assurance, developers will need to demonstrate that these documents and agreements contain effective data protection requirements, and that these requirements are met.

We updated our position on engineering individual rights into generative AI models, as set out in the consultation chapter four.[14]

- Organisations acting as controllers must design and build systems that implement the data protection principles effectively and integrate necessary safeguards into the processing. This would put organisations in a better place to comply with the requirement to facilitate people's information rights.
- Article 11 (on processing which does not require identification) may have some relevance in the context of generative AI. However, organisations relying on it need to demonstrate that their reliance is

appropriate and justified. For example, they must demonstrate they are not able to identify people. They must also give people the opportunity to provide more information to enable identification.

5 Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models

6 Generative AI second call for evidence: Purpose limitation in the generative AI lifecycle

7 Generative AI third call for evidence: accuracy of training data and model outputs

8 Generative AI fourth call for evidence: engineering individual rights into generative AI models

9 Generative AI fourth call for evidence: engineering individual rights into generative AI models

10 Generative AI second call for evidence: Purpose limitation in the generative AI lifecycle

11 Generative AI third call for evidence: accuracy of training data and model outputs

12 Generative AI fifth call for evidence: allocating controllership across the generative AI supply chain

13 Generative AI first call for evidence: The lawful basis for web scraping to train generative AI model

14 Generative AI fourth call for evidence: engineering individual rights into generative AI models

Context

Generative AI is a type of AI system that can generate new text, image, audio and video content. Generative AI models are typically trained on vast amounts of personal data. In a generative AI context, this training is called 'pre-training' because the core model is often fine-tuned with additional training data that can also contain personal data.

Generative AI gives rise to new and challenging data protection questions. It involves vast amounts of personal data, often processed invisibly and across a complex supply chain. It also poses risks to people's rights and freedoms, particularly when they are not aware that the processing is happening. The pace of technological development means that technical aspects of generative AI models may not always be designed with data protection in mind, making it challenging for organisations to demonstrate compliance with the law.

At the same time, the principles of data protection apply wherever personal data is processed – including with new or novel technologies. We continue to monitor major developers of generative AI, who need to ensure that they demonstrate ongoing compliance when developing and deploying new technology.

Methodology

Our methodology was underpinned by the principles set out in our Policy Methodology[15] and consultation policy.[16]. We have a documented, consistent, and transparent approach to policy making where we deliver evidence-based decisions that focus on achieving clearly stated regulatory outcomes.

The public consultation received responses via both survey responses as well as written responses submitted via email. We received response from 192 organisations and 22 members of the public.

The open text survey responses and email responses were systematically reviewed and analysed to identify key points across the responses. We also held three roundtables, held a workshop with counterparts at the CNIL, and engaged a wider range of stakeholders on particular issues.

The ICO is committed to making timely, informed and impactful decisions, drawing on evidence and insight, as well as understanding the impacts of our interventions to ensure that we are making a material positive difference.[17] Impact considerations formed part of our public consultation. A summary of the impact related responses is provided in the Annex.

15 ICO - Policy Methodology

16 ICO - Consultation-policy

17 Our impact approach ensures that in making a decision to intervene our actions are both proportionate to the issue at hand and not unduly burdensome on the those that we regulate. For more information, see: Measuring our impact

Tackling misconceptions

This section clarifies several data protection issues to address misconceptions highlighted in the consultation responses. Some of them are about generative AI, while others are about AI (including 'narrow' AI) or data protection in general. We have set out existing ICO positions to offer clarity.

1) The "incidental" or "agnostic" processing of personal data still constitutes processing of personal data. Many generative AI developers claimed they did not intend to process personal data and that their processing of that data was purely incidental. Our view is clear: data protection law applies to processing of personal data (which includes special category data), regardless of whether this is 'incidental' or unintentional.

2) Common practice does not equate to meeting people's reasonable expectations. Organisations should not assume that a certain way of processing will be within people's reasonable expectations, just because it is seen as "common practice". This applies particularly when it comes to the novel use of personal data to train generative AI in an invisible way or years after someone provided it for a different purpose (when their expectations were, by default, different).

3) "Personally identifiable information" (PII) is different to the legal definition of "personal data". Many organisations focus their generative AI compliance efforts around PII. However, to ensure compliance in the UK they should be considering processing of any "personal data" (which is a broader and legally defined concept in the UK). Organisations must not undertake compliance based on a fundamental misunderstanding or miscommunicate their processing operations.

4) Organisations should not assume that they can rely on the outcome of case law about search engine data protection compliance when considering generative AI compliance. A few respondents sought to rely on these case outcomes, arguing that as the initial collection of data was substantively the same (ie crawling the web), the decisions should also apply to the generative AI context. However, while we can see there are similarities in terms of data collection, there are key differences which means that the logic of these decisions may not be applicable. For example, while a search engine intends to index, rank and prioritise information and make this available to the public, generative AI goes beyond this. It synthesises

information and creates something new in its outputs. Traditional search engine operators also enable people to exercise their rights, in particular the right to erasure through 'delisting'. The current practices of generative AI developers make it difficult for people to do the same.

5) Generative AI models can themselves have data protection implications. Some developers argued that their models do not "store" personal data. Our 2020 guidance on AI and data protection stated that AI models can contain personal data.[18] In the generative AI consultation chapters, we explained that generative AI models may embed the data they have been trained on in a form that may allow their retrieval or disclosure by transmission. In particular, this may have implications for open-access models. This needs further research to understand when and how this risk may materialise. We intend to explore this issue in more detail in future, taking account of ongoing developments in the technological and academic spaces.

6) The ICO cannot determine or provide guidance on compliance with legal requirements which are outside the scope of our remit (ie data protection and information law). There was a perception by some respondents that the lawfulness principle[19] meant that we could provide views or guidance on lawfulness under regimes other than data protection. Some also thought that data protection could be a useful lever to address issues within other regulatory remits. To be clear, data processing that is unlawful because it breaches other legal requirements (such as Intellectual Property law) will also be unlawful under data protection law. However, that does not mean we will or can be the arbiter of what is lawful under legislation outside of our remit.

7) There is no 'AI exemption' to data protection law. Some of the respondents argued that data protection law should not complicate generative AI development. While we support responsible AI development and deployment, it is important for organisations to be aware that there are no carve-outs or sweeping exemptions for generative AI. If an organisation is processing personal data, then data protection law will be applicable. We encourage organisations that are uncertain about compliance to adopt a "data protection by design approach", considering compliance issues before they start processing.

18 We said that model inversion and membership inferences show that AI models can inadvertently contain personal data: How should we assess security and data minimisation in AI?

19 Principle (a): Lawfulness, fairness and transparency

The lawful basis for web scraping to train generative AI models

In brief: Our position on the lawfulness of using web-scraped data to train generative AI largely remains the same. See the original call for evidence for the full analysis.

The consultation enabled us to refine it in the following way:

- Legitimate interests remains the sole available lawful basis for training generative AI models using web-scraped personal data based on current practices. However, this is only when the model's developer can ensure they pass the three-part test, including necessity. We received consultation responses which suggested that alternative data collection methods may be feasible. We therefore expect controllers who develop generative AI to evidence why other available methods for data collection are not suitable.
- The three-part test also includes the balancing test. Web scraping for generative AI training is a high-risk, invisible processing activity. Where insufficient transparency measures contribute to people being unable to exercise their rights, generative AI developers are likely to struggle to pass the balancing test.

Respondents

In January 2024, we published the first chapter of our consultation series. This chapter set out our policy position on the lawful basis for processing web-scraped data to train generative AI models.

We received 77 responses from organisations and 16 responses from members of the public. 31 responses came via our survey, with a further 47 received directly via email. The sectors most represented were:

- creative industries (18);
- law firms (nine);
- the technology sector (eight); and
- trade or membership bodies (eight).

Of the survey responses, 19 respondents (61%) agreed with our initial analysis.

Original call for evidence

In our original call for evidence,[20] we set out our positions on the lawful basis for using web-scraped data to train generative AI models. For the full analysis we advise consulting the original call for evidence, but our key positions were as follows.

Firstly, we determined that five (consent, contract, legal obligation, vital interests, public task) of the six available lawful bases are not likely to be applicable in this context. This means that, in practice, legitimate interests (LI) is realistically the only lawful basis developers could explore. However, they still need to pass the three-part test[21] to demonstrate this lawful basis is actually valid.

Secondly, our initial view was that it was likely that most generative AI training would require the volume and kind of data obtained through web scraping – but that we welcomed views on this. If organisations can reasonably achieve their purpose without the high-risk,[22] invisible processing[23] involved in web scraping, then they wouldn't pass the necessity part of the legitimate interest test.

Thirdly, we explained that using web-scraped data to train generative AI is essentially a combination of two high-risk processing activities. This is because it involves innovative technology as well as constituting invisible processing. It therefore hits two triggers on the ICO list of high-risk processing activities.[24]

Finally, we set out various considerations that generative AI developers could explore to potentially help them pass the third part of the legitimate interests test in this context.

Key points from the responses

When it came to identifying a valid legitimate interest ('the purpose test') in processing web-scraped data to train generative AI, the following points arose in the responses:

- Respondents regularly cited either commercial or societal interests, or both, as a legitimate interest. This was particularly the case for AI

developers within the technology sector and some law firms. Several respondents argued that all generative AI is innovative by default and any innovation was innately beneficial for society.
- Some respondents challenged societal interests as a valid legitimate interest. This included representatives from civil society and the creative industries, who highlighted the detrimental impacts generative AI may have on people. These impacts included lack of transparency and the inability to exercise information rights.
- Many respondents from the creative industries argued that web-scraped content would constitute unauthorised and unlawful use of copyright data (specifically under Chapter II of the Copyright, Designs and Patents Act (1988)).[25] They added the processing would therefore also lack a lawful basis under the lawfulness principle of data protection.

In terms of the need to use web-scraped data for training a generative AI model (the 'necessity test'), the following points arose in the responses:

- Generative AI developers and the wider technology sector stated that, because of the quantity of data required, training generative AI models cannot happen without the use of web-scraped data. Similarly, they argued that large datasets with a wide variety of data help ensure the effective performance of models and avoid biases or inaccuracies.
- On the other hand, many respondents, especially from the creative industries, argued that there were alternative ways to collect data to train generative AI models, such as licensing datasets directly from publishers. Therefore, they argued that using web-scraped data couldn't meet the necessity test.
- A variety of responses mentioned synthetic data.[26] However, the extent to which developers can rely on it to provide large amounts of training data remains an open question.

In terms of the impact of the processing on people's interests (the 'balancing test'), the following points arose in the responses:

- Many respondents, especially from civil society and creative industries, argued that in most cases the legitimate interests of generative AI developers would be overridden by people's rights. This is because of the loss of control over personal data that invisible processing involves and that people are unable to understand the impact of that processing on them. This was also raised in both roundtables with these sectors.
- On the other hand, AI developers within the technology sector, among others, argued that the societal benefits of generative AI are a

significant factor in passing the balancing test. They cited innovation and potential beneficial uses.
- On safeguards and mitigations that may help in passing the balancing test, all respondents recognised the key role of transparency. In particular, they mentioned making clear the extent to which personal data is being processed, where this data came from, how it was processed and communicating this clearly to people. One civil society organisation advocated for an AI registry that could perform this.
- They also suggested using licences and terms of use as effective safeguards for generative AI developers, to ensure that "open-access" models used by downstream deployers comply with data protection.

Respondents commented on the processing of special category data and compliance with article 9 of the UK GDPR:

- Civil society respondents, and others, identified the processing of special category data[27] in the training data as a significant concern. This was because of the greater risks to people's rights and freedoms. They argued it was difficult to see how developers could ever meet an article 9 condition[28] when training generative AI.

Our response

Why legitimate interests is the only available lawful basis

Numerous respondents questioned why we determined that legitimate interests was the only valid lawful basis for using web-scraped personal data to train generative AI. For example, some suggested public task as a lawful basis that the public sector could use. The creative industries often raised consent as an option. To help provide clarity, our reasoning for this in the specific context of web-scraping to train generative AI is as follows:

- **Consent:** This is unlikely to apply here because the organisation training the generative AI model has no direct relationship with the person whose data is scraped. In addition, when people first provided their personal data, they could not have anticipated another organisation would later use it for this purpose. People are unlikely to be able to revoke their consent if removing their data requires model re-training, which is currently an extremely cost and time intensive process.[29]

- **Performance of contract:** This is not available because the person (whose personal data is in the training data) does not have a contract with the controller undertaking the web-scraping.[30]
- **Complying with a legal obligation:** The organisations who are training generative AI are under no legal obligation to collect data via web-scraping.
- **Protecting the vital interest of a person:** Training generative AI does not protect a person's life.
- **Public task:** Controllers can only rely on this when the activity is set out in law or is part of their official role. This is not the case with commercial generative AI developers and would be highly unlikely to apply to public sector developers.

When does data protection law apply to creative content?

It is important to clarify that data protection will only apply to creative content if it constitutes personal data.[31] The identifiability of any person will determine whether the content is personal data. This needs to be evaluated on a case-by-case basis, depending on the availability of information[32] and tools[33] to identify the person.

The 'purpose test'

When articulating a legitimate interest in the 'purpose test', our view remains that it is important for controllers to set out a specific and clear interest, even for models that can be used for various downstream purposes. This can help controllers pass the third part of the legitimate interests test, known as the 'balancing test'. Organisations can use interests that are generic, trivial or controversial. However, if they do, they are less likely to pass the balancing test or override someone's right to object.[34]

Controllers should ensure that the specified purposes make it possible to meaningfully assess the necessity of the processing to achieve that purpose. Even when the processing is necessary, they need to ensure that their interest is not overridden by the interests or fundamental rights and freedoms of the person whose data is processed. Generative AI developers should not assume that general societal interests will be sufficient to rely on as a legitimate interest when considering their lawful basis for web-scraping. As we said in the original call for evidence, developers should evidence the

likely benefits rather than assume them.[35] Just because certain generative AI developments are innovative, it does not mean they are automatically beneficial or will carry enough weight to pass the balancing test.[36]

To demonstrate that the chosen approach for achieving a legitimate interest is reasonable, controllers should properly define all of their purposes and justify the use of each type of data collected.

The 'necessity test'

The consultation responses also clearly showed that the necessity of using web-scraped personal data to train generative AI is not a settled issue. The creative industries in particular challenged our initial position that web-scraping is necessary. We received evidence that other methods of data collection exist, for example where publishers collect personal data directly from people and license this data in a transparent manner. As a result, we encourage developers to seek out other sources of data where possible. Where controllers are seeking to evidence that web-scraping is necessary, they should explain why they are unable to use a different source of data.

We will engage further with developers, tech companies, academic researchers and NGOs on the necessity of web-scraping for the purpose of training generative AI.

The 'balancing test'

As part of this consultation, we asked for evidence on the potential technical and organisational safeguards organisations could deploy to mitigate any identified risks. We thought this was especially relevant when considering the legitimate interests balancing test.

Some respondents — particularly from industry — argued that using licenses and ToU for "open-access" models provided a safeguard. They said it helped to mitigate risks that people could otherwise be exposed to from downstream deployment of generative AI models. However, when developers want to rely on licenses and ToU to mitigate risks posed by downstream deployment, they will need to demonstrate that these arrangements contain data protection requirements. They also need to assure these requirements are met, so that the safeguard is effective in practice.

We also received many suggestions for technical safeguards that could be deployed. However on the whole, apart from theoretical or proof-of-concept

suggestions, the consultation responses did not provide us with sufficient verifiable evidence to properly assess the efficacy of these safeguards in practice.

Additionally, we are aware that many controllers are not meeting their basic transparency obligations under article 14 when relying on web-scraped data to develop generative AI models.[37] To demonstrate that the chosen approach for achieving a legitimate interest is reasonable, controllers should properly define all of their purposes and justify the use of each type of data collected. They must express those purposes in a way that allows people to better understand why an organisation is using their data, and, in line with our guidance, what happens with the data in question. Controllers must take into account the modality requirements of article 12, articulating the purpose concisely, transparently, intelligibly and in clear and plain language.

We therefore encourage developers to consider how they can use new and innovative transparency mechanisms and safeguards to enable people to have an enhanced understanding of the data processing and put them in a stronger position to exercise their information rights. This is an area we will continue to monitor, and we strongly encourage generative AI developers to engage with us further on this.

Further, where generative AI model developers are using personal data, they should assess the financial impact on people in the balancing test. For example, a fashion model could lose their income if a generative AI model uses their personal data to create a digital version of them (such as an avatar) to replace them in a fashion show.

Special category data

We did not focus on special category data in our initial consultation. However, many respondents raised the processing of this data for generative AI training as a salient issue. We are currently scrutinising the use of special category data by generative AI developers based on our existing positions.[38]

Finally, as with processing non-special category personal data, whether a controller intends to process special category data remains irrelevant in determining whether that data falls within article 9. The only exception is if an article 9 category could be inferred but there is no intention to make that inference.[39]

20 Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models

21 Legitimate interests

22 Examples of processing 'likely to result in high risk'

23 See glossary.

24 Examples of processing 'likely to result in high risk'

25 Copyright, Designs and Patents Act 1988

26 See glossary and our related guidance: Synthetic data

27 What is special category data?

28 An article 9 condition is necessary in addition to the article 6 lawful basis when someone processes SCD. See: What are the rules on special category data?

29 In the creative industries context, consent is not likely to be valid in the context of generative AI training as it is not practically revokable.
In addition, in this professional context, it can be challenging for consent to be freely given. Any power imbalance between the data subject (the creator) and the controller may render consent unworkable during deployment too. It should also be noted that the concept of 'consent' in data protection is distinct from the concept of consent or 'permission' in other regimes such as copyright.

30 For the creative industries, the contract lawful basis is very unlikely to apply as it is unlikely that an organisation is under a contractual obligation to use a creator's content to train its generative AI.

31 See ICO's guidance on personal data: What is personal information: a guide

32 Such as metadata.

33 For example, facial recognition search engines.

34 What is the 'legitimate interests' basis?

35 Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models

36 For example, generative AI is used to create harmful deepfakes, or can leak personal information in certain contexts.

37 Right to be informed

38 Special category data

39 What is special category data?

Purpose limitation in the generative AI lifecycle

In brief: Our position on purpose limitation in the context of generative AI remains the same. See the original call for evidence for the full analysis.

Respondents

In February 2024, we published the second chapter of our consultation series. This set out our policy position on the interpretation of the purpose limitation principle in the generative AI lifecycle. This built on our existing 2020 position as set out in our core guidance on AI and data protection.[40]

We received 46 responses from organisations, with just one respondent identifying as a member of the public. 16 responses came via our survey, with a further 30 received directly via email. The sectors most represented were:

- creative industries (14);
- trade or membership bodies (eight); and
- the research sector (six).

A general consensus emerged among respondents about the consultation on purpose limitation. 12 (75%) respondents agreed with our overall analysis.

Original call for evidence

In our original call for evidence,[41] we set out our positions on the interpretation of the purpose limitation principle in the generative AI lifecycle. To recap, the key positions were as follows:

Firstly, we determined that the different purposes of training and deploying a generative AI model must be explicit and specific. This is so that controllers, processors and data subjects have a clear understanding of why and how personal data is processed.

Secondly, we explained that developers who are reusing personal data for training generative AI must consider whether the purpose of training a model is compatible with their original purpose of collecting that data. This is called a compatibility assessment.

Thirdly, we explained that developing a generative AI model and developing an application based on such a model (fine-tuned or not) constitute different purposes. These purposes are in addition to the initial separate purpose that an organisation may pursue when collating repositories of web-scraped data.

Key points from the responses

The following key points arose in the consultation responses in terms of defining explicit and specified purposes:

- Respondents, including generative AI developers and some law firms, argued that the 'open-ended' downstream uses of generative AI make it challenging for developers to be explicit about the purpose of the processing at the development stage.
- On the other hand many respondents, including civil society and the creative industries, argued that "developing a model" is too broad a purpose without making the uses of the model clearer. Some civil society respondents argued that it would be impossible to define an explicit and specific purpose at the time of processing data to train a model.
- Most respondents agreed that the purpose of using data to fine-tune a model was more likely to meet the requirements of the purpose limitation principle. This is because the purpose for processing is more explicit and specific and involves less data processing.
- Some downstream AI deployers in the financial sector raised concerns that their power to choose use cases would be limited due to purposes being determined, and limited, at the development stage. They argued this could inhibit innovation and beneficial uses.

Building on this, people highlighted the importance of transparency in complying with the purpose limitation principle:

- Respondents, particularly from the creative industries, argued that clear documentation about the purpose of processing personal data was necessary. This is in addition to the source and context of data collection, the lawful basis, completing a DPIA and demonstrating how they fulfil article 13 and 14 of the UK GDPR.

- Deployers of generative AI also outlined the need for developers to provide guidance on the foreseeable downstream applications of their models, including their capabilities and limitations. This was linked to questions about how developers could monitor the use of 'off-the-shelf' or 'open-access' products and services in practice.
- Trade and membership bodies raised concerns about trade secrets and called for a balanced approach to transparency. For example, revealing too much detail about the purpose of model training could reveal product information to a competitor.
- Respondents also suggested the wider use of contracts and ToU to ensure that the purpose of processing personal data, at each stage of the lifecycle, is being made clear to people involved.

The following key points arose on the reuse of personal data to train generative AI:

- Developers desired a broad interpretation of data protection's purpose limitation principle around the re-use of data to train generative AI, citing greater innovation as a justification.
- But creative industry respondents consistently argued that generative AI developers do not understand the importance of articulating an explicit purpose for the reuse of personal data. They also argued developers did not appreciate that personal data is often embedded in copyright-protected works, which creatives rely on for their remuneration as professionals.
- Numerous respondents, particularly from the creative industries, argued that companies are not meeting people's reasonable expectations when they take and use content containing personal data without permission. This is because people do not expect companies to use this information for free, without payment or acknowledgment.
- Technology trade and membership bodies raised that reusing personal data to train generative AI is crucial to innovation.

The following key points arose about initially developing a generative AI model and then developing an application based on such a model (fine-tuned or not) constituting different purposes:[42]

- Most respondents felt that having separate purposes for development and deployment is necessary to ensure data minimisation and to check that the appropriate lawful basis is in place at each stage.
- A broad range of respondents highlighted that separating purposes supports a risk-based approach. This is because they can offer a

proportionate level of review, assurance and oversight to potential harms at each stage of the lifecycle.
- There was a view from technology trade and membership bodies that a rigid distinction between development and deployment could stifle innovation as, in practice, development and deployment are cyclical (for example, when a model is engaged in continuous learning).

Our response

We welcome the support we received for our initial position on how the purpose limitation principle should be interpreted in the generative AI lifecycle. We are reassured by the positive response to our view that developing a generative AI model and developing an application based on such a model constitute different purposes under data protection.

We recognise that respondents have different views on what qualifies as defining explicit and specified purposes under article 5(1)(b) for training generative AI. A common request from all respondents was for the ICO to develop guidance that includes examples of how developers could demonstrate a sufficiently detailed and specific purpose when training generative AI. We will consider this request for the next iteration of our core guidance on AI, while maintaining our consultation position.

We have consistently said that data protection law recognises the need to protect intellectual property and trade secrets. For example, our guidance on explainability says that providing people with a meaningful explanation about the processing doesn't mean including source code or proprietary algorithms.[43]

Respondents suggested that developers could use contracts or ToU to ensure that they are making the purposes clear to people whose data they're using to train a model, as well as the people whose data is used during deployment, and the ICO. Contracts or ToU could detail the purpose or purposes of the processing and set out expectations on who communicates what to which group. Where parties are relying on contracts or ToU to ensure that purpose or purposes of processing are communicated to individuals and others, they will need to ensure that the requirements in these contracts or ToU are effective. We note that these parties will continue to process data using the legitimate interests lawful basis, even where they have a contract or ToU in place.

40 See: How do we ensure lawfulness in AI?

41 Generative AI second call for evidence: Purpose limitation in the generative AI lifecycle

42 Our 2020 guidance outlines that this is the case: How do we ensure lawfulness in AI?

43 The basics of explaining AI: Benefits and risks

Accuracy of training data and model outputs

In brief: Our position on the application of the data protection principle of accuracy in generative AI remains largely the same. See the original call for evidence for the full analysis.

Respondents

In April 2024, we published the third chapter of our consultation series. This set out our policy position on the accuracy of training data and model outputs.

We received 25 responses from organisations, with just one respondent identifying as a member of the public. Nine responses came via our survey, with a further 17 received directly via email. The most represented sectors were:

- creative industries (eight);
- trade or membership bodies (three); and
- finance (three).

Of the survey respondents, four (44%) agreed with our initial analysis and another four (44%) were unsure. There was a clear consensus about the importance of accuracy. However, respondents disagreed about how far developers or deployers should be primarily responsible for ensuring and communicating accuracy.

Original call for evidence

In our original call for evidence,[44] we set out our positions on the application of the accuracy principle to generative AI. We drew substantially on our existing positions. To recap, the key positions were as follows:

Firstly, we stated that developers will need to know whether their training data contains:

- accurate, factual and up to date information;

- historical information;
- inferences;
- opinions; or
- AI-generated information about people.

In other words, the developer should curate the training data accordingly to ensure sufficient accuracy for the purpose for which it is processed.

Secondly, we determined that the appropriate level of statistical accuracy of the generative AI model is linked to the specific purpose organisations will use the model for. For example, creating or using generative AI models to create non-factual outputs as a source of inspiration will have different accuracy requirements than models whose outputs users rely on as a source of factual information.

Finally, we set out that developers should assess and communicate the risk and impact of incorrect and unexpected outputs. They should also provide clear information about the application's statistical accuracy and its intended use. We provided a list of possible measures, including labelling the outputs as generated by AI (or not factually accurate) and providing information about the output's reliability, for example by using confidence scores.

Key points from the responses

Firstly, in response to our view that developers should ensure training data is made up of accurate, factual and up to date information, the following key points arose:

- Respondents – including a large generative AI developer – stated that it would be impossible to verify whether personal data in training datasets is factually accurate, because there is a lack of 'ground truth'[45] to measure against. They also added that limiting training data to accurate personal data would negatively affect model performance because of a lack of diverse data.
- Many respondents from the creative industries, finance sector and a media organisation all raised concerns that a fundamental lack of transparency about training data by developers means that it is challenging to determine, and question, its accuracy. The creative industries also raised concerns that web-scraped data is highly likely to contain inaccurate personal data.
- The creative industries and a media organisation emphasised the importance of high quality, accurate training data for downstream

purposes where organisations rely on the model as a source of factual information.
- Researchers and academics emphasised the role of independent external audits in ensuring accurate training data and model development.

Secondly, in response to our view that the specific purpose for which a generative AI model will be used is what determines whether the outputs need to be accurate, the following points arose:

- There was general agreement that the degree of accuracy is linked to the specific purpose of a generative AI model. We saw cross-cutting agreement on this from the technology sector, the creative industries and the financial services sector.
- However, creative industry respondents raised specific concerns about the likeness of creatives being generated. They questioned the extent to which creative purposes require less accuracy because it may cause misrepresentation. In contrast, a law firm argued that accuracy was not as important for creative purposes, because the content being generated is not presented as factual.
- Generative AI developers claimed that deployers and end users (ie consumers) were chiefly responsible for the accuracy of outputs, as they solely define the purpose. One industry group highlighted that deployers may also be better placed to understand a model's purpose.
- In contrast, a respondent from the financial services sector argued that, while they agree that developers cannot fully anticipate all of the potential uses, they are still primarily accountable for the quality of the output.

Finally, in terms of assessing and communicating the risk and impact of incorrect and unexpected outputs, as well as providing clear information about statistical accuracy, the following key points arose:

- There was general support for developers to provide clear information about a model's statistical accuracy. One industry and trade body described informing deployers and users of generative AI systems about the statistical accuracy of the model as essential. The technology sector identified examples such as the use of technical reports, model cards and clarifying to users that results may be inaccurate. They also cited retrieval augmented generation (RAG) as a method of ensuring accuracy.

- The creative industries firmly supported measures like labelling and watermarking, including efforts such as embedding metadata into outputs. One law firm stated that these types of measures would be necessary to meet accuracy obligations.
- Another law firm raised concerns that if communication and disclaimers alone are sufficient for compliance, this may incentivise developers to put less resource into ensuring a model's robust statistical accuracy.
- However some respondents, including from the technology sector, argued that providing too much information places unrealistic expectations on developers. They may not anticipate all of the downstream uses of a model or even have a relationship with the end user. Some also raised the limitations of watermarking, such as limited use in text outputs and a lack of durability.
- A number of respondents emphasised the role of technical and organisational measures that developers could deploy, such as monitoring outputs and engaging in content authentication (eg C2PA).[46] One generative AI developer argued that it would not always be feasible for developers to monitor user-generated content, such as analysing inputs or tracking publicly-shared outputs.

Our response

We welcome the overall support we received for our position on accuracy. We are reassured by the support for our position that accuracy is closely linked to a model's specific purpose. This aligns with our established position on the accuracy principle.[47] We appreciate that there is disagreement among respondents about the degree of the statistical accuracy required, meaning how often a model's output is accurate. As we said in our consultation, clear communication between developers, deployers and end-users of models is key to ensuring that the degree of statistical accuracy is proportionate to the model's final application.[48]

In terms of ensuring developers use accurate data to train a model, we accept that there can be limitations to validating ground truth.[49] However, this does not negate the relationship between inaccurate training data and inaccurate model outputs. As we said in our consultation, generative AI developers will need to understand and be transparent about the accuracy of the training data used to train generative AI. Even though accurate training data will not stop generative AI models from hallucinating,[50] it can constrain the margin of error.

Many respondents, including both the creative industries and technology sector, asked us to provide use cases and examples of what would constitute appropriate and sufficient means of assessing and communicating the risk and impact of incorrect and unexpected outputs. They also wanted clear information about statistical accuracy. We will consider this for future guidance updates.

We acknowledge the concerns raised about deepfakes, misinformation and the use of people's likeness. While data protection law can apply to the creation and dissemination of deepfakes in some cases, other legislation such as the Online Safety Act, copyright law and criminal law are also relevant to addressing the harms caused by these technologies.

We also understand that many of the safeguards and measures we initially proposed may have technical limitations or constitute emerging rather than robustly tested practices, such as text watermarking. Unfortunately, we received little evidence to demonstrate a technical measure's viability or lack thereof in practice. We will continue to monitor safeguards and measures and expect generative AI stakeholders and innovators to provide novel solutions to ensure that people can use models in ways that are appropriate to the level of statistical accuracy that the developer knows them to have.

44 Generative AI third call for evidence: accuracy of training data and model outputs

45 In this context, it would mean checking training data against its source and validating that the source of the data was itself accurate (for example, a news article).

46 Coalition for Content Provenance and Authenticity (C2PA)

47 Principle (d): Accuracy

48 Statistical accuracy relates to fairness under data protection. See What do we need to know about accuracy and statistical accuracy?

49 'Ground truth' refers to content that is verifiable by trusted sources.

50 See glossary.

Engineering individual rights into generative AI models

In brief: Our position on controllers' response to people's information rights remains largely the same. See the original call for evidence for the full analysis.

The consultation enabled us to refine it in the following way:

- Organisations must design and build systems that implement the data protection principles effectively and integrate necessary safeguards into the processing. This would, in turn, put organisations in a better place to comply with the requirement to facilitate people's information rights.
- Even though we referenced article 11 in this chapter, organisations should not view it as a way to avoid obligations. Organisations need to demonstrate that any reliance on article 11 is appropriate and justified in the circumstances. They must still give people the opportunity to provide more information to enable the identification of their data.

Respondents

In May 2024, we published the fourth chapter of our consultation series. This set out our policy position on engineering information rights into generative AI models.

We received 28 responses from organisations, with three respondents identifying as a member of the public. Six responses came via our survey, with a further 22 received directly via email. The most represented sectors were:

- creative industries (11);
- industry groups (four); and
- the technology, insurance and civil society sectors (all three).

Of the survey respondents, three (50%) agreed with our initial analysis. There was clear agreement about the importance of information rights.

However, there was both disagreement and a lack of evidence, particularly from technology firms, about how to ensure information rights are engineered into generative AI.

Original call for evidence

In our original call for evidence,[51] we set out several positions on information rights and generative AI. To recap, the key positions were as follows:

Firstly, when generative AI developers and deployers are controllers, they need to show that they have clear and effective processes for enabling people to exercise their rights over their personal data. This applies whether their data is contained in the training, fine-tuning or output data, but also the model itself.

Secondly, developers and deployers need to evidence how they are making sure people have meaningful, concise and easily accessible information about the use of their personal data.

Thirdly, developers and deployers need to clearly justify the use of any exemptions and demonstrate how they are safeguarding people's interests, rights and freedoms.

Finally, we covered article 11. We said that if developers argue they cannot respond to requests because they cannot identify individuals (in the training data or anywhere else), the law requires them to explain this to the requester and demonstrate why this is the case.

Key points from the responses

In terms of generative AI developers and deployers needing to show they have a clear and effective process for enabling people to exercise their rights, the following points arose in the responses:

- Creative industry respondents felt the development of generative AI was not respecting the information rights of creators. They argued that generative AI developers are chiefly responsible for ensuring information rights are being exercised.
- Several respondents, particularly generative AI developers and industry bodies, consistently argued that is difficult to facilitate people's information rights once data is ingested and compiled into a training dataset and is used to train a generative AI model. One large

generative AI developer argued that deployers should be mainly responsible for facilitating information rights.
- Generative AI developers and industry respondents mainly argued that certain measures, such as retraining a model to erase the influence of personal data, would be impractical or not technically feasible. They instead argued that issues such as rectification and erasure should be exercised at the application level, particularly via the use of output filters.
- Civil society respondents argued that if generative AI developers cannot uphold information rights, then their development and deployment is unlawful. They said that the non-compliant models must be retrained on compliant data. They argued that non-compliant models cannot exist just because some claim they are innovative. They also said that while they accept that new technologies require new ways of meeting information rights, untested and unproven methods are not acceptable.

On the need for generative AI developers and deployers to evidence how they are making sure people have meaningful, concise and easily accessible information about the use of their personal data, the following key points arose:

- Generative AI developers and industry bodies argued that it would be disproportionate to inform every person about processing web-scraped data. They reference the exemption at article 14(5)(b) of the UK GDPR. This provision provides an exception from controllers' obligations under the right to be informed when receiving personal data from a source other than the individual, if providing this information proves impossible or would involve disproportionate effort; or would likely render impossible or seriously impair the processing's objectives.[52]
- However, civil society groups and the creative industries argued that web-scraping is invisible processing. They argued that this processing cannot meet people's reasonable expectations, even taking into account article 14(5)(b).
- Generative AI developers and industry bodies argued that they meet transparency and notice requirements through public notices generally explaining that they are using publicly-accessible data. Some also argued that broad categories of source data, such as 'publicly-accessible information' provide appropriate levels of transparency, as opposed to exhaustive lists of sources.

On the need for generative AI developers and deployers to clearly justify the use of any exemptions and demonstrate how they are safeguarding people's interests, rights and freedoms, the following key points arose in the responses:

- Tech industry trade bodies raised article 11 and argued that if a developer cannot identify who the data in a model relates to, it is not personal data.
- One law firm argued that data minimisation is important. They said that precise collection criteria and excluding certain sources play a key role in safeguarding rights and freedoms.
- This was linked to another point made by many respondents, that most people are ill-equipped to access information about the processing or understand the technicalities. Organisations should not rely on them to find the correct information and exercise their rights.
- We received numerous arguments about machine unlearning.[53] These mainly pointed to its theoretical application and not any current practical usage.
- The technology sector and industry groups strongly emphasised input and output filters as a suitable safeguard. Civil society expressed strong doubts about the effectiveness of these filters, citing the ease of jailbreaking through prompts injection.[54]

Our response

We welcome the agreement from all respondents about the importance of ensuring information rights in the context of generative AI. It is vital that, across the generative AI lifecycle, organisations have processes in place to enable and record people exercising their information rights. However, we did not receive clear and verifiable evidence from generative AI developers or the wider industry about the practical measures that could enable people to exercise their rights.

Data protection by design and by default is a legal requirement. This means that, when organisations develop generative AI systems, they must adopt appropriate measures to protect people's rights from the outset. We are increasingly concerned that many organisations developing and deploying generative AI models and systems do not have measures in place to effectively respond to people's information rights requests, particularly where those requests concern web-scraped personal data. In the absence of

effective tools to comply with people's requests, and depending on the organisation's lawful basis, their processing may be unlawful.

Not all information rights are absolute. For example, if an organisation is relying on legitimate interests as the lawful basis and there remains an overriding legitimate interest in continuing the processing (which also passes the three-part test), then the right to erasure would not apply.[55] However, organisations must consider rights requests on a case-by-case basis. There may be some cases where they do not have compelling legitimate grounds which override the individual's rights.

Many respondents mentioned output filters as a useful tool for implementing information rights. However, these may not be sufficient, as they do not actually remove the data from the model.

Organisations must therefore have mechanisms in place to fulfil information rights[56] requests for both the training data and, if a model contains personal data, the trained model itself. The controller is accountable for complying with people's information rights. If a developer and a deployer are joint controllers, they must determine which of them is best positioned to respond to information rights requests.

We expect generative AI developers and deployers to substantially improve how they fulfil their transparency obligations towards people, in a way that is meaningful rather than a token gesture. Testing both whether the measures in place actually work, as well as new and more innovative solutions, can help organisations to comply with the transparency principle. It may help them with the principles of lawfulness and purpose limitation. It will also enable people to obtain meaningful information about the processing so that they can exercise their information rights. We will continue to engage with stakeholders on promoting effective transparency measures, without shying away from taking action when our regulatory expectations are ignored.

Finally, controllers should not apply article 11 so broadly that it has the effect of undermining people's rights. To rely on article 11, controllers would need to establish that they cannot identify a person. Controllers must assess their ability to identify a person on a case-by-case basis. In circumstances where a controller is unable to identify a person, they should inform the person and offer easy ways for the person to provide additional information. This may enable the organisation to identify that person's personal data. In addition, article 11 serves to guard against unnecessary retention of data, in line with the data minimisation principle.

51 Generative AI fourth call for evidence: engineering individual rights into generative AI models

52 Exemptions

53 See glossary and A Survey of Machine Unlearning

54 Here jailbreaking refers to techniques used to bypass the input and output filters that restrict or control what the AI model can process and produce. When an entity “jailbreaks" a generative AI system, they exploit weaknesses in these filters, often by crafting specific prompts or inputs that cause the AI to ignore or circumvent its constraints. This can result in the AI providing responses that it would otherwise block, such as sensitive information, prohibited content, or outputs that could be potentially unsafe.

55 This exemption would not apply to the right to rectification, provided the individual has the required evidence that the data in question is inaccurate.

56 It should be noted that controllers need effective tools that enable them to respond to all information rights requests, not only the right to erasure.

Allocating controllership across the generative AI supply chain

In brief: Our position on the allocation of controllership remains the same. See our original call for evidence for the full analysis.

Respondents

In August 2024, we published the fifth chapter of our consultation series. This set out a policy position on allocating controllership across the generative AI supply chain.

We received 16 responses from organisations, with one respondent identifying as a member of the public. Seven responses came via our survey, with a further nine received directly via email. In contrast to previous calls, the technology sector (five) was the most represented. There was just one creative industry response. Industry groups (four) and law firms (two) were the next most represented.

Of the survey respondents, five (71%) agreed with our initial analysis.

Original call for evidence

In our original call for evidence,[57] we set out our positions on controllership and generative AI. To recap those positions were as follows:

Firstly, we said that a contract does not necessarily determine whether an organisation is a controller, joint controller or processor. Instead, this is determined by the practical realities of the processing. In generative AI, the roles of 'developers' and 'deployers' don't always neatly map onto the concepts of controllers and processors. Roles and responsibilities under data protection law are also not determined by other legal regimes, such as intellectual property or competition law.

Secondly, the allocation of controller, joint controller or processor roles must reflect the actual levels of control and influence over the purposes and means of each different processing activity taking place. We emphasised the importance of this in complex supply chains such as generative AI.

Thirdly, we set out that the relationship between developers and third-party deployers towards the "closed-source" end of the generative AI model release spectrum will mean that there are often shared objectives and influence from both parties for the processing. Joint controllership, instead of a processor-controller arrangement, may be more likely.

Finally, we set out that, based on a lot of current practices in "closed access" scenarios, it is unlikely that such developers can claim to be processors at the deployment stage. This is because the generative AI developer is likely to influence overarching decisions (such as the training data, model architecture, risk mitigations and model distribution) that predetermine how data will be processed during deployment. This means that joint controllership remains likely.

Key points from the responses

On determining means and purposes for models closer to the "closed-access" end of the model access spectrum:

- There was clear recognition from all respondents that the varying degrees of control and influence at each stage of processing activity are key factors. These make allocating controllership a challenging task.
- Generative AI developers accept they are controllers for the development stage of models, but do not want to accept controllership responsibility for downstream deployment of their models. One generative AI developer stated that, in some cases, they may consider themselves a processor if they are acting on instructions from a third party.
- The prevailing argument from the technology sector and trade membership bodies is that generative AI developers do not exert a sufficient degree of control or influence over the purposes or means at the deployment phase to be considered controllers or joint controllers for this phase. Instead, they stated that they are processors.

On joint controllership for "closed-access" models:

- Both an industry body and the respondents from the technology sector (including two generative AI developers) agreed that, in some scenarios, they may be joint controllers. However, they argued that in most cases, deployers determine entirely separate purposes for processing personal data (eg when fine-tuning a model). Therefore, they consider themselves a separate controller.

- One industry and trade membership body raised that, in joint controllership scenarios, there is potential for one party to hold a dominant negotiating position. They may unfairly push obligations onto the other party. They also raised that, as a complex arrangement, it may cause delays in responding to subject access requests.
- One technology company argued that joint controllership does not improve legal certainty or accountability. Instead, they argued it created confusion. They advocated for clear contracts setting out defined roles and liabilities as a preferred solution.

Our response

We understand that allocating accountability in complex supply chains such as generative AI may be more challenging than in simpler contexts. This is why, in the call for evidence, we welcomed real-life evidence on the complex decisions deployers and developers make when allocating accountability. Based on the evidence we received, we maintain our consultation position and continue to engage with the sector where they have queries.

Generative AI developers' overarching decisions may influence how a model operates at the deployment stage. As a result, there is a risk that deployers of "closed-access" models do not have meaningful control and influence over all the processing at deployment. In such circumstances, when considering the role of the parties in the processing, it will be important to consider what information the deployer has and the level of control that the developer provides. Joint controllership is more likely where both parties retain shared objectives or influence during the deployment stage.

There was also some stakeholder confusion around what joint controllership entails. It does not mean that both parties have joint and equal responsibility. A joint controllership agreement sets out what each party is responsible for. This means that they do not have to have joint responsibility for everything involved in the processing. This is a fact-based assessment. Any other contracts that are set up between developers and deployers will not override the fact-based assessment that determines controllership.

Many respondents asked us to provide further practical examples of processing activities that clearly demonstrate examples of controllership, to help organisations understand their roles. We welcome further engagement with stakeholders on this issue, given the different contexts and variety of processing activities.

Some respondents suggested that developers could be processors when they use a deployer's data to improve their own models, because the deployers would benefit from the improvement. We do not accept this argument. It contradicts our established position that if a developer is processing data for their own purposes, they are a controller for that processing. The fact that clients may also benefit is not enough to make them the sole controller for all processing.

57 Generative AI fifth call for evidence: allocating controllership across the generative AI supply chain

Next steps

This consultation response confirms our current position and clarifies our regulatory expectations on generative AI. Those interested in data protection compliance still need to consult our core guidance on AI and data protection. Following the changes to data protection law through the Data (Use and Access) Bill, we will update and consult on our guidance to reflect the changes and include generative AI.

Our final positions will also align with our forthcoming joint statement on foundation models with the Competition and Markets Authority (CMA). This statement will touch on the interplay of data protection and competition and consumer law in this complex area.

Glossary

Artificial intelligence (AI): The theory and development of computer systems able to perform tasks normally requiring human intelligence.

Fine-tuning: The process of training a pre-trained model on a specific dataset to adapt it for a specialised task, adjusting its parameters slightly instead of starting from scratch. Common in AI applications like NLP and image recognition, it saves time and resources while leveraging the model's foundational knowledge.

Generative AI: Generative AI is a type of AI that can generate outputs that resemble human-created content. Most of the current generative AI systems are based on the Transformer architecture.

Hallucination (also referred to as confabulation): When a generative AI model produces incorrect, misleading or fabricated content due to reliance on patterns in its training data rather than factual understanding.

Invisible processing: Processing where people are not aware their personal data is being processed, undermining their information rights among others.

Machine unlearning: Adjusting a model’s parameters to remove the influence of specific data, though current techniques are not scalable for large AI models like generative AI.

Synthetic data: This is ‘artificial’ data generated by data synthesis algorithms. It replicates patterns and the statistical properties of real data (which may be personal data). It is generated from real data using a model trained to reproduce its characteristics and structure.

Transformer architecture: Transformer architecture is a type of deep learning model designed to process sequences of data, eg text, by focusing on the relationship between different parts of the input. The transformers are fast, flexible, easily scalable and are highly efficient in understanding context, making them the foundation for generative systems.

Annex: Summary of impact responses

Generally, the levels of engagement with the impact aspects of our calls for evidence were low and limited impact evidence was provided.

Across the five calls for evidence, the overall balance of sentiment around the impact of our regulatory position was inconclusive. The majority (35, 39%) of respondents indicated that our regulatory approach would result in both costs and benefits for organisations.

However, for our positions on lawful basis, purpose limitation and accuracy of training, the net anticipated impact was positive excluding those who were unsure.

The main benefits identified that the proposals would result in improved regulatory certainty for their organisation (39 respondents) and public trust in the adoption of generative AI models (30 respondents).

The majority of respondents that identified costs highlighted that the proposals would result in increased:

time costs of understanding and implementing the regulatory approach (20 respondents);

costs associated with making changes to organisations' business models (10 respondents); and

resource costs associated with people exercising their information rights (seven respondents).

Overview

In developing our policy position, it is important to consider whether regulatory action is proportionate and not unduly burdensome on those that we regulate. Through the calls for evidence, we have collected impact responses from those we regulate to ensure our final policy positions are informed and evidence-based.

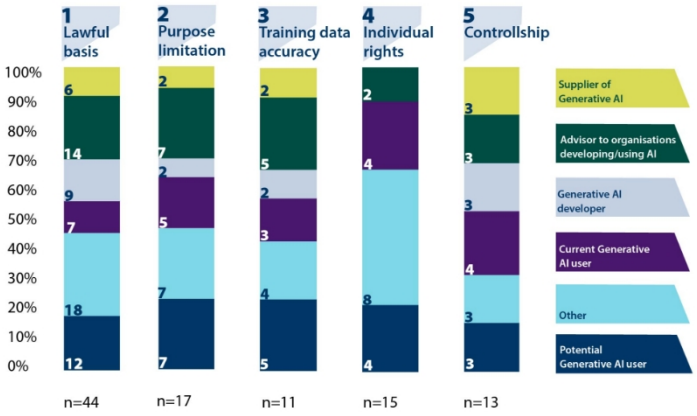
Overview of impact respondents

We received 192 responses from organisations and 22 from members of the public. Respondents could submit evidence through an online survey link or by email. Of the total respondents, approximately 100 answered the impact-related questions in some form. Of these:[58]

- 16 respondents were developers of generative AI;
- 14 respondents were suppliers of generative AI;
- 24 respondents were current users of generative AI;
- 32 respondents were potential users; and
- 33 respondents were advisors to organisations using or developing generative AI.

Figure 1 below shows the breakdown of respondents across the five calls for evidence.

Figure 1: Which of the following describes your organisation?[59]



Source: ICO analysis

58 Multiple responses were permitted.

59 Totals do not add to total sample size as multiple answers were permitted.

Views on the impacts of our proposals

Across the five calls for evidence, we asked respondents answering on behalf of organisations about the impact of our proposals. This asked respondents to provide information on the likely costs and benefits of our proposals and whether these would affect organisations' ability to offer services to the UK market (if at all).

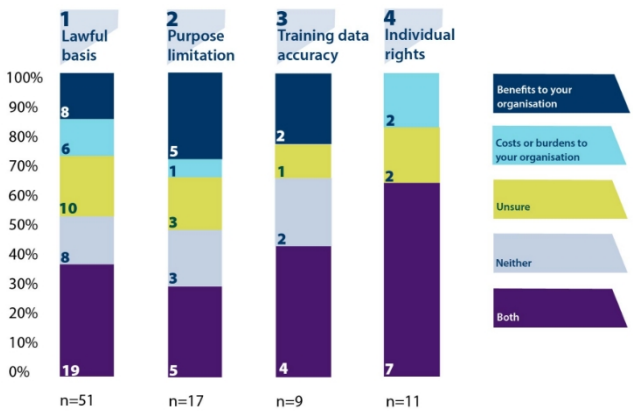
Some comments provided by respondents referred to issues based in the law, rather than the regulatory approach we proposed in the calls for evidence.

When we asked respondents whether our regulatory position would result in additional impacts:

- 15 respondents thought the proposals would result in benefits for their organisation;
- nine respondents thought the proposals would result in costs or burdens to their organisation;
- 35 respondents thought there would be costs and benefits;
- 13 respondents thought there would be neither; and
- 16 respondents were unsure.

Figure 2, below, shows the breakdown across the calls for evidence. Please note that sample sizes varied throughout according to the number of respondents that provided a response to each question.[60]

Figure 2: Do you think the proposed regulatory approach will result in benefits or costs for your organisation?



Source: ICO analysis.

Note: Controllership is not included in Figure 2 as the framing of impact related questions differed for the controllership call for evidence. Refer to section 4.1.5 for discussion of impact responses on controllership.

Across the five calls for evidence, the overall balance of sentiment was inconclusive in terms of who provided a response around whether our approach would result in additional costs or benefits. As shown in Figure 2, the majority (35, 39%) of respondents that answered the impact questions identified that our regulatory approach would result in both costs and benefits for organisations. However, for our positions on lawful basis, purpose limitation and accuracy of training, the net anticipated impact was positive excluding those who were unsure.

Table 1 highlights the main impacts of the proposals that respondents identified across the five calls for evidence. We asked respondents to select from multiple choice questions to identify which impacts would result from our proposals.[61] Few respondents elaborated further on the details of anticipated costs or benefits beyond the level of detail set out in Table 1.

The main benefits identified noted that the proposals would result in improved:

- regulatory certainty for their organisation (39 respondents); and
- public trust in the adoption of generative AI models (30 respondents).

The majority of respondents that identified costs highlighted that the proposals would result in increased:

- time costs of understanding and implementing the regulatory approach (20 respondents);
- costs associated with making changes to organisations' business models (10 respondents); and
- resource costs associated with people exercising their information rights (seven respondents).

A small number (three) of respondents commented on how the proposals would affect their ability to offer services to the UK market. Two of these were received for the lawful basis call for evidence, where respondents highlighted that:

"We consider it inappropriate to require the developer to impose downstream controls, which would have a fundamental chilling effect on any development of AI that is open source or open innovation in the UK."

(The need to detail specific processing purposes at each stage of the AI lifecycle may) "inhibit firms' ability to develop and deploy Gen-AI in practice".

Similarly, one respondent to the controllership call for evidence also commented that:

"if the understanding of when a developer/provider of AI is to be considered as (joint) controller was to be broadened, this may significantly increase the obligations to comply with data protection law and constitute an additional obstacle to making AI services available on the market."

On the basis of the feedback received, it is challenging to make inferences on what the proposals mean for the market as a whole given the small sample size and representativeness of respondents.

Table 1: Summary of impact responses[62]

Impacts		1: Lawful basis	2: Purpose limitation	3: Accuracy of training data	4: Individual rights	Total
Benefits	Providing regulatory certainty to your organisation	16	10	5	8	39
	Improved public confidence in the adoption of	12	9	4	5	30

	generative AI models					
	Reputational benefits from a reduced risk of data protection harms	12	7	3	5	15
	Other	7	1	1	3	12
Costs	Resource costs of defining purpose for each processing activity		5			5
	Resource costs of people exercising their information rights				7	7
	Time-costs of understanding and implementing the regulatory approach	4	6	3	7	20
	Costs associated with making changes to your organisation's business model	3	2	2	3	10
	Costs of assessing the accuracy of training data			3		3
	Costs of			2		2

	communicating the accuracy of model outputs to end-users					
	Costs of accessing proprietary datasets for model training	2				2
	Other	3	1	1	2	7

Source: ICO analysis. Multiple responses permitted.

Note: The impact categories that organisations were asked to identify varied across each call for evidence. Impacts which were not relevant for a specific call for evidence are denoted by the shaded boxes.

Respondents that selected "other" costs or benefits provided some additional details. For benefits, this included:

- a better ability to define organisational strategy;
- increased confidence around risk management; and
- reduced regulatory burden.

Some of the costs identified included:

- additional legal costs;
- reduced ability to innovate; and
- increased development costs.

We explain these in more detail in the next section.

60 Some respondents only provided responses to certain questions which is reflected in the sample size as a result.

61 Impacts that organisations were asked to identify varied across each call for evidence. Impacts which were not relevant for a specific call for evidence are denoted by the shaded boxes in Table 1.

62 Controllership is not included in Table 1 as the framing of impact related questions differed for the Controllership call for evidence, limiting

comparability with previous calls for evidence. Refer to section 4.1.5 for discussion of impact responses relating to controllership.

Further exploration of impact feedback

In addition to the impacts identified in Table 1, we also asked respondents for further information on their selected response and evidence on how impacts could be quantified. This section discusses the feedback that was received across the five calls for evidence where respondents provided further information.

Lawful basis

As shown in Table 1, the first call for evidence received the highest level of engagement, with around 50 respondents. The majority of respondents agreed with the benefits we identified in Table 1, although only 10 respondents provided further clarification. Out of those respondents who provided further information, one noted that:

> "the…approach will benefit our organisation and other responsible developers of generative AI. The (ICO's) approach validates our core belief that only developers…who respect IP and privacy rights should be able to benefit from the development and deployment of generative AI models."

Another respondent argued that a potential benefit would mean an improved ability to define organisational strategy and approach to the use or adoption of generative AI models.

Several respondents highlighted the importance of regulatory certainty and the impact this can have on organisations' willingness to invest. One respondent commented that:

> "the lower the certainty an organisation can enjoy, the less likely they are to enter that market."

Four different organisations responding to the first call for evidence identified the following costs:

- mitigating against unauthorised scraping;
- establishing legitimate interests on personal data used in training;
- assessing consent and other activities associated with data protection impact assessments; and
- deployers of generative AI models in obtaining transparency information around the sourcing of training data from third party developers.

Elaborating further on potential costs, one respondent stated that:

> "We consider it inappropriate to require the developer to impose downstream controls, which would have a fundamental chilling effect on any development of AI that is open source or open innovation in the UK."

Another respondent felt that the proposals were likely to result in additional time costs for organisations implementing the ICO's approach.

> "We also believe that organisations, data controllers and processors, will struggle to understand and apply the ICO's approach, resulting in time and resources costs for them."

One organisation highlighted the importance of web-scraped data for model efficacy. They made a number of suggestions around how to quantify impacts.

> "Estimating the benefits of utilising web-scraped data for training generative AI models involves considering several factors:
>
> Firstly, leveraging publicly accessible data through web scraping enables a more extensive and diverse dataset, enhancing the model's ability to generalise and generate realistic outputs. This broad dataset contributes to the model's accuracy and performance during training, fine-tuning, and post-deployment phases.
>
> The benefits can be quantified by assessing the efficiency gains in model development, as large-scale web scraping provides a substantial volume

of data crucial for training generative AI models effectively. Additionally, the diverse data sources contribute to the model's adaptability across various applications, potentially expanding its market usability.

Furthermore, the potential for innovation and competitive advantage arises from the improved capabilities of generative AI models trained on comprehensive datasets. This may lead to the development of cutting-edge applications, enhancing the organisation's product or service offerings.

To calculate these benefits, one can consider the reduction in data acquisition costs compared to alternative methods, the increased efficiency in model development, and the potential revenue growth resulting from the superior performance of generative AI models. It's essential to weigh these benefits against the legal and ethical considerations outlined in the ICO consultation, ensuring compliance and responsible use of web-scraped data."

Other responses noted that model development had already reached a stage where "the genie is out of the bottle", given the prevalence of web-scraped data in existing frontier models. They said that the need to detail specific processing purposes at each stage of the AI lifecycle may "inhibit firms' ability to develop and deploy Gen-AI in practice".

Despite the overall balance of responses indicating a net anticipated benefit, more detailed feedback suggests that, for developers, the impact of a more limited use of web-scraped data is likely to be on balance a net-cost. However, it is challenging to draw conclusions for the market as a whole due to the limited sample of survey responses.

Purpose limitation

With 17 responses there was limited engagement on the impact considerations of the second call for evidence. Only two respondents elaborated further on potential impacts, beyond those set out in Table 1. They noted that the approach may result in additional benefits and costs.

One respondent who highlighted an additional benefit suggested that it will increase their confidence in risk management. However, another respondent which highlighted an additional cost suggested:

> "A key cost may be reduced ability to innovate with general purpose Gen-AI models…due to incompatibility with tightly defined (and potentially contractually imposed) purposes higher up the lifecycle."

None of the respondents were able to provide a cost or supporting evidence on quantifying the impacts.

While the net anticipated impact of this call for evidence was positive,[63] it is challenging to draw conclusions on the impact for the market as a whole given the limited levels of engagement.

Accuracy

As with the previous call for evidence, there was limited engagement with only nine respondents answering the impact-related questions on whether the approach would result in costs or benefits. Only three respondents provided further information beyond the impacts identified in Table 1.

One respondent, representing the creative industries, highlighted:

> "one of the benefits our members would incur is a more equal playing field, when being forced to compete with synthetic outputs."

Another respondent highlighted the potential cost impact of information sharing between developers and deployers of generative AI.

> "measuring accuracy and addressing any risks or challenges requires detailed information sharing and cooperation between developers and deployers. It is not yet clear whether the market will evolve in a way that will facilitate such an approach. This will have an impact on the costs of implementing Gen-AI use cases."

Most respondents could not provide an estimate for the impacts. However, one organisation suggested that the benefits of the approach could be estimated by:

> "considering market demand for compliant solutions along with cost savings from avoiding fines and a potential competitive advantage."

While the overall balance of respondents suggests a net positive impact, the impact evidence received is inconclusive, given the small sample size and representativeness of respondents.

Individual rights

With 11 responses received there was limited engagement on the impact considerations of the call for evidence on individual rights. While seven respondents agreed that there are likely to be costs associated with people exercising their information rights, only one respondent elaborated further on the cost implications of compliance.

> "A resource cost exists already for Data Subjects exercising their Rights. Additional FTE costs incurred through the introduction and production of Policies, Procedures, Forums and other associated Framework construction."

Overall the impact evidence on this call for evidence is inconclusive given the small number of responses received.

Controllership

Like other calls for evidence, there was limited engagement on the impact of the proposals on controllership, with only eight responses received. When asked about the impact[64] of the proposals on controllership:

- One respondent thought they would have a major impact on their organisation.
- Two respondents thought there would be a moderate impact.
- Three respondents thought the impact would be minimal.
- Two respondents thought there would be no impact.

Of these, only two provided further information.

Respondents that indicated the proposed regulatory guidance would have a major or moderate impact highlighted:

> "Depending on the regulatory approach adopted, the concepts of controllers and processors will be defined accordingly and consequently their responsibilities and liabilities."
>
> "if the understanding of when a developer/provider of AI is to be considered as (joint) controller was to be broadened, this may significantly increase the obligations to comply with data protection law and constitute an additional obstacle to making AI services available on the market."

While the responses received indicate that the proposals are, on balance, a net-cost to organisations, it is challenging to draw conclusions for the market as a whole due to the limited sample of survey responses.

63 On the basis of the absolute number of respondents that answered the approach would result in costs or benefits for their organisation

64 Respondents were asked what scale of impact the proposals would have on their organisation and prompted to provide further details. The controllership call for evidence did not ask whether impacts would be positive or negative.

Actioning the impact feedback

Generally, the levels of engagement with the impact aspects of our calls for evidence were low and provided limited impact evidence. Thus, there were no major changes in our positions attributable to this feedback. Lower levels of engagement could imply limited concerns about the potential impact. This is something we will monitor through our continued stakeholder engagement.

As highlighted throughout the main body of our response, we have taken on board the overall feedback received during the calls for evidence and have amended our regulatory position where appropriate and proportionate. We will also address the feedback received when producing updated guidance on AI. We will follow best practice and guidance in our 'Impact Assessment Framework'[65] and will consider the proportionality of further assessment of impacts as we move towards updating our AI guidance. This will build on the impact evidence raised during our calls for evidence.

65 ICO (2023) The ICO's Impact Assessment Framework. Available at: https://ico.org.uk/media/about-the-ico/documents/4027020/ico-impact-assessment-framework.pdf (accessed 21 November 2024)