Privacy International's response to the Information Commissioner's Office's call for evidence on "Generative AI first call for evidence: the lawful bases for web scraping to train generative AI models"

March 2024

Introduction

Privacy International (PI)[1] welcomes the opportunity to provide input to the Information Commissioner's Office (ICO)'s call for evidence on the lawful basis for web scraping to train generative AI models. PI has a history of advocating for the strict application of data protection laws to mass web scraping, for example in the Clearview AI case.[2] We also have a history of working on "invisible processing" in both the offline and online contexts.

We welcome the ICO's initiative to evaluate the application of a legal basis to web scraping for generative AI training (the Analysis). However, we believe that the ICO's Analysis overlooks several important implications of a simplistic declaration that web scraping for generative AI training may rely on legitimate interests, without further details as to how developers can balance individuals' information rights against their own interests.

Together, these mean that relying on legitimate interests to scrape data from the web to build generative AI models is rife with problems. We are not convinced that existing practice (where data collection is indiscriminate and outputs are unpredictable) stands

[1] Privacy International (PI) is a London-based non-profit, non-governmental organization (Charity Number: 1147471) that researches and advocates globally against government and corporate abuses of data and technology. It exposes harm and abuses, mobilises allies globally, campaigns with the public for solutions, and pressures companies and governments to change. PI challenges overreaching state and corporate surveillance so that people everywhere can have greater security and freedom through greater personal privacy. Within its range of activities, PI investigates how peoples' personal data is generated and exploited, and how it can be protected through legal and technological frameworks. It has advised and reported to international organisations like the Council of Europe, the European Parliament, the Organisation for Economic Cooperation and Development and the UN Refugee Agency.

[2] PI, "Challenge Against Clearview AI in Europe" (2021)

up to the scrutiny and standards established by the GDPR, DPA 2018, and by authorities like the ICO for the protection of people's rights and the rigour of legitimate interest assessments.

In our submission, we discuss three key areas the ICO should further consider:

1) The risks of an overly permissive approach to the "legitimate interests" test. Lack of precision here may leave a door wide open for personal data to be misused or abused in the future in wider contexts as technology develops;

2) The barriers to exercising information rights in the context of "invisible processing" activities like web scraping; and

3) A public registry system for generative AI models.

The approach taken by the ICO towards web scraping for generative AI models may have important downstream repercussions for the future of people's information rights online. Certain current online practices already present a problematic ecosystem: behavioural advertising notably creates an environment with considerable uncertainty and opacity about how, and for what purpose, people's data is captured, transferred and processed online by large but faceless organisations.

If the balance is got wrong with respect to emergent AI practices, then people stand to have their rights to privacy and to protection of their personal data further violated by new and emerging technologies. The growth of generative AI and LLMs is likely to further drive business models that depend on large scale scraping, processing, and potentially exploitation, of personal data with limited regard for people's rights and interests. There is little reason to think that LLMs will not become easier, cheaper and more accessible to develop and run in the future – increasing their prevalence and potentially their harm, unless steps are taken to require less intrusive practice.

1. Risks of an overly permissive approach to the "legitimate interest" test

The ICO Analysis recognises that there may be both business interests and societal interests that could qualify as a legitimate interest (LI) for scraping data from the web to train a generative AI model. It rightly acknowledges – and we agree - that assessment of whether these interests are being met in practice is challenging because, as the Analysis articulated, "if you don't know what your model is going to be used for, how can you ensure its downstream use will respect data protection and people's rights and freedoms?"

This challenge is fundamental and inherent to the very design of most generative AI models: which is intentionally to be of a general and indiscriminate nature rather than only for a specific purpose. They can be used to draft legal submissions to courts, to

generate harmful pornographic content, to provide instructions on building bombs with limited materials, or to produce misleading content about high-profile and/or elected personalities. The "specific purpose and use" of a generative AI model may therefore be impossible to determine at the point of scraping. Recently developed services already illustrate this aspect of their model, OpenAI for example offers a "GPT store"[3] that provides access to a variety of GPT-based chatbots with widely different purposes, from academic research assistants to text-to-speech tools for maths tutors.

Collaterally justifying other forms of large-scale data scraping

Given the unavoidably generic nature of web scraping for generative AI, the Analysis must further engage with the wide-ranging and far-reaching implications of its assessment of how it engages the LI test. If the ICO considers that legitimate interest can be a lawful basis for training generative AI models on web-scraped data, then it must also consider what other forms of large-scale web scraping of personal data it is allowing the LI test to permit.

Permitting developers to scrape large amounts of personal data from the web to train AI models could risk collaterally opening the doors for other entities to justify large scale collecting of personal data under the same pretence of the "legitimate interests" of the business. It may even further incentivise and/or legitimise the development of new business models that depend on web scraping and other large scale and indiscriminate means of data collection, such as that of Clearview AI.

In any case, the ICO ought to more fully articulate how developers and deployers of generative AI models should assess whether their scraping does constitute a legitimate interest in theory, what steps they should take to ensure that that interest is being met in practice, and what response is necessary in the event that evidence in practice does not match up to aspirations in theory.

Different types of data collected

The LI test is also difficult, if not impossible, to properly apply to large-scale web scraping in part because blanket scraping cannot easily discern between the types of data it collects. It therefore cannot properly assess the consequences of processing for relevant data subjects. Consequently, PI calls for the ICO to consider limits to permitting the LI legal basis to apply to large-scale data scraping.

One such solution might be to set out types of data that should never be scraped under the LI test, whether for generative AI training or any other reason. Other regulatory frameworks are restricting what is or is not allowed to be scraped from the web, such as

[3] https://chat.openai.com/gpts

public images. The forthcoming EU AI Act, for example, prohibits applications of AI that may threaten democracy and citizens' rights including the "untargeted scraping of facial images from the internet" to create facial recognition databases.[4] A similar limitation ought to apply in the UK, if only for consistency between regulatory approaches to a technology that is cross-border by nature. It is essential that the Analysis is 'future-proofed' against other tools and techniques that are yet to come.

Third-party deployment of models

The Analysis also recognises that the LI assessment is further complicated by developers making models available to third parties (whether on an open-source or closed-source basis). However, it fails to adequately address the implications and responsibilities of this. As the ICO provides in its guidance on AI and data protection,[5] the purpose limitation principle requires that developers define why they are processing personal data and only process data for that purpose. However, how can developers demonstrate that their models are meeting their identified purpose when they release their models to third parties who might tailor them to their unique needs beyond the developers' intended purpose? This is particularly problematic for open-source models which are freely accessible, fine-tuneable and can be used in a wide variety of scenarios.

The Analysis places sole responsibility on the developers to ensure their models meet the purpose test even when the model is sold or distributed to third parties. This risks creating a "safe harbour" for third parties who should also have some responsibility. The various responsibilities and obligations for both developers and third parties need further consideration, for example in terms of what kind of transparency, accountability and redress mechanisms are expected of all parties. This should not diminish developers' responsibilities but rather prevent the appearance of accountability vacuums.

Safeguards against jailbreaking are inadequate

Even if an LI can be established, web scraping for generative AI must still be balanced against risks to individuals' rights (such as outputting harmful materials and/or personal information scraped from the web, inferred – accurately or not – from training data or entirely hallucinated).

---

[4] Article 5(1)(db) of the 26 January 2024 final compromise text. See also European Parliament, "Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI" (12 September 2023), https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

[5] ICO, "Guidance on AI and data protection" (15 March 2023), https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/

The Analysis proposes that developers implement "technical", "organisational" and "monitoring" measures and controls and contractual restrictions to guard against these risks. However, evidence abounds showing that technical guardrails are not robust enough to protect against inevitable misuse.

There are countless examples of successful jailbreaking[6] methods like DAN[7] and SneakyPrompt[8] that provoked chatbots like ChatGPT and Bard into delivering harmful outputs that bypass safety guardrails[9] (notwithstanding that these outputs are learned from such harmful content the LLM has processed into its training dataset) or outputted personal data scraped in its training dataset.[10]

Researchers from Carnegie Mellon University (CMU) similarly developed 'adversarial attack' methods[11] and concluded that jailbreaking can be automated[12] in such a way that there is an unknown and unlimited number of ways to break in. The CMU research concluded: "it is unclear whether such behaviour can ever be fully patched by LLM providers … It is possible that the very nature of deep learning models makes such threats inevitable. Thus, we believe that these considerations should be taken into account as we increase usage and reliance on such AI models."

Similarly, the AI Safety Institute found that LLM safeguards can easily be bypassed[13] where "users were able to successfully break the LLM's safeguards immediately" using basic jailbreaking techniques, and "more sophisticated jailbreaking techniques took just a couple of hours and would be accessible to relatively low skilled actors".

---

[6] Jailbreaking is a process of "design[ing] prompts that make the chatbots bypass rules around producing hateful content or writing about illegal acts." Matt Burgess, "The Hacking of ChatGPT Is Just Getting Started" (13 April 2023), https://www.wired.co.uk/article/chatgpt-jailbreak-generative-ai-hacking

[7] "New jailbreak! Proudly unveiling" (2023), https://www.reddit.com/r/ChatGPT/comments/10tevu1/new_jailbreak_proudly_unveiling_the_tried_and/

[8] Yuchen Yang, et al., "SneakyPrompt: Jailbreaking Text-to-image Generative Models" (10 November 2023), https://arxiv.org/abs/2305.12082

[9] Rhiannon Williams, "Text-to-image AI models" (17 November 2023), https://www.technologyreview.com/2023/11/17/1083593/text-to-image-ai-models-can-be-tricked-into-generating-disturbing-images/

[10] Jason Koebler, "Google researchers' attack prompts" (29 November 2023), https://www.404media.co/google-researchers-attack-convinces-chatgpt-to-reveal-its-training-data/

[11] Andy Zou, et al., "Universal and Transferable Adversarial Attacks" (20 December 2023), https://llm-attacks.org/

[12] Clint Rainey, "Computer scientists claim" (2 August 2023), https://www.fastcompany.com/90932325/chatgpt-jailbreak-prompt-research-cmu-llms

[13] AI Safety Institute, "AI Safety Institute approach to evaluations" (9 February 2024), https://www.gov.uk/government/publications/ai-safety-institute-approach-to-evaluations/ai-safety-institute-approach-to-evaluations

These examples raise two important considerations that the ICO must factor into its analysis of the balancing questions of the LI test for web scraping: 1) encouraging developers to implement safeguards in their AI models is not robust enough a mitigation solution based on the inherent fallibility of such "after-the-event" patches to protect against the harms to individuals - they are closing the stable door after the horse has bolted; and 2) the dependence of AI on web scraped data and the lack of real-time human oversight over its outputs creates unavoidable risks of output harms, including from a data protection perspective. While no technology is entirely secure from hacks or breaches, the key difference in the case of AI models is that the black-box nature of the algorithm means that there are an infinite number of potential vulnerabilities as opposed to "hard-coded" algorithmic logic which can be manually fixed and secured after, for example, a security audit.

A permissive tilt to the UK data protection regime

Two important legislative developments contextualise the risk of too broad an approach being taken to the LI lawful basis. While neither of these directly apply to web-scraping by LLMs, they demonstrate how an overly permissive approach could potentially be exploited by attempts to shift analysis away from a careful contextual consideration of the impact on people's rights and towards blanket approvals of problematic practice.

- The Data Protection and Digital Information Bill seeks to introduce the concept of a "recognised legitimate interest". The difficulty of performing a meaningful LI assessment for web-scraping for generative AI may result in developers and others seeking a similar short-cut mechanism to bypass proper analysis;

- The Investigatory Powers Act (Amendment) Bill seeks to introduce the concept of bulk personal datasets where there is "no, or only a low, reasonable expectation of privacy". While this only applies within a regime for the Intelligence Services, AI developers or others may seek to argue that the data they obtain through web-scraping ought to be similarly treated as subject to lower standards.

PI opposes the introduction of both of these new legal definitions in part because of how they may negatively affect the wider landscape of data protection law in the UK. This is a key example of that.

2. Exercising information rights in the context of invisible processing

The Analysis identifies web scraping as an 'invisible processing' activity where individuals are not aware that their personal data may be scraped to train a generative AI model. Invisible processing can restrict people's knowledge about, and frustrate their ability to exercise, their rights and the ICO has previously challenged invisible processing by companies that have web scraped personal data (see examples below). However, the

Analysis does not go further to address the tension between invisible processing and the exercise of information rights in the context of web scraping and generative AI.

Invisible Processing Case Studies: Clearview and Experian

In 2022, the ICO fined Clearview AI for scraping images of people from the web to create a database of faces for law enforcement use.[14] The ICO found Clearview to be in breach of UK data protection laws in several ways, including the failure to use people's information in a way that is fair and transparent, where individuals were not made aware or would not reasonably expect their personal data to be used in such a way and a failure to have a lawful basis for collecting personal data.[15] This is no different from the type of web scraping used to build generative AI training data sets as concerns this call for evidence: it is inherently problematic for people's data to be gathered without their knowledge, as is the case when it is scraped off the internet.

The ICO has also found invisible processing by Experian to be violation of data protection law for reasons that should similarly apply to generative AI web scraping.[16] This processing entails:

- Lack of transparency;
- Individuals not being properly notified of Experian processing their data;
- Creation of databases built by collecting publicly available data and data obtained by third parties; and
- Experian's legitimate interests assessment failed due to its lack of regard for the intrusive nature of its profiling and implications on transparency.

In the above cases, the ICO rightly held a high standard of the LI test for invisible processing. While PI appreciates the complexities in applying the LI test to web scraping for generative AI training, the application of laws ought to be consistent across use cases. Invisible processing is a high risk activity. The purpose of innovation cannot justify

[14] PI, "Challenge Against Clearview AI Europe" (2021), https://privacyinternational.org/legal-action/challenge-against-clearview-ai-europe
[15] ICO, "ICO fines facial recognition database" (23 May 2022), https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/
[16] ICO, "Enforcement powers of the information commissioner" (12 October 2020), https://ico.org.uk/media/2618467/experian-limited-enforcement-report.pdf

exceptional applications of legal standards nor any heightened risks to individuals' rights and freedoms, in particular as the application of legal standards in the early days of a technology have a critical influence on the development and propagation of practices, with the potential to encourage similarly abusive behaviour in future innovations.

The ICO's position regarding web scraping for generative AI (which could also be deployed for law enforcement or commercial use) should build on its position established in the above two cases. The information rights established in the UK GDPR we wish to highlight as particularly relevant to invisible processing in this case are:

- The right to information (Arts 12 to 14)
- The right of access by the data subject (Art 15)
- The right to erasure (Art 17)

Individuals simply cannot exercise these rights when their personal data is scraped into a training dataset for generative AI. Respectively, individuals are not informed that their personal data has been scraped into a dataset; they consequently may not have the appropriate means for accessing the data scraped as they are not even aware their data has been scraped; and current research[17] demonstrates that individual data deletion proves difficult for existing technology, as deleting individual data points from a training dataset requires, for most standard LLMs, the whole model to be retrained from scratch.

3. Public registry for generative AI models

A key complication for applying the LI test to generative AI models is the lack of transparency and certainty as to (a) what data is being processed and (b) for what purposes training data is being processed. We encourage the ICO to consider the implementation of a public registry system for generative AI models (including LLMs) that use web-scraped data in their training dataset. Such a system should provide a comprehensive list of all such systems, detailing each system's usage, data elements and ownership, technical details about scraping technique (including identifying user-agents) among other things, to ensure more transparent oversight and accountability.

Such a registry could help meet two needs for both the individual and the developer:

[17] Antonio A. Ginart, et al., "Making AI Forget You" (2019), https://proceedings.neurips.cc/paper_files/paper/2019/file/cb79f8fa58b91d3af6c9c991f63962d3-Paper.pdf

1. Individuals can find out whether personal data about them is being collected by registered models, and they have a direct, accessible opt-out method to exercise their right to access, their right to be informed and their right to erasure (if this can be effected in practice).

2. Developers registering their models can meet the transparency requirement without having to navigate the "impractical" feasibility of individually informing data subjects whose data form part of the training set,"[18] an argument typically brought in as the transparency exception under UK GDPR Article 14(5)(b).

Public Registry Case Studies

California's Data Broker Registration statute (SB 362)[19] show that it is possible to hold private bodies accountable to the public for their data collection practices. SB 362 requires data brokers (defined as "a business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship") to register with the state and appear in a publicly available data broker registry accessible by any member of the public.[20] The registry enables consumers to find data brokers and request to opt out of the collecting and sharing of their personal information.[21] Recently, the California Privacy Protection Agency responsible for the registry has been exploring a one-time deletion button that would allow consumers to send one deletion request at once to all registered data brokers rather than having to do so individually.

EU laws are also mandating greater public transparency and accountability for VLOPs and VLOSEs via the Digital Services Act. While not exactly a registry system, this demonstrates the need for, and value of, the public being able to easily access information about how powerful data-intensive companies are operating.

A registration model provides one potential way to improve oversight and compliance for implementing robust accountability measures with reasonable demands that meet both the rights of individuals and the interests of developers. It would not, however, absolve in any way developers from carrying out a full legitimate interests test that should lead to strict safeguards against the blanket scraping of web data and negative impacts on

[18] Claudio Novelli, et al., "Generative AI in EU Law" (17 January 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4694565
[19] Senate Bill No. 362 (12 October 2023), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB362
[20] Data Broker Registry, https://oag.ca.gov/data-brokers
[21] Titus Wu, "California's new data broker law: explained" (14 November 2023), https://news.bloomberglaw.com/in-house-counsel/californias-new-data-broker-law-explained

people's rights. As noted above, it is far from clear that existing practice is able to pass this test.

If web scraping for the purposes of generative AI development can be shown to have a lawful basis, then its extremely high risks mean that strict monitoring and abundant transparency would be proportionate.

Therefore, whatever position the ICO adopts with respect to web scraping, it is imperative that it sets out a way for people to be informed about what their personal data is being used for so they can exercise their rights.

Conclusion

Governance of large language models and other forms of generative AI is likely to be a key battleground for people's data and privacy rights. As the incentives to amass and process ever more data intensify, so do the risks for people's rights as enshrined in national, regional and international laws. The UK GDPR is technologically neutral, but its interpretation and guidance must keep pace with new and emerging developments.

The Analysis provides a good basis, but it overlooks several important risks of web scraping by generative AI models. In particular:

- The risk that an overly permissive approach for scraping by generative AI will further intensify harmful online data extraction and processing practices.

- The additional risks that the invisible nature of web scraping pose for people's ability to exercise their legal rights.

- The extent and scale of AI driving and normalising new intrusive practices.

We urge the ICO to more deeply consider its position on the scope, safeguards and context of the LI test for web scraping and the possibility of a registry system for generative AI models as a form of shaping greater transparency, oversight and compliance.

Additional topics to discuss in the future series

As mentioned throughout our submission, topics we recommend for discussion in a future series are:

- Bans on the collection of certain types of data, notably special categories data that should in no circumstances be scraped and processed for a training data set;
- Technical feasibility of the right to erasure in LLMs;
- Use of user-inputted data in generative AI models;

- New technical measures to govern web-scraping.

We welcome the opening of a second chapter in the consultation series concerning purpose limitation, which is also a key consideration in this matter.