

RESEARCH ARTICLE

Private ordering, generative AI and the ‘platformisation paradigm’: What can we learn from comparative analysis of models terms and conditions?

Lilian Edwards^{1,2,3}, Igor Szpotakowski⁴ , Gabriele Cifrodelli², Joséphine Sangaré² and James Stewart⁵

¹Newcastle Law School, Newcastle University, Newcastle upon Tyne, UK; ²CREATe, University of Glasgow, Glasgow, UK; ³Pangloss Consulting Limited, Edinburgh, UK; ⁴The School of Law, University of Leeds, Leeds, UK and ⁵Department of Science and Technology Studies, University of Edinburgh, Edinburgh, UK

Corresponding author: Igor Szpotakowski; Email: i.m.szpotakowski@leeds.ac.uk

(Received 9 August 2024; revised 23 October 2024; accepted 6 November 2024)

Abstract

Large or “foundation” models are now being widely used to generate not just text and images but also video, music and code from prompts. Although this “generative AI” revolution is clearly driving new opportunities for innovation and creativity, it is also enabling easy and rapid dissemination of harmful speech and potentially infringing existing laws. Much attention has been paid recently to how we can draft bespoke legislation to control these risks and harms; however, private ordering by generative AI providers, via user contracts, licenses and privacy policies, has so far attracted less attention. Drawing on the extensive history of study of the terms and conditions (T&C) and privacy policies of social media companies, this paper reports the results of pilot empirical work conducted in January–March 2023, in which T&C were mapped across a representative sample of generative AI. With the focus on copyright and data protection, our early findings indicate the emergence of a “platformisation paradigm,” in which providers of generative AI attempt to position themselves as neutral intermediaries. This study concludes that new laws targeting “big tech” must be carefully reconsidered to avoid repeating past power imbalances between users and platforms.

Keywords: generative AI; private ordering; large language models; platformisation paradigm; terms and conditions; privacy policy

The choice is never ‘government regulation’ or ‘no regulation’ – it’s always ‘government regulation’ or ‘corporate regulation’..... You either live by rules made in public by democratically accountable bureaucrats, or rules made in private by shareholder-accountable executives.

Cory (Doctorow, 2024)

1. Introduction

In November 2022, ChatGPT was released by OpenAI and the world changed. Although large, general purpose or “foundation” models and their generative products had already been in the research arena for several years (OpenAI, 2018; see also: Brown et al., 2020),¹ it was ChatGPT’s debut which captured

¹The first large language model (LLM) was published in 2018 by Open AI. Foundation models are deep learning models based on very large training datasets, used as general-purpose technologies that can support a diverse range of use cases.

the public and media's imagination as well as large amounts of venture capital funding. Large models generating not just text and image but also video, games, music and code were widely promoted as set to revolutionise innovation and democratise creativity, against a background of media obsession. We have lived in the socio-cultural and economic hype bubble thus created since, for better or worse, but there are now signs that disillusionment is setting in, with the AI hype bubble beginning to burst (see Abdoullaev, 2023; Pratim Ray, 2023).

However, challenges and complexities remain: literature already emphasises that foundation models may create serious societal risks, including embedding and outputting bias; generating fake news, illegal or harmful content and inadvertent “hallucinations”; infringing existing laws relating, e.g. to copyright and privacy; as well as environmental, security and workplace concerns (Bird *et al.*, 2023; Birhane *et al.*, 2023; House of Lords, Communications and Digital Committee, 2024; Weidinger *et al.*, 2021). Most developed nations are now considering regulation to address these worries, whether via mandatory comprehensive legislation, e.g. the EU AI Act (Official Journal of the European Union, 2024); siloed or vertical legislation (Cyberspace Administration of China, 2023; Parliament of Canada, 2024); adapting existing law (see the many copyright lawsuits underway (Samuelson, 2024); or by “soft law” such as codes of conduct (Department for Science, Innovation & Technology, 2024), “blueprints” (White House, 2022) or industry guidelines (Veale *et al.*, 2023).

What has had less attention has been *self-regulation* (sometimes known as *private ordering* in the contractual context – see Gunningham, 1991) by model providers via a variety of instruments which range from the arguably more legally binding terms and conditions (T&C) imposed on users²; privacy policies or notices; and licenses of copyright material; to the fuzzier and more public relations-friendly but less enforceable “acceptable use” policies, stakeholder “principles” and codes of conduct. These terms, binding or otherwise, are also often cascaded down from model providers to downstream deployers as part of their agreement with ultimate users.

Such conditions have been widely studied, and often reviled (Wauters *et al.*, 2014) in the history of e-commerce, especially in the business-to-consumer or “B2C” context, as largely unread, not understood and accepted without possibility of negotiation through imbalance of power in monopolistic or oligopolistic markets (Micklitz & Palka 2017; Betkier, 2019; Marique & Marique, 2019; Palka & Lippi, 2021). As such they form part of a general history of abuse of power in consumer contracts generally, and relationships with digital platforms specifically (Micklitz, 2014). Social media networks in particular display network effects which have made it impossible for a real marketplace of choices to operate, displacing consumer choice. Privacy policies have become notorious for inordinate lengthiness (Loos & Luzak, 2016) and requiring reading comprehension abilities at university level (Edwards & Brown, 2013; Jensen & Potts, 2004; McDonald & Faith Cranor, 2008), and users have no incentive to read them anyway (Obar & Oeldorf-Hirsch, 2020) as they often change frequently without additional consents sought.

As such, Palka has named T&C of online platforms “terms of injustice” and argued they should no longer be tolerated (Palka, 2023). This nuclear option is unlikely; but as discussed below, unfair terms legislation, and particularly the EU Digital Services Act (DSA) have attempted to curb and expose

There are a number of phrases used in this arena, including general-purpose AI models (GPAI, the choice of the EU AI Act, see below), “frontier” models, generative AI and LGAIMs (Large General AI Models). We do not choose here to go into the lengthy debates on definitions and what separates these, as they are not, in general, terms of art and some are promotional or brand-recognition terms. It is important to note however that generative AI predates large models, e.g., the use of Generative Adversarial Networks (GANs). We have chosen instead to identify models and services provided by *function* as most germane to our study: Text to Text (T2T); Text to Image (T2I); Text to Audio or Video (T2A/V), see Section 3.

²We adopt here the definition of terms and conditions adapted from the DSA art 3(u) as “all clauses, irrespective of their name or form, which govern the contractual relationship between the provider and the recipients of the service.” Licenses and privacy policies may not always be regarded in strict law as contractual terms but do inform the behaviour of users of the service provider and possibly also provide transparency to third parties; we subsume them into T&C for brevity except where clearly pointed out otherwise.

their worst excesses. Meanwhile, T&C, privacy policies, etc. remain interesting, not just for providing transparency about provider data practices but also for exposing noncompliance with relevant mandatory laws including consumer and due process rights (Suzor, 2016). Advocates and regulators have often found this work useful as a way to defend consumer and societal interests against tech giants.

Arguably, the risks of generative AI should thus be controlled by democratically made legislation not self-preferencing private ordering (see for a strong rejection though of the idea that terms of service are a valid way of allowing providers to govern themselves; Palka, 2018). But legislative process moves slowly, and although the first wave of AI legislation is underway – the EU AI Act, for example, has now passed as of April 2024 – their bedding in, interpretation and enforcement will still take time. In the USA, the home of most foundation models, so far the only mandatory Federal law concerning AI and large models is an Executive Order which affects only public agencies (White House, 2023). In China, by contrast the national internet regulator, the Cyberspace Administration of China (CAC), has taken a global lead by announcing on 13 July 2023 the Interim Measures for the Management of Generative Artificial Intelligence which took effect on 15 August 2023 (Cyberspace Administration of China, 2023). According to Article 22 (1) of the Measures, “generative artificial intelligence technology refers to models and related technologies that have the ability to generate content such as text, pictures, audio, and video” (Cyberspace Administration of China, 2023, Article 22). The most interesting feature of the Chinese rules is the premarket licensing of the generative AI models, discussed in Section 3 (Cyberspace Administration of China, 2023, Article 23). Despite these developments, private ordering remains probably the most significant current form of governance of foundation models.

1.1. Inspiration and method

Our initial provocation in January 2023 – only three months after the ChatGPT *coup de foudre* – was that social media platform T&Cs had been extensively studied for decades, but almost no work had yet been done on the T&C of foundation models. Systematic collection of datasets of T&C and privacy policies, such as ToSBack (ToSBack, 2023), and the Princeton-Leuven Longitudinal Corpus of Privacy Policies (Princeton-Leuven, 2023), has been historically a strong feature of US research on online platforms but in Europe; such projects are less prevalent, perhaps because stronger regulation (privacy, consumer law) replaced the need to rely wholly for user remedies on publicising and enforcing T&C. However, recent times have seen arrivals such as CLAUDETTE (European University Institute, 2024), which analyses consumer contracts and privacy policies for unfair terms using machine learning, and in 2021, the Platform Governance Archive, an open-source “data repository and platform that collects and curates policies of major social media platforms in a long-term perspective” (Efferenn, 2023).

The EU itself now collects the T&C of major digital players as part of its DSA efforts (Code.europa.eu, 2024) which have become mandatory since February 2024 (European Commission, 2024a). In January 2023, however, no-one seemed to be systematically collecting and analysing the T&C of foundation model providers. A generative AI dataset is now collected by the Open Terms Archive, which commenced 9 October 2023, but no analysis has apparently appeared, merely the raw terms, and only 18 providers are included (4 in October 2023), and from a much narrower jurisdictional basis than our study (see OpenTerms Archive, 2024).

Accordingly, we sought funding from the UK Research & Innovation (UKRI) Trustworthy Autonomous Systems programme to pilot empirical work in January–March 2023 on the T&C ecology of generative AI model providers (A number of papers have been given on this work. See CREATE Team, 2023). We decided to conduct an empirical study to determine whether self-regulation of large language models (LLMs) is sufficient to protect users. Ambitiously, as a team of legal researchers with one co-author working in the field of technology studies, we decided to map T&C across a representative sample of generative AI providers. We planned to study different *modes* of model (e.g. text,

image, video). We were aware of patent differences between models like Stable Diffusion which cultivated an open-source community-minded image, and the proprietary and somewhat secretive nature of market leader models, such as ChatGPT (though Meta's LLAMA (Chalkidis & Brandl 2024), perhaps unintentionally, provided a counter-example). Although we could not do a full comparison of proprietary vs open-source models, we did incorporate a number of each type of model. Finally, we were keen to find small as well as large providers and to explore a range of countries of origin, not just, as was typical, the USA, UK or EU. These choices reflect our desire to gain a comprehensive understanding of the landscape and how it affects user protections.

After extensive scoping, we found 13 generative AI models which cut across many of these criteria. We examined the T&C, privacy policies, and, in some cases, additional documents such as community guidelines for

- Text-to-Text (T2T) services (ChatGPT, ERNIE, Bard, CLOVA Studio, AI Writer, DeepL)
- Text-to-Image (T2I) services (LENSA, Midjourney, Nightcafe, Stable Diffusion)
- Text-to-Audio/Video (T2A/V) services (Gen-2, Synthesia, Colossyan).

We also considered examining the T&Cs of downstream deployers creating applications based upon top level foundation models, on the basis that governance within the genAI value chain is now too complex simply to focus on model providers alone (Cobbe, Veale & Singh, 2023). We looked primarily at the AI and legal services area which was showing exciting development. For example, much media attention was being paid in early 2023 to Harvey AI (2023), who were partnering with law firms to create bespoke models for them using client and firm data built on top of a GPT LLM. In the end, though, it was impossible at that time to obtain the T&C for Harvey AI due to commercial trade secrecy, and it was difficult to conclusively identify from websites and media clippings whether other prominent legal service providers (e.g. DoNotPay – see Donotpay.com, 2023a, 2023b; Roth, 2022) were in fact using a top-level foundation model as opposed to simply coding their own ML system or rule base. Accordingly, we left the deployer angle for a later time.

In each of these categories, though not for every model, we found specific clauses regarding copyright, privacy or data protection (DP), illegal and harmful content, dispute resolution, jurisdiction and enforcement. We chose to mainly focus on copyright, privacy and dispute resolution. We found little difference in jurisdiction and ouster clauses from those typical of US standard form consumer contracts, and so for reasons of space, this chapter does not deal with that area. This was a very complicated project design, and our results are best seen as qualitative rather than quantitative.

Our multidimensional approach was probably overly ambitious, given timeframe, especially as the T&Cs of many of the models were constantly changing. The large-scale EU and US tracker projects use automated scraper bots, and in future, so would we. However, we do feel that our preliminary findings represent a “line in the sand” worth recording, of valuable historical significance. As such, we choose to present them here as of end March 2023, rather than attempting to update them a year later. Only one other team of researchers in Europe, led by Natali Helberger, reported results analysing T&C from this very early phase of generative AI, and their work, though valuable, was limited to only five providers, and focused mainly on impacts on journalism and the media (Helberger, 2023). Helberger and Samuelson note: “The transparency obligations in the DSA can usefully be sorted into four categories: 1) consumer-facing transparency obligations; 2) mandatory reporting and information access obligations to national regulators and the European Commission; 3) rights of access to data; and 4) obligations to contribute to public-facing databases of information” (Helberger & Samuelson, 2024). Article 14, which requires platforms to enforce their T&C and in so doing have due regard to the fundamental rights of users under the Charter of Rights, is especially important and applies to all platforms defined under the DSA not just to the Very Large Online Platforms (VLOPs) or Very Large Online Search Engines (VLOSEs). We anticipated that our handcrafted dataset would be displaced shortly thereafter by the mighty EU DSA transparency regime (see DSA Observatory, 2024) but in

fact, Section – 3 it is unlikely that generative AI models do in fact fit within the scope of DSA rules (see also Quintais et al., 2023).

2. Analysis of T&C and other provider documents by topic

2.1. Copyright

Our research questions here were informed by the considerable debate on copyright and large, models in academe, the courts and the media (see Friedmann, 2024; Frosio, 2025; Guadamuz, 2024). The key issues that emerged were

1. Who owns the copyright over the outputs of the model? Is it a proper copyright ownership or an assigned license?
2. If output works infringe copyright, who is responsible (e.g. user, service)?
3. Is there any procedure in force (e.g. notice and takedown [NTD], prompt filtering) to avoid or at least minimise the risk of copyright infringement? If yes, what?

2.1.1 Key takeaways

- In almost every model or service studied, ownership of outputs was assigned to the user, but, in many cases, an extensive license was also granted back to the model for co-existing use of the outputs.
- Similarly, in almost every case, risk of copyright infringement in the output work was left, with some decisiveness, with the user.
- Licenses assigning copyright were mostly bespoke, though some use of Creative Commons and OpenRAILs was observed, and boilerplate clauses reminiscent of those used in social media T&C were common. There was a lack of industry norms as to definition of some key terms.
- Even at this early stage of foundation models, every model provider undertook some content moderation, and notice and take down arrangements were the norm.

2.1.2. *Who owns the copyright over the outputs? Is it a proper copyright ownership or an assigned license?*

In every case studied but (possibly) one, ownership over output works was granted to users (Table 1).

Two T2I services (LENSA, 2023a; Midjourney, 2023) assigned ownership to the user but demanded back an extremely wide co existing license, e.g. in the case of LENSА, a Russian service, a “perpetual, revocable, nonexclusive, royalty-free, worldwide, fully-paid, transferable, sub-licensable license to use, reproduce, modify, adapt, translate, create derivative works.” Nightcafé’s T&C (2023) simply stated that the user owns all the IP rights related to outputs. Stable Diffusion (Dezgo, 2023) adopted not a bespoke license but a commonly known open-source license, a version of the BLOOM license, CreativeML Open RAIL-M license (Ferrandis & M., 2023). With regard to T2A/V services (Synthesia, 2023), Gen-2 (Runway, 2023) and (Colossyan Creator, 2023), the scenario was substantially similar.

With regard to T2T services, the scenario differed a little more. For ChatGPT (OpenAI, 2023), OpenAI assigned to the user all the “right, title and interest in and to Output” and also the “inputs” which seems to mean prompt material. Bard (Google, 2023), (Simplified, 2023) and CLOVA Studio (Naver Cloud Platform, 2023) also assigned ownership to users. By contrast, the Chinese company Baidu – proprietor of Ernie Bot (Baidu, 2022) – declared itself the owner of all IP rights of the API service platform and its related elements, such as “content, data, technology, software, code, user interface.” However, this probably referred only to infrastructure, not output works – but it is not entirely clear yet (The Chinese Beijing Internet Court judgment of November 27, 2023, in the case of Li v. Liu might be a possible indicator of policy changes in China toward providing ownership of copyright

Table 1. Analysis of Copyright clauses

Models	Who owns the copyright over the outputs and (if any indication is found) over the inputs?	If a copyright infringement is committed, who is responsible?	Is there any procedure in force to avoid or at least minimise the risk of copyright infringement?
LENSA	User BUT user grants extensive license to provider	User	Yes
Midjourney	User BUT user grants extensive license to provider	User	Yes
Nightcafe	User	User	Yes
Stable Diffusion	User	User	Not really (it reserves the right to restrict the usage of the model)
Synthesia	User BUT user grants extensive license to provider	User	Yes
Gen-2	User BUT user grants extensive license to provider	User	No
Colossyan	User	User	Not really (it reserves the right for access to and removal from the service)
ChatGPT	Input (e.g. prompt): user; output: assignment of “right, title and interest” from company to user	User	Yes
Bard (now Gemini)	User	N/a	Yes
Ernie Bot	Provider	User	Yes
Simplified AI writer	Unclear	User	Yes
CLOVA Studio	User BUT only for purpose of the service’s provision	N/a	No
DeepL	User	N/a	No

of AI-generated content to the users. See translation of the judgment: Wang, 2024). Lastly, (DeepL, 2023) “does not assume any copyrights to the translations made by Customer using the Products.”

2.1.3. *If output works infringe copyright, who is responsible (e.g. user, service)?*

In every model studied (Table 1), the liability for copyright infringement was laid entirely at the door of the user.

Midjourney’s T&C (2023) used entertainingly colourful language:

[i]f you knowingly infringe someone else’s intellectual property, and that costs us money, we’re going to come find you and collect that money from you.

LENSA’s (2023a) was more diplomatic: the user is responsible for any content that “may infringe, misappropriate or violate any patent, trademark, trade secret, copyright or other intellectual or proprietary right of any person.”

Stable Diffusion (Dezgo, 2023) asserted (in US legalistic CAPITAL LETTERS) that the model was provided “on an ‘as is’ basis, without warranties or conditions of any kind, either express or implied, including, without limitation, any warranties or conditions of title, non-infringement, merchantability, or fitness for a particular purpose.” Indeed, the user was held “solely responsible” also for the appropriateness of distributing the *model* not just the outputs, which was possible under Stability’s open-source policies – something it seems unlikely they could control or assess. Any conceivable liability of Stability was limited to US\$100.

In none of these services was there any admission that copyright infringement liability might come from the provider's training of the *model* (e.g. by using copyright works without consent) rather than any bad actions by the user. Some T&C, e.g. ChatGPT, not only asserted that the user was solely responsible for outputs but required the user to indemnify the provider against any liability arising as a result of the user's interactions with their service. These combinations of exclusion, limitation and indemnity clauses may well be wholly or partially invalid in consumer contracts in many jurisdictions but of course the user would have to have the resources to challenge. We saw no real differences across the different modes of model studied, nor by national origin or size.

2.1.4 *Is there any procedure in force (e.g. notice and takedown, prompt filtering) to avoid or at least minimise the risk of copyright infringement? If yes, what?*

Most the models studied (Table 1), including all the T2I services, provided for some sort of mitigation or enforcement measures in relation to copyright infringement liability, e.g. prompt or keyword or proper name blocking. In many cases, the aim may have been to prevent a wider range of harms than just copyright infringement, such as production of child sexual material or other illegal, harmful or adult content; or production of misinformation (fake news).

Many models threatened to ban users who broke the rules of the site. Repeat infringers were especially mentioned. Midjourney's T&C (2023) were the most forthright, stating, without much legal decorum, that:

[a]ny violations of these rules (i.e. the ones indicated both in T&C and Community Guidelines with regard to content restrictions) may lead to bans from our services. We are not a democracy. Behave respectfully or lose your rights to use the Service.

Midjourney also automatically blocked some prompts, and certain words, and implemented flagging of infringing content by users to moderators. Similarly, Nightcafé's T&C (2023) provided for a series of enforcement measures in case of breach or suspicion of breach, including deletion of any user content; suspension or termination of the account; suspension or permanent ban to the site; and disclosure of some prohibited content to appropriate government authorities.

NTD was implemented on most sites, in exactly the same ways as it is on most platform or hosting sites, to immunise themselves from liability under instruments like the Digital Millennium Copyright Act (DMCA) or EU DSA (formerly Electronic Commerce Directive, arts 12–15 – see European Parliament, 2000). Midjourney, ChatGPT, Bard (Google) and LENSEA all offered means to request removal of infringing content, the latter by “written notification to our copyright agent [...], [i]n accordance with the Digital Millennium Copyright Act (17 U.S.C. § 512) (“DMCA”).” It is interesting that a Russian provider was willing to namecheck a US law, showing the effective global scope of DMCA warnings. By contrast, China's Ernie Bot asked for notice of infringing content, “in accordance with the laws and regulations of the People's Republic of China.” This is an unusual reference to a local national law other than the DMCA.

Midjourney had the most developed T2I model moderation scheme. Users were encouraged to flag infringing content to moderators in a well-developed model of community quasi-self-enforcement, although who these moderators were, and how they made their enforcement decisions was (typically) unclear. Nightcafé stated the company had the right (not the obligation) “to appoint community moderators or automoderators from time to time” who “will flag creations and comments for human moderation.”

Invariably, however, even where community moderation was in operation, the provider reserved the right to exercise its own discretion in assessing whether outputs or behaviour violated the T&C, and what the sanction might be (site ban, for example). Gen2 stated unusually it had *no* obligation to review or monitor users or content. CLOVA Studio and DeepL did not apparently provide for enforcement procedures. These kinds of arbitrary denials of due process, unclear terms and sanctions,

and haphazard applications of enforcement are exactly the kind of problems the DSA was drafted to address in relation to moderation failures on conventional platforms: yet as we will see in Section 3, it is likely that foundation models do not, *per se*, fall under the DSA.

2.2. DP rights

2.2.1. Takeaways

- DP rights are the Cinderella sister of foundation model governance, with little attention paid in the early stage of industry development either in T&C or in literature. In April 2023, by no means all model providers had privacy policies.
- Only basic options to exercise rights, such as an email address to complain to, were found in most models studied in early 2023, and not always that.
- Another sweep to December 2023 indicated improvement in recognition of DP rights and more detailed complaint mechanisms for users.

This is a shorter section than the preceding for a reason. While copyright has been a subject of furious debate since the inception of large image and language models, and the DMCA, alongside US copyright law, is well known to carry the risk of punitive sanctions if requests for takedown are ignored, DP, was at the early stage we studied, much less mentioned if at all (Quite possibly this was because of the lack of overarching DP law akin to the GDPR in the US, see Gal, 2023). Arguably it was only the action of the Italian state DP regulator, the Garante, against Open AI and ChatGPT in April 2023 (Lomas, 2023) – after our sampling period – that made the world of foundation models wake up to the likelihood of DP infringement (LNB News, 2024). Yet the EU General Data Protection Regulation (GDPR) (European Parliament, 2016), the leading global DP instrument, is extraterritorial: if services have bases in Europe, sell into Europe or use the personal data of European data subjects, then these rights, at least in theory, must be operationalised for data subjects (GDPR, art 3).

Privacy policies have become effectively universal for digital and platform services, even where not in theory required by local law (Hayne 2007). All the T2T models in our study, regardless of jurisdiction, had a privacy policy. However, at the early stage of industry evolution we covered, in many cases they were that of the parent company's general privacy policies, rather than specially tailored for foundation models. In the wider ecology of models, some had no privacy policy at all and others bore evidence of having been cut and pasted from large social media network or e-commerce templates without much thought of applying them to the new world of foundation models.

The GDPR offers several significant user rights against data controllers including information and subject access rights, the right to erasure of personal data (“to be forgotten”), the right to rectification, rights to data portability, the right to object to processing including profiling and the right to object to solely automated decision-making (GDPR, Chapter III). If personal data are used in training sets to build large models, then there must be a lawful ground of processing (GDPR, art 6). Purpose limitation, DP impact assessments and the age at which children can consent will also, *inter alia*, be relevant. Many of these issues were raised in the (ongoing) Italian action, and the European Data Protection Board is currently constructing guidance on how large models should (if it is even possible) comply with GDPR rules. However, we found evidence of relative failure of model providers in early 2023 to engage with DP user rights.

Of the 13 models we studied (Table 2), 9 models referenced the GDPR in their privacy policies (if they had one). However, seven explicitly mentioned the California Consumer Privacy Act (CCPA). Of these, LENSEA, a Russian service, stood out with a special California Notice and Collection and Privacy Notice (LENSEA, 2023b).

All of the T2T models we studied (Table 2) offered at least an email address to enable objection to and removal of user data. The privacy policies of the other models studied were however far less consistent. In most models studied, providers used boilerplate clauses to talk about the processing of

Table 2. Analysis of privacy policies of GenAI models

Models	Mention CCPA rights (California), EU or UK GDPR?	Mention rights other than erasure explicitly, and do they give a form to claim your rights?	Offer an email address to claim DP rights?
Open AI	No CCPA, EU and UK GDPR	Yes Form	Yes privacy@openai.com
Baidu (ERNIE)	Nothing mentioned	Yes Form	No Specific form
Google (Bard, now Gemini)	CCPA, EU and UK GDPR	Yes Form	Yes Special form
CLOVA Studio	EU GDPR No CCPA no UK GDPR	Yes Email, no form	Yes dl_ncloud_privacy@naver-corp.com
AI Writer	Nothing mentioned	Yes No form	Yes hello@simplified.com
DeepL	EU GDPR	Yes	No
LENSA	CCPA EU and UK GDPR	Yes, Access, modification, correction and erasure.	Yes privacy@prisma-ai.com
Midjourney	CCPA, EU and UK GDPR	Yes Email, no form	Yes privacy@midjourney.com
Nightcafe	EU and UK GDPR	Yes Email, no form	Yes hello@nightcafe.studio
Stable Diffusion	EU GDPR	Yes No easy option to do so	No
Gen-2	Nothing mentioned	Yes Email, no form	Yes support@runwayml.com
Synthesia	UK and EU GDPR	Yes Email, no form	Yes support@synthesia.io
Colossyan	Nothing mentioned	Yes Email, no form	Yes info@colossyan.com

data supplied by subscribers such as registration data and prompts. The much more difficult questions around processing personal data from the public Internet as input to training datasets were not raised. It can be argued that T&C relate only to the relationship between provider and user so issues around the provenance of the training set are irrelevant – but this is patently not true of DP notices which should give third parties as well as contractual partners notice of processing of their data and subject access and other user rights (GDPR, art 14 and Chapter III).

Eleven out of thirteen models studied (Table 2) offered clear contact information, such as an email address or a link to a form within their privacy policies. In first quarter 2023, none of these 13 models included much information about data rights other than subject access rights and the right of erasure. Open AI, providers of ChatGPT, gave no lawful ground for processing data and said only to email them about erasure rights. Compared to the well-trodden path for fulfilling the “right to be forgotten” (Zhang et al., 2024) since 2014 in the wake of *Google Spain* (Court of Justice of the European Union, 2014; Court of Justice of the European Union, 2023) for platforms and search engines, this seemed little more than a token effort.

Due to the lack of information about DP in our survey in April 2023, we chose to do another sweep at end 2023 (Table 2). By then, 12 out of 13 models had decided to update their privacy policies. All 12 now make more, though not completely comprehensive, mention of the range of DP user rights. This is a positive trend away from the DP “wild west.” But most still only offer an email address as

means of complaint (9 out of 13). Two still do not do even that. For larger providers, because privacy policies are generally applicable to all generative AI services provided, and sometimes all services offered *in toto* by a company such as Google (Google generally only have one privacy policy and set of T&Cs across all their services; however, they altered that policy on 22 May 2024 to incorporate some extra terms relating to generative AI into their main conditions including notably their statement that (in line with this paper's findings) ownership in outputs belongs to the user. See the extra terms; Google, 2024), it is extremely difficult to work out what particular models are doing with users' data. Almost certainly, further action by DPA regulators and privacy advocates as well perhaps as the Federal Trade Commission in the USA, will be necessary to force more than lip service compliance. It is notable that Open AI only threw up a RTBF form and added a lawful ground for processing straight after they were banned in Italy by the Garante and they still, as of April 2024, declare that the right of rectification is beyond their abilities (As we went to press, NoYB the privacy NGO announced they were suing ChatGPT for failing to comply with the right to rectification: "ChatGPT provides false information about people, and OpenAI can't correct it," NoYB point out that OpenAI themselves admit that "factual accuracy in large language models remains an area of active research" – showing the importance of documentary admissions on model provider websites – see NoYB, 2024). It is fair to say that DP compliance by large models is still very much a work in progress.

3. Conclusions and next work

Sometimes you have to do the work to know *how* to do the work, and we learnt a lot from this pilot. In future work, we would aim to explore in greater depth several aspects that were either overlooked in their significance or beyond the scope of our current investigation. There is obviously a vast amount to be done regarding prohibited content and behaviour, and how it is policed on these sites, beyond just copyright and privacy. We felt it was vital to look at how T&C changed over time, but we did not have the technology or time to do this. The automated scraper bots powering datasets like those held by Open Terms Archive would in future simplify this task tremendously. Indeed, future work could simply rely on analysing the GenAI database in the Open Terms Archive. Our complex project design in the end seemed unnecessary as we found very similar clauses and issues coming up, regardless of mode of model, size of provider, or country of business. We did not anticipate this; but might hypothesise that large providers such as Google already have well-crafted internationally targeted T&C which were largely applied *mutatis mutandis* to generative models; while small providers across the globe seem largely to have copy pasted from familiar commercial styles seen in social media T&Cs wherever they are located. Even the Chinese services had fairly familiar T&Cs. A key problem was gaining access to B2B T&Cs and this would require careful cultivation of trusted relationships which we could not do in three months.

Substantively, our main finding across the T&Cs examined was a general paradigm in which no ownership was claimed over outputs, but no risk was accepted in relation to them either. Instead risk (e.g. of copyright infringement, of privacy breaches, of production of illegal content) was assigned firmly and unconditionally to the user. This might just be seen as a typical commercial land grab in B2C clauses – deny everything and wait for them to sue you! – but it remains surprising that providers were willing so happily to give up claims to monetise outputs (As noted in Section 2.1., some but not all providers did at least demand back from users a non-exclusive license over outputs). Again, this can be explained as a smart commercial choice to waive rights over output works in return for collecting the input data of users to train bigger and better models – but that makes less sense now given providers more recently seem to have accepted that legally users probably have rights to opt out from (or object to) processing of their data under DP law, even where they are not paying enterprise customers.

We suggest instead that what is happening here is a *platformisation paradigm*, in which model providers attempt to position themselves as “neutral intermediaries” in a style very familiar to those who have studied the case law battles around the Electronic Commerce Directive and the DMCA in the early years of this millennium (For a general history of how this paradigm emerged, see Edwards, 2018; Husovec, 2023). Model providers are seeking all the benefits of neutrality in terms of deferring liability and responsibility to users, while still gaining all the advantages of their position in terms of profit and power. This suggestion is bolstered by the way all or most of the providers in our sample behaved as if they were indeed platforms under the ECD (now DSA) and the DMCA in terms of content moderation; accepting DMCA notices for takedown, removing repeat infringers, etc., as if this would provide them with safe harbours like any other “platform.”

Yet foundation model providers simply are *not* platforms; or certainly were not in the simpler days of early 2023 (Open AI’s GPT Store muddies the waters somewhat – see OpenAI, 2024). Conceptually, a platform was originally an online hosting service which stored and/or made content provided by third parties available to the world. The original policy justification for viewing platforms as neutral actors was to balance the possibly unlimited risk for acts of users that might render the platforms economically unviable, with the need to provide some redress to those affected by legal violations in user-generated content – notably, the copyright industries. Morally, at the turn of 2000, though probably not by only a few years later, there was also a case that it was wrong to “shoot the messenger” unless and until they received notice that their “premises” were being used for no good. Legally, this was crystallised into the familiar safe harbour liability exemptions, and notice and take down (NTD) obligations, introduced via the DMCA and ECD c 2000. Now, in the DSA, the heir to the ECD, a series of definitions continue to describe an “online intermediary service” which is a provider of “information society services,” which here exclusively include hosting, acting as a mere conduit, caching and (originally in the DMCA, and belatedly and with limits in the DSA) providing search engine tools (see DSA arts 3–6. From here on, we will use the DSA only as the legal paradigm for discussing platformisation). Some types of larger or more complex intermediary services then have further obligations placed on them: “online platforms” and “very large online platforms” (VLOPs, or online platforms with over 45 million annual users). But crucially, platforms are still, at root, hosts, which store or make available content “provided by a recipient of the service” (DSA, art 6).

This description simply does not match foundation models. The only content the user supplies is the prompt or other input (e.g. image, database) and storage of it is not the relevant information society service that the model provider is offering; that is (surprise) access to the model. As Hacker et al. (2023) note, “users . *request* information from LGAIMs via prompts, they can hardly be said to *provide* this information” [italic added]. As Botero Arcila puts it, “they [provider sites] neither consist of the merely technical transmission of information nor host *user*-generated content. Rather they host *AI*-generated content.” (Botero Arcila, 2023) agree, pointing out that with LLMs, the relevant content is decidedly not provided by the user, but by the LLM itself.

In policy terms, relieving a model provider of liability is inappropriate because they are not a mere hapless victim of risks deriving from user-created content, but the creator of the content *themselves* by allowing users to query the model. Botero Arcila makes a spirited attempt to argue that LLMs might sneak into the DSA hierarchy as search engines, since VLOSEs do not have to be based on intermediary hosts and this evade its definitional constraints; and LLMs such as ChatGPT are indubitably often used very like search engines. But this ingenious idea ignores foundation models used to generate art, or code, which are rarely if ever used as search engines, and it also ignores those models with too few users to fit into VLOSEs. In fact, it is already becoming common to incorporate LLMs into VLOSEs anyway and in that case, they will fall into the umbrella of the DSA (this is currently true of Bing [using GPT] and Google Search [using Gemini] – see European Commission, 2024b, 2024c).

Although there can be more debate, so far so relatively clear. Foundation models do not *per se* fall under the DSA. The problem then is that the DSA does not, as the ECD did, just provide liability

exemptions; it also demands *positive* steps of hosts, platforms and VLOPs of varying natures (Husovec, 2023). And these are steps that, judging by our research above, are exactly what are needed to protect the B2C users of the generative AI sector. Content moderation actions of model providers are opaque, dictatorial, unclear and unjustified, and opportunities to meaningfully contest arbitrary decisions and sanctions on users are few. To meet these problems on genuine platforms, the DSA *inter alia* provides that,

- *all* hosting services shall provide clear information in plain English about their terms and conditions and content moderation practices (art 14). Content moderation rules must be clear and predictable and based on existing policies (art 14(1)). Importantly, services must also *enforce* their terms in a way that has “*due regard*” for the “*fundamental rights*” of users (art 14(4) – Quintais *et al.*, 2023). It could be argued for example that placing all the risk on model users, as models uniformly do, is not consonant with their rights of free expression.
- *all* hosting services shall meet transparency reporting obligations relating to their content moderation decisions and give reasons for such (art 15).
- online platforms (a subset of hosts) must provide ad archives, implement trusted flaggers and respect due process in internal and external appeals against moderation decisions.

These provisions would be extraordinarily useful in meeting the procedural vices identified above in model T&Cs and would transform their generally hostile and unfair governance approach to disempowered users. There seems no good policy reason why these rules should *not* be applied to foundation models. Currently, model providers are in a position to both benefit from and control the outcomes of their offerings; they assert exemption from liability by passing risk via their terms and conditions to users, but evade the new positive obligations of the DSA. This is unjust. We suggest therefore that the DSA is already not fit for purpose and should be amended to bring foundation models within its scope as soon as possible.

The DSA is only a European instrument of course, if a relevant one. One of the suppositions of this study was that generative AI models are being developed globally not just in Silicon Valley, and perhaps too much attention has been spent on models from a few Western countries, and Western regulation. As of April 2024, according to the CAC announcement, 117 generative AI models have been registered in the People’s Republic of China. We noted earlier that China has legislated extensively in relation to AI and has in fact taken an early lead on regulation of generative AI and in ways quite distinct from the West (Zhang, 2024, p. 291; Abiri & Huang 2023, p. 3). A key feature of the Chinese regulation is that generative AI services are only allowed to provide their services to the public after pre-approval by the regulator (Cyberspace Administration of China, 2024). In order to be approved, developers need to provide documentation of the model, including T&C, to the regulator (CAC). This premarket approval of T&C and Privacy Policies of generative AI services, commonly thought of as state licensing, might be considered as another possible way to mitigate the vices of market-driven private ordering. Indeed, such an idea, of premarket supervision leading to “regulated contracts,” might be regarded as a new spin on the familiar notion of “a Food and Drugs Agency (FDA) for models” frequently floated by US academics and recently investigated by UK thinktank, the (Ada Lovelace Institute, 2024). As the Ada researchers note, while there is substantial support for the idea of state agency pre-licensing, less thought has been put into what it might actually *do*. It is not impossible to imagine that this process might be used not just to regulate how models are *built*, as the EU AI Act does, but also to standardise its terms of *use* when put on the market. By these means, conceivably private ordering practices within a sector could be harmonised, and abusive terms of use and content moderation brought under control (Szpotakowski, 2024).

Looking beyond the EU and China, many jurisdictions have general rules declaring abusive rules in consumer contracts void and restricting unfair commercial practices, which might be extended to the T&C of generative AI models. The DSA itself is to some extent at least partly a consumer

instrument, an extension of these familiar ideas of consumer protection hitherto found in the Consumer Rights Directive and Unfair Commercial Practices Directive, For non-EU states though, unfair terms and commercial practices rules may be a first port of call in considering how to control generative AI contracts. One issue though is whether controls tailored to consumers go far enough. Although we were unable to study B2B contracts, it seems quite likely similar imbalances of power operate between tech giants like Google or incumbents like Open AI and small or medium sized enterprises (SMEs). In an earlier draft of the EU AIA, art 28a of the European Parliament's compromise text did indeed regulate unfair contractual terms unilaterally imposed on an SME or startup by a general purpose AI provider, but this text seems sadly to have slipped out of the final draft.

In future work, we would like to examine if competition in the market itself produces fairer and more balanced T&Cs or if, as seems sadly more likely, a de facto cartel continues to impose the clauses most favourable to a small handful of extremely powerful tech companies. It cannot be equitable that the vices of the extractive social media era – only now being challenged effectively by the DSA, DMA and its ilk – can now slip through the cracks almost accidentally into the era of generative AI.

Acknowledgements. The authors would like to thank the Horizon Centre for Digital Economy Research (University of Nottingham) for funding Professor Lilian Edwards, and the Trustworthy Autonomous Systems Hub for the project funding.

Funding statement. This work was supported by the UKRI Trustworthy Autonomous Systems Hub; the Horizon Centre for Digital Economy Research (University of Nottingham).

Competing interests. The authors declare none.

References

- Abdoullaev, A.** (2023, December 30). AI Bible: Why Generative AI bubble is to burst and Interactive AI is to rise... LinkedIn. <https://www.linkedin.com/pulse/why-generative-ai-bubble-burst-2024-world-embeddings-vs-abdoullaev-14ycf/> last accessed 30 August 2024.
- Abiri, G., & Huang, Y.** (2023). A red flag? China's generative AI dilemma. *Harvard Journal of Law & Technology*, 37, 1–20.
- Ada Lovelace Institute.** (2024). Safe before sale: Learnings from the FDA's model of life sciences oversight for foundation models. <https://www.adalovelaceinstitute.org/report/safe-before-sale/> last accessed 03 May 2024.
- Baidu.** (2022, December 9). Wenxin Large Model Documentation - Service Agreement. Baidu Intelligent Cloud Documents. <https://wenxin.baidu.com/AIDP/wenxin/Yl6th25am> last accessed 04 December 2023.
- Betkier, M.** (2019). *Privacy online, law and the effective regulation of online services*. Intersentia.
- Bird, C., Ungless, E., & Kasirzadeh, A.** (2023). Typology of Risks of Generative Text-to-Image Models. ARXIV. arxiv.org/abs/2307.05543 last accessed 04 July 2024.
- Birhane, A., Kasirzadeh, A., Leslie, D., and Wachter, S.** (2023). Science in the age of large language models. *Nature Reviews Physics*, 5(5), 277–280.
- Botero Arcila, B.** (2023). Is it a platform? Is it a search engine? It's Chat GPT! The European liability regime for large language models. *Journal of Free Speech Law*, 3(2), 453–488.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D.** (2020). Language models are few-shot learners, In *Advances in Neural Information Processing Systems*, 34th Conference on Neural Information Processing Systems, 33, NeurIPS, 1877–1901.
- Chalkidis, I., & Brandl, S.** (2024). Llama meets EU: Investigating the European Political Spectrum through the Lens of LLMs. ARXIV. arxiv.org/pdf/2403.13592 last accessed 20 July 2024.
- Cobbe, J., Veale, M., & Singh, J.** (2023). Understanding accountability in algorithmic supply chains. In *FACt '23: ACM Conference on Fairness, Accountability, and Transparency*. 10.1145/3593013.3594073 last accessed 01 May 2024.
- Code.europa.eu.** (2024). Contrib-versions. code.europa.eu/p2b/contrib-versions last accessed 01 May 2024.
- Colossyan Creator.** (2023). Terms & Conditions. <https://www.colossyan.com/terms> last accessed 04 December 2023.
- Court of Justice of the European Union.** (2014). Judgment of the Court (Grand Chamber) of 13 May 2014. *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*. EU: Case C-131/12.
- Court of Justice of the European Union.** (2023). Judgment of the Court (Grand Chamber) of 4 July 2023, *Meta v Bundeskartellamt*. EU:C-252/21.

- CREATe Team.** (2023, September 13). EPIP 2023 annual conference (11-13 September 2023). <https://www.create.ac.uk/blog/2023/09/08/epip-2023-annual-conference-11-13-september-2023/> last accessed 01 May 2024.
- Cyberspace Administration of China.** (2023, July 13). Interim Measures for the Management of Generative Artificial Intelligence Services (生成式人工智能服务管理暂行办法). https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm last accessed 20 July 2024.
- Cyberspace Administration of China.** (2024, April 2). Announcement of the Cyberspace Administration of China on the release of registered information on generative artificial intelligence services. https://www.cac.gov.cn/2024-04/02/c_1713729983803145.htm last accessed 30 July 2024.
- DeepL.** (2023). Terms and conditions. <https://www.deepl.com/pro-license>, last accessed 04 December 2023.
- Department for Science, Innovation & Technology.** (2024, February 6). Consultation outcome A pro-innovation approach to AI regulation: government response. www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response#:~:text=In%20the%20AI%20regulation%20white,Appropriate%20transparency%20and%20explainability last accessed 01 March 2024.
- Dezgo.** (2023). Stable Diffusion License. <https://dezgo.com/> last accessed 04 July 2024.
- Doctorow, C.** (2024, October 23). In defense of bureaucratic competence. Pluralistic. <https://pluralistic.net/2023/10/23/getting-stuff-done/> last accessed 23 February 2024.
- Donotpay.com.** (2023a). Terms of Service. <https://donotpay.com/learn/terms-of-service/> last accessed 24 July 2023.
- Donotpay.com.** (2023b). Privacy Policy. <https://donotpay.com/learn/privacy-policy/> last accessed 24 July 2023.
- DSA Observatory.** (2024). DSA and Platform Regulation Conference 2024, 15-16 February 2024. Amsterdam. <https://dsa-observatory.eu/the-dsa-and-platform-regulation-conference-2024>, last accessed 24 July 2024.
- Edwards, L.** (2013). Privacy, law, code and social networking sites. In I. Brown (Ed.), *Research handbook on governance of the internet*. Cheltenham: Edward Elgar Publishing 309–352.
- Edwards, L.** (Ed.). (2018). *Law, policy and the internet*. Hart Publishing.
- Efferenn, F.** (2023). The first open access repository of platform policies: HIIG launches new Platform Governance Archive. <https://www.hiig.de/en/the-first-open-access-repository-of-platform-policies-hiig-launches-new-platform-governance-archive/> last accessed 24 March 2023.
- European Commission.** (2024a). How the Digital Services Act enhances transparency online. <https://digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency> last accessed 01 May 2024.
- European Commission.** (2024b, May 17). *European Commission compels Microsoft to provide information under the Digital Services Act on generative AI risks on Bing*. <https://digital-strategy.ec.europa.eu/en/news/commission-compels-microsoft-provide-information-under-digital-services-act-generative-ai-risks> last accessed 22 May 2024.
- European Commission.** (2024c, August 5). Supervision of the designated very large online platforms and search engines under DSA. <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses#ecl-inpage-google> last accessed 05 August 2024.
- European Parliament.** (2000). Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'). <https://eur-lex.europa.eu/eli/dir/2000/31/oj> last accessed 23 February 2024.
- European Parliament.** (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC - General Data Protection Regulation (GDPR). <https://eur-lex.europa.eu/eli/reg/2016/679/oj> last accessed 24 July 2024.
- European University Institute.** (2024). CLAUDETTE machine learning powered analysis of consumer contracts and privacy policies. claudette.eui.eu/index.html last accessed 20 March 2024.
- Ferrandis, C., & M.** (2023). The BigScience OpenRAIL-M License. <https://bigscience.huggingface.co/blog/bigscience-openrail-m> last accessed 01 May 2023.
- Friedmann, D.** (2024). Creation and Generation Copyright Standards. *New York University Journal of Intellectual Property & Entertainment Law*, SSRN. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4770924_code1572004.pdf?abstractid=4770924&mirid=1&type=2
- Frosio, G.** (2025). Should we ban generative AI, incentivise it or make it a medium for inclusive creativity? In E. Bonadio, and C. Sganga (Eds.), *A research agenda for EU copyright law*. Cheltenham: Edward Elgar Publishing Forthcoming.
- Gal, U.** (2023, February 8). ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. The Conversation. <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283> last accessed 29 April 2024.
- Google.** (2023). Terms of Service – Privacy & Terms. <https://policies.google.com/terms> last accessed 04 December 2023.
- Google.** (2024). Generative AI. <https://policies.google.com/terms/generative-ai> last accessed 04 June 2024.
- Guadamuz, A.** (2024). A scanner darkly: Copyright liability and exceptions in artificial intelligence inputs and outputs. *GRUR International*, 73(2), 111–127.
- Gunningham, N.** (1991). Private ordering, self-regulation and futures markets: A comparative study of informal social control. *Law & Policy*, 13(4), 297–326.

- Hacker, P., Engel, A., & Mauer, M.** (2023). Regulating ChatGPT and other large generative AI models. ARXIV. <https://arxiv.org/abs/2302.02337>, last accessed 04 March 2024.
- Harvey AI.** (2023). Harvey Generative AI for Elite Law Firms. <https://www.harvey.ai/>, last accessed 04 month 2023.
- Hayne, A.W.** (2007). Online privacy policies: Contracting away control over personal information?. *Dickinson Law Review*, 111(3), 597–624.
- Helberger, N.** (2023, July 18). Generative AI in media & journalism: Think big, but read the small print first. Medium. <https://generative-ai-newsroom.com/generative-ai-in-media-journalism-think-big-but-read-the-small-print-first-375f2ecb1256> last accessed 24 July 2023.
- Helberger, N., & Samuelson, P.** (2024). Will the EU's Digital Services Act Become a Global Transparency Regime? *Verfassungsblog: From the DMCA to the DSA—A Transatlantic Dialogue on Online Platform Regulation and Copyright*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4783341 last accessed 29 April 2024.
- House of Lords, Communications and Digital Committee.** (2024, February 2). Large Language Models and Generative AI. <https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/5402.htm> last accessed 04 March 2024.
- Husovec, M.** (2023). Rising above liability: The Digital Services Act as a blueprint for the second generation of global internet rules. *Berkeley Technology Law Journal*, 38(3), 883–920.
- Jensen, C., & Potts, C.** (2004). Privacy policies as decision-making tools: An evaluation of online privacy notices. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 471–478.
- LENSA.** (2023a). Terms of Use. <https://T&C.LENSA-ai.com/terms-dec-2-2022> last accessed 04 December 2023.
- LENSA,** (2023b, July 10). App California Notice at Collection and Privacy Notice. T&C.LENSA-ai.com/CCPAPP last accessed 24 July 2023.
- LNB News.** (2024, April 12). CO launches consultation on accuracy of generative AI models. <https://www.lexisnexis.co.uk/legal/news/ico-launches-consultation-on-accuracy-of-generative-ai-models> last accessed 01 May 2024.
- Lomas, N.** (2023, March 31). Italy orders ChatGPT blocked citing data protection concerns. TechCrunch. <https://techcrunch.com/2023/03/31/chatgpt-blocked-italy/>, last accessed 01 April 2024.
- Loos, M., & Luzak, J.** (2016). Wanted: A bigger stick. On unfair terms in consumer contracts with online service providers. *Journal of Consumer Policy*, 39(1), 63–90.
- Marique, E., & Marique, Y.** (2019). Sanctions on digital platforms: Beyond the Public–Private divide. *Cambridge International Law Journal*, 8(2), 258–281.
- McDonald, A., & Faith Cranor, L.** (2008). The cost of reading privacy policies. *A Journal of Law and Policy for the Information Society*, 4(3), 543–568.
- Micklitz, H.** (2014). Unfair Terms in Consumer Contracts. In N. Reich (Ed.), *European Consumer Law* (2nd edn ed., pp. 125–164). Cambridge: Intersentia.
- Micklitz, H. W., Palka, P., and Panagis, Y.** (2017). The empire strikes back: Digital control of unfair terms of online services. *Journal of Consumer Policy*, 40(3), 367–388.
- Midjourney.** (2023). Terms of Service. <https://docs.midjourney.com/docs/terms-of-service> last accessed 04 December 2023.
- Naver Cloud Platform.** (2023). Terms of Service. <https://www.ncloud.com> last accessed 04 December 2023.
- NightCafe.** (2023). NightCafe Creator Terms of Service. <https://nightcafe.studio/policies/terms-of-service> last accessed 04 December 2023.
- NoYB.** (2024, April 29). ChatGPT provides false information about people, and OpenAI can't correct it. <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it> last accessed 01 May 2024.
- Obar, J. A., & Oeldorf-Hirsch, A.** (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1), 128–147.
- Official Journal of the European Union.** (2024, July 12). Artificial Intelligence Act (P9_TA(2024)0138). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689 last accessed 30 July 2024.
- OpenAI.** (2018, June 11). Improving language understanding with unsupervised learning. <https://openai.com/index/language-unsupervised/> last accessed 04 December 2023.
- OpenAI.** (2023, March 1). Terms of Use - March 2023. <https://openai.com/policies/mar-2023-terms> last accessed 04 December 2023.
- OpenAI.** (2024, January 10). Introducing the GPT Store. <https://openai.com/index/introducing-the-gpt-store> last accessed 02 May 2024.
- OpenTerms Archive.** (2024). OpenTerms Archive GenAI-versions. <https://github.com/OpenTermsArchive/GenAI-versions> last accessed 01 May 2024.
- Palka, P.** (2018). Terms of service are not contracts – Beyond contract law in the regulation of online platforms. In S. Grundmann (Ed.), *European Contract Law in the Digital Age* (vol 3, pp. 135–162). Intersentia.
- Palka, P.** (2023). Terms of injustice. *West Virginia Law Review*, 126, 134–184.
- Palka, P., & Lippi, M.** (2021). Big data analytics, online terms of service and privacy policies. In R. Vogl (Ed.), *Research Handbook on Big Data Law*. Cheltenham: Edward Elgar Publishing 115–134.

- Parliament of Canada.** (2024, May 29), Amendments to the Artificial Intelligence and Data Act (“AIDA”) and the Consumer Privacy Protection Act in Canada. <https://ourcommons.ca/committees/en/INDU/StudyActivity?studyActivityId=12157763> last accessed 30 July 2024.
- Pratim Ray, P.** (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154.
- Princeton-Leuven.** (2023). Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset. <https://privacypolicies.cs.princeton.edu/> last accessed 24 July 2023.
- Quintais, J., Appelman, P., and Fathaigh, R. Ó.** (2023). Using terms and conditions to apply fundamental rights to content moderation. *German Law Journal*, 24(5), 881–911.
- Roth, E.** (2022, December 13). DoNotPay is launching an AI chatbot that can negotiate your bills. *The Verge*. <https://www.theverge.com/2022/12/13/23505873/donotpay-negotiate-bills-ai-chatbot> last accessed 02 May 2024.
- Runway.** (2023). Terms of Use. <https://runwayml.com/terms-of-use/> last accessed 04 December 2023.
- Samuelson, P.** (2024, April 11). Thinking about possible remedies in the generative AI copyright cases. *Communications of the ACM*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4770671 last accessed 11 July 2024.
- Simplified.** (2023). *Terms of Service*. <https://simplified.com/terms-of-service> last accessed 04 December 2023.
- Suzor, N.** (2016). The responsibilities of platforms: A new constitutionalism to promote the legitimacy of decentralized governance. *AoIR Selected Papers of Internet Research*, 6, 1–4.
- Synthesia.** (2023). Terms of Service. <https://www.synthesia.io/terms/terms-of-service> last accessed 04 December 2023.
- Szpotakowski, I.** (2024). Self-regulation and licensing as methods of control over generative artificial intelligence in the People’s Republic of China – Preliminary observations. In P. Grzebyk (Ed.), *New Technologies and Law*. Warsaw: Scholar Publishing Press Forthcoming (in Polish).
- ToSBack.** (2023). The terms-of-service tracker. <https://tosback.org/> last accessed 24 July 2024.
- Veale, M., Matus, K., & Gorwa, R.** (2023). AI and global governance: Modalities, rationales, tensions. *The Annual Review of Law and Social Science*, 19(1), 255–275.
- Wang, X.** (2024). Copyright protection for ‘AI-Generated’ images. *GRUR International*, 73(4), 360–368.
- Wauters, E., Lievens, E., & Valcke, P.** (2014). Towards a better protection of social media users: A legal perspective on the terms of use of social networking sites. *International Journal of Law and Information Technology*, 22(3), 254–294.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. H., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., Isaac, W., Legassick, S., Irving, G. and Gabriel, I.** (2021). Ethical and social risks of harm from language models. *DeepMind*. ARXIV. arxiv.org/abs/2112.04359 last accessed 29 March 2024.
- White House.** (2022, October 30). Blueprint For An AI. Bill Of Rights Making Automated Systems Work For The American People. <https://www.whitehouse.gov/wp-content/uploads/2022/10/blueprint-for-an-ai-bill-of-rights.pdf> last accessed 30 July 2024.
- White House.** (2023, October 30). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> last accessed 30 July 2024.
- Zhang, A. H.** (2024). *High wire: How China regulates big tech and governs its economy*. Oxford University Press.
- Zhang, D., Finckenberg-Broman, P., Hoang, T., Pan, S., Xing, Z., Staples, M. and Xu, X.** (2024, June 5). Right to be forgotten in the era of large language models: Implications, challenges, and solutions. ARXIV. <https://arxiv.org/pdf/2307.03941> last accessed 04 July 2024.

Lilian Edwards is an Emerita Professor at Newcastle University, Honorary Professor in Law to the University of Glasgow and Director of Pangloss Consulting. She is a leading academic in the field of Internet law. She has taught information technology law, e-commerce law, privacy law and Internet law since 1996 and been involved with law and artificial intelligence (AI) since 1985. She has worked at the University of Edinburgh, University of Southampton, University of Sheffield, University of Strathclyde and Newcastle University. She is the editor and major author of *Law, Policy and the Internet*, one of the leading textbooks in the field of Internet law (Edwards, 2018). A partner in the Horizon Digital Economy Hub at Nottingham, the lead for the Alan Turing Institute on Law and AI, and a fellow of the Institute for the Future of Work. Edwards has consulted for inter alia the EU Commission, the OECD, and WIPO.

Igor Szpotakowski is a Lecturer (Assistant Professor) in Intellectual Property Law at the School of Law of the University of Leeds. Member of the research team in the project “Mapping Contracts and Licenses around Large Generative Models: private ordering and innovation.” Igor held the position of Visiting Scholar at Peking University (2024), where he researched regulations on Generative AI in China. Previously, he worked at the University of Edinburgh and was a PhD candidate and Deputy Convenor of the Law and Futures Research Group at Newcastle University. He was a Yenching Scholar at Peking University and served as the Principal Investigator for the project “Supply of Services Contracts in Private Law of the People’s Republic of China: Codification in the Era of Decodification” (2020–2024) at Jagiellonian University. Igor’s areas of expertise span across Copyright Law, AI Regulation, Data Protection, and Comparative Private Law.

Gabriele Cifrodelli is a PhD candidate and Research and Teaching Assistant at the University of Glasgow, CREATE Centre. His research interests mostly lie in the intersection between Intellectual Property, Technology and Innovation. Gabriele is the Convenor of the IP Reading Group in CREATE and Coordinating Board Member of SCOTLIN – Scottish Law and Innovation Network. He holds a LLM in Intellectual Property and Digital Economy at the University of Glasgow, and his dissertation “Patent System and Artificial Intelligence: Towards a New Concept of Inventorship?” has been awarded as one of the Outstanding LLM Dissertations in 2021. Gabriele is also a Law Graduate from the University of Trento where his dissertation “Can you patent the sun? The Covid-19 Vaccine as a chance to rethink the relationship between Intellectual Property and Commons” has been recognized as one of the Outstanding Dissertations in the Open Science Field written in the years 2020–2021–2022.

Joséphine Sangaré, LL.M. Joining CREATE in 2022, Joséphine is a PhD researcher, and tutor for law and sociology at the University of Glasgow. Her research focusses on information and communication technology, infrastructure integrity, public-private partnerships, and global knowledge. Joséphine holds an LL.M. in International Law and Security Studies (University of Strathclyde) and a B.A. (Hons) in Law and Political Science (Leibniz University Hanover).

Dr James Stewart is a Lecturer in Science and Technology Studies at the University of Edinburgh. He specialises in use, adoption, development and governance of ICTs across a range of contexts, with a recent focus on data-enabled computational systems such as “AI.” He has published on fields such as user and policy aspects of elder care technology, the internet in Science, Government use of targeted advertising, digital games and play, innovation and governance of Chinese telecommunications, and the digital divide.

Cite this article: Edwards L, Szpotakowski I, Cifrodelli G, Sangaré J and Stewart J. (2025). Private ordering, generative AI and the ‘platformisation paradigm’: What can we learn from comparative analysis of models terms and conditions?. *Cambridge Forum on AI: Law and Governance* 1, e2, 1–17. <https://doi.org/10.1017/cfl.2024.11>