

PRIVATE ORDERING AND GENERATIVE AI

WHAT CAN WE LEARN FROM
MODEL TERMS AND
CONDITIONS?

CREATe Working Paper 2024/5

LILIAN EDWARDS
IGOR SZPOTAKOWSKI
GABRIELE CIFRODELLI
JOSÉPHINE SANGARÉ
JAMES STEWART



CREATe

Private Ordering and Generative AI: What Can We Learn From Model Terms and Conditions?

Lilian Edwards, Igor Szpotakowski, Gabriele Cifrodelli,
Joséphine Sangaré, and James Stewart*

Forthcoming in the Cambridge Handbook of Generative AI and the Law (CUP 2024)

Abstract

Large or “foundation” models, sometimes also described as General Purpose Artificial Intelligence (GPAI), are now being widely used to generate not just text and images but also video, games, music and code from prompts or other inputs. 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 such as deepfakes, hate speech and disinformation, as well as potentially infringing existing laws such as copyright and privacy. Much attention has been paid recently to how we can draft bespoke legislation to control these risks and harms, notably in the EU, US and China, as well as considering how existing laws can be tweaked or supplemented. However private ordering by generative AI providers, via user contracts, licenses, privacy policies and more fuzzy materials such as acceptable use guidelines or “principles”, has so far attracted less attention. Yet across the globe, and pending the coming into force of new rules in a number of countries, T&C may be the most pertinent form of governance out there.

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 providers as well as some downstream deployers. Our study looked at providers of multiple modes of output (text, image, etc), small and large sizes, and varying countries of origin. Although the study looked at terms relating to a wide range of issues including content

* Lilian Edwards is Professor of Law, Information and Society, Newcastle Law School; Igor Szpotakowski is PhD candidate, Newcastle Law School; Gabriele Cifrodelli is PhD candidate, CREATE, University of Glasgow; Joséphine Sangaré is PhD candidate, CREATE, University of Glasgow; and James Stewart is Senior Lecturer, Science and Technology Studies, University of Edinburgh. The authors would like to thank the Horizon Centre for Digital Economy Research (University of Nottingham) for funding Professor Edwards, and the Trustworthy Autonomous Systems Hub for the project funding.

restrictions and moderation, dispute resolution and consumer liability, the focus here is 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 similarly to search and social media platforms, but without the governance increasingly imposed on these actors, and in contradistinction to their function as content generators rather than mere hosts for third party content. This study concludes that in light of these findings, new laws being drafted to rein in the power of “big tech” must be reconsidered carefully, if the imbalance of power between users and platforms in the social media era, only now being combatted, is not to be repeated via the private ordering of the providers of generative AI.

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 Doctorow¹

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 years², it was ChatGPT’s debut which captured 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, as the rest of this book evidences³.

However, all is not rosy: 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 eg to copyright and privacy; as well as environmental, security and workplace concerns⁴. Most developed

¹ C. Doctorow, Oct 23, 2023, in ‘Pluralistic: In defense of bureaucratic competence (23 Oct 2023), <https://pluralistic.net/2023/10/23/getting-stuff-done/>, last accessed 23.02.2024.

²The first large language model (LLM) was published in 2018 by Open AI “Improving language understanding with unsupervised learning”. openai.com. June 11, 2018. See also T. B. Brown et al., ‘Language Models are Few-Shot Learners’, In: Advances in Neural Information Processing Systems, 34th Conference on Neural Information Processing Systems, 33 (NeurIPS, Vancouver, 2020),1877–1901. 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. 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 eg the use of Generative Adversarial Networks or 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 below section 3.

³ P. Pratim Ray, ‘ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope’, Internet of Things and Cyber-Physical Systems, 3, (2023), 121-154. At time of writing there are some signs that disillusionment is setting in with the AI bubble, at least in its hype version, bursting a little: see a pessimistic version of the Gartner Hype Cycle for generative AI at Azamat Abdoullaev, ‘AI Bible: why Generative AI bubble is to burst and Interactive AI is to rise...’, LinkedIn 2023, <https://www.linkedin.com/pulse/why-generative-ai-bubble-burst-2024-world-embeddings-vs-abdoullaev-14ycf/>, last accessed 23.02.2024.

⁴ L. Weidinger et al., ‘Ethical and social risks of harm from Language Models’, DeepMind (2021), arxiv.org/abs/2112.04359, last accessed 29.03.2024; A. Birhane, A. Kasirzadeh, D. Leslie, S. Wachter, ‘Science in the age of large language models’, Nature Reviews Physics, 5 (2023), 277-280; C. Bird, E. Ungless et al “Typology of Risks of Generative Text-to-Image Models”, Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2023) <https://arxiv.org/pdf/2307.05543> last accessed

nations are now considering regulation to address these worries, whether via mandatory comprehensive legislation (eg the EU AI Act⁵); siloed or vertical legislation⁶; adapting existing law (see the many copyright lawsuits underway⁷); or by “soft law” such as codes of conduct⁸, “blueprints”⁹, or industry guidelines¹⁰.

What has had less attention has been *self-regulation* (sometimes known as *private ordering* in the contractual context¹¹) by model providers via a variety of instruments which range from the arguably more legally binding terms and conditions (T&C) imposed on users¹²; privacy policies or notices; and licenses of copyright material; to the fuzzier and more PR-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.

28.04.2024; House of Lords, Communications and Digital Committee, Paper 54 Large Language Models and Generative AI, 2 February 2024.

⁵ Artificial Intelligence Act (P9_TA(2024)0138), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf, last accessed 29.03.2024 (in future, “AI Act”).

⁶ Chinese 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) which took effect on 15th August 2023; 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 01.03.2024.

⁷ See P. Samuelson, ‘Thinking About Possible Remedies in the Generative AI Copyright Cases’, Communications of the ACM (2024), SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4770671, last accessed 11.04.2024.

⁸ Department for Science, Innovation & Technology, Consultation outcome A pro-innovation approach to AI regulation: government response Updated 6 February 2024, 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 01.03.2024.

⁹ Blueprint For An AI Bill Of Rights Making Automated Systems Work For The American People. October 2022. <https://www.whitehouse.gov/wp-content/uploads/2022/10/blueprint-for-an-ai-bill-of-rights.pdf>, last accessed 01.03.2024.

¹⁰ See: M. Veale, K. Matus, R. Gorwa, ‘AI and Global Governance: Modalities, Rationales, Tensions’, The Annual Review of Law and Social Science, 19 (2023), 255-275.

¹¹ See: N. Gunningham, ‘Private Ordering, Self-Regulation and Futures Markets: A Comparative Study of Informal Social Control’, Law & Policy, 13/4, (1991), 297-326.

¹² We 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.

Such conditions have been widely studied, and often reviled¹³ 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 markets¹⁴. As such they form part of a general history of abuse of power in consumer contracts generally, and relationships with digital platforms specifically¹⁵. 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 lengthiness¹⁶ and requiring reading comprehension abilities at university level¹⁷ and users have no incentive to read them anyway¹⁸ 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 tolerated¹⁹. 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 their worst excesses. Meanwhile T&C, privacy policies, et al 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 rights²⁰. Advocates and regulators have often found this work useful as a way to defend consumer and societal interests against tech giants.

¹³ E. Wauters, E. Lievens, and P. Valcke, '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 (2014): 254–94.

¹⁴ E. Marique and Y. Marique, 'Sanctions on Digital Platforms: Beyond the Public-Private Divide', *Cambridge International Law Journal* 8/2 (2019), 258–81; P. Pałka and M. Lippi, 'Big Data Analytics, Online Terms of Service and Privacy Policies', in R. Vogl et al. *Research Handbook on Big Data Law* (Edward Elgar Publishing, Cheltenham 2021). M. Betkier, *Privacy Online, Law and the Effective Regulation of Online Services* (Intersentia, Bristol 2019).

¹⁵ H. Micklitz, 'Unfair Terms in Consumer Contracts' in, N. Reich et al (Eds.), *European Consumer Law*, 2nd, (Intersentia, Antwerp 2014), 125–64.

¹⁶ M. Loos and J. Luzak, 'Wanted: a Bigger Stick. On Unfair Terms in Consumer Contracts with Online Service Providers', *Journal of Consumer Policy* 39/ 1(2016), 63–90.

¹⁷ See C. Jensen and C. Potts, 'Privacy policies as decision-making tools: an evaluation of online privacy notices' In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, (2004) 471–478; A. M. McDonald and L. Faith Cranor, 'The Cost of Reading Privacy Policies', *Am Journal of Law and Policy for the Information Society*, 4/3 (2008), 543–568; L. Edwards and I. Brown 'Privacy, Law, Code and Social Networking Sites' in I. Brown (Ed.) *Research Handbook On Governance Of The Internet* (OUP, Oxford 2013).

¹⁸ J. A. Obar and A. Oeldorf-Hirsch, 'The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services', *Information, Communication & Society* 23/1(2020), 128–47.

¹⁹ P. Palka "Terms of Injustice" *West Virginia Law Review*, Vol. 126, (2023) 133–184.

²⁰ N. Suzor, 'The responsibilities of platforms: A new constitutionalism to promote the legitimacy of decentralized governance', *AoIR Selected Papers of Internet Research*, 6 (2016), 1–4.

Arguably, the risks of generative AI should thus be controlled by democratically made legislation not self-preferencing private ordering²¹. 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 US, the home of most foundation models, so far the only mandatory Federal law concerning AI and large models is an Executive Order which affect only public agencies²². 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 2023²³. 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’²⁴. The most interesting feature of the Chinese rules is the pre-market licensing of the generative AI models, discussed below in the conclusions²⁵. Despite these developments, private ordering remains probably the most significant current form of governance of foundation models.

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 ToSBack²⁶, and the Princeton–Leuven Longitudinal Corpus of Privacy Policies²⁷, 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

²¹ For a strong rejection though of the idea that terms of service are a valid way of allowing providers to govern themselves, see: P. J. Palka, '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, (Intersentia, Bristol 2018), 135–62.

²² See: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Oct.30.2023 at <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 28.04.2024.

²³ CAC, Interim Measures for the Management of Generative Artificial Intelligence Services, July 13, 2023, http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

²⁴ In Chinese: 第二十二条 本办法下列用语的含义是：（一）生成式人工智能技术，是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

²⁵ Article 23 of Chinese Interim Measures for the Management of Generative Artificial Intelligence Services.

²⁶ See: ToSBack. The terms-of-service tracker, <https://tosback.org/>, last accessed 24.07.2023.

²⁷ See: Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset, <https://privacypolicies.cs.princeton.edu/>, last accessed 24.07.2023.

T&C. However, recent times have seen arrivals such as CLAUDETTE²⁸, which analyses consumer contracts and privacy policies for unfair terms using machine learning, and in 2021, the Platform Governance Archive (PGA), an open source “data repository and platform that collects and curates policies of major social media platforms in a long-term perspective”²⁹.

The EU itself now collects the T&C of major digital players as part of its Digital Services Act (DSA) efforts³⁰ which have become mandatory since February 2024³¹. 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 study³².

Accordingly, we sought funding from the UK Research & Innovation (UKRI) Trustworthy Autonomous Systems (TAS) programme to pilot empirical work in January–March 2023 on the T&C ecology of generative AI model providers³³. Ambitiously, we decided to map T&C across a representative sample of generative AI providers. We planned to study different *modes* of model (eg 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, perhaps more by accident than plan, was 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.

²⁸ CLAUDETTE Machine Learning Powered Analysis of Consumer Contracts and Privacy Policies, claudette.eui.eu/index.html, last accessed 20.03.2024.

²⁹ See: F. Efferenn, 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 24.03.2023.

³⁰ See: <https://code.europa.eu/p2b/contrib-versions>, last accessed 01.05.2024.

³¹ European Commission, ‘How the Digital Services Act enhances transparency online’, <https://digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency>, last accessed 01.05.2024. See below, section 4 and notes 41 and 42.

³² See: OpenTerms Archive, GenAI-versions <https://github.com/OpenTermsArchive/GenAI-versions>, last accessed 01.05.2024.

³³ A number of papers have been given on this work: For example, themed panel session “Generative AI and Intellectual Property: Assessing the Regulatory Landscape of the Terms & Conditions as a Start for Responsible Innovation and Adoption” at EPIP (European Policy for Intellectual Property), Krakow, Poland, 11-13 September 2023, see: <https://www.create.ac.uk/blog/2023/09/08/epip-2023-annual-conference-11-13-september-2023/>, last accessed 01.05.2024.

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

- 6 Text-to-Text (T2T) services (ChatGPT, ERNIE, Bard, CLOVA Studio, AI Writer, DeepL)
- 4 Text-to-Image (T2I) services (LENSA, Midjourney, Nightcafe, Stable Diffusion)
- 3 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 alone³⁴. 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 AI³⁵, who were partnering with law firms to create bespoke models for them using client and firm data built on top of a GPT large language model. 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 (eg DoNotPay³⁶) 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, illegal and harmful content, dispute resolution, jurisdiction, and enforcement. We chose to mainly focus on copyright, privacy, and dispute

³⁴ See: eg J. Cobbe et al, 'Understanding accountability in algorithmic supply chains', FAccT '23: 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, June 2023 DOI: <https://doi.org/10.1145/3593013.3594073>, last accessed 01.05.2024.

³⁵ Harvey AI, Harvey | Generative AI for Elite Law Firms, <https://www.harvey.ai/>, last accessed 04.12.2023. Harvey was under development at the time of the survey (and remains so with waitlist for early adoption) so details were vague but such was the hype around it seemed an important actor to include. Harvey is backed financially by a number of genAI providers including OpenAI.

³⁶ Famously, in February 2023, Joshua Browder the CEO of DoNotPay (DNP), advertised he would supply DNP in an earpiece to a party litigant so that it could argue his case in court as a kind of remote "robot" lawyer. When warned however this might result in 6 months in jail for contempt, he terminated the stunt. Since DNP has by his own admission proved to be very inaccurate about law (hallucination problem) this was probably for the best for all. M. Cerullo, 'AI-powered "robot" lawyer won't argue in court after jail threats', CBS News, 26.01.2023, <https://www.cbsnews.com/news/robot-lawyer-wont-argue-court-jail-threats-do-not-pay/>, last accessed 01.05.2023. The inclusion of DNP as a foundation model-based deployer service would have been uncertain as it is not clear how far DNP relied on its own simplistic rule-based AI in first quarter 2023 and how far it was using upstream foundation models to extend its services. See: E. Roth 'DoNotPay is launching an AI chatbot that can negotiate your bills', The Verge, 13.12.2022, <https://www.theverge.com/2022/12/13/23505873/donotpay-negotiate-bills-ai-chatbot>, last accessed 02.05.2024, which announces a demo to show the use of OpenAI's GPT-3 API to build a "convincingly human-like" negotiating chatbot on to DNP's platform.

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 media³⁷. We anticipated that our handcrafted dataset would be displaced shortly thereafter by the mighty EU Digital Service Act (DSA) transparency regime³⁸, see above³⁹; but in fact – see below, section 4 – it is unlikely that generative AI models do in fact fit within the scope of DSA rules.

³⁷ N. Helberger, ‘Generative AI in Media & Journalism: Think Big, But Read the Small Print First’, in *Generative AI in the Newsroom*, Medium, July 18 2023 <https://generative-ai-newsroom.com/generative-ai-in-media-journalism-think-big-but-read-the-small-print-first-375f2ecb1256>, last accessed 28.04.2024. Subsequent research published by the AI for Media and Democracy team (<https://www.aim4dem.nl/>) and AlgoSoc (<https://algosoc.org/>) has incorporated this work but not with a key focus on T&C

³⁸ See early results advertised from the DSA Observatory, the Platform and the AlgoSoc project at the first DSA and Platform Regulation Conference 2024, 15-16 February 2024, Amsterdam, see: <https://dsa-observatory.eu/the-dsa-and-platform-regulation-conference-2024/>, last accessed 28.04.2024.

³⁹ The transparency regimes of the DSA are complex. The main content moderation decisions database can be found at <https://transparency.dsa.ec.europa.eu/>, last accessed 28.04.2024. 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”. 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). See N. Helberger and P. Samuelson “Will the EU’s Digital Services Act Become a Global Transparency Regime?” In *Verfassungsblog: From the DMCA to the DSA—A Transatlantic Dialogue on Online Platform Regulation and Copyright*, forthcoming March 2024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4783341, last accessed 29.04.2024; J. P. Quintais, N. Appelman, R. Ó Fathaigh, ‘Using Terms and Conditions to apply Fundamental Rights to Content Moderation’, *German Law Journal*, 24/5 (2023), 881 – 911.

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 media⁴⁰. 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 (eg user, service)?
3. Is there any procedure in force (eg notice and takedown, prompt filtering) to avoid or at least minimise the risk of copyright infringement? If yes, what?

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.

2.1.1. 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.

⁴⁰ See: eg A. Guadamuz, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs', *GRUR International*, 73 (2), 2024, 111-127; D. Friedmann, 'Creation and Generation Copyright Standards', *NYU Journal of Intellectual Property & Entertainment Law*, forthcoming 2024, SSRN, https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4770924_code1572004.pdf?abstractid=4770924&mirid=1&type=2, last accessed 29.04.2024; G. Frosio, 'Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity?', in: E. Bonadio, C. Sganga (Eds.), *A research agenda for EU copyright law*, Edward Elgar Publishing 2023.

Two T2I services LENSEA⁴¹ and Midjourney⁴² assigned ownership to the user but demanded back an extremely wide co existing license, eg in the case of LENSEA, 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”. Nightcafe’s T&C⁴³ simply stated that the user owns all the IP rights related to outputs. Stable Diffusion⁴⁴ adopted not a bespoke license but a commonly known open-source license, a version of the BLOOM license, CreativeML Open RAIL-M license⁴⁵. With regard to T2A/V services Synthesia⁴⁶, Gen-2⁴⁷, and Colossyan⁴⁸, the scenario was substantially similar.

With regard to T2T services, the scenario differed a little more. For ChatGPT⁴⁹, 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. Bard⁵⁰, Simplified⁵¹ and CLOVA Studio⁵² also assigned ownership to users. By contrast, the Chinese company Baidu – proprietor of Ernie Bot⁵³ – 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 yet⁵⁴. Lastly, DeepL⁵⁵ “does not assume any copyrights to the translations made by Customer using the Products”.

⁴¹ LENSEA Terms of Use’ <https://tos.LENSA-ai.com/terms-dec-2-2022>, last accessed 04.12.2023.

⁴² ‘Midjourney Terms of Service’ <https://docs.midjourney.com/docs/terms-of-service>, last accessed 04.12.2023.

⁴³ ‘NightCafe Creator Terms of Service’ <https://nightcafe.studio/policies/terms-of-service>, last accessed 04.12.2023.

⁴⁴ ‘Stable Diffusion License on Dezgo.Com’ <https://dezgo.com/>, last accessed 04.12.2023.

⁴⁵ C. M. Ferrandis, The BigScience OpenRAIL-M License, <https://bigscience.huggingface.co/blog/bigscience-openrail-m>, last accessed 01.05.2023.

⁴⁶ ‘Terms of Service | Synthesia’ <https://www.synthesia.io/terms/terms-of-service>, last accessed 04.12.2023.

⁴⁷ ‘Runway Terms of Use’ <https://runwayml.com/terms-of-use/>, last accessed 04.12.2023.

⁴⁸ ‘Terms & Conditions | Colossyan Creator’ <https://www.colossyan.com/terms>, last accessed 04.12.2023.

⁴⁹ ‘Terms of Use – March 2023’ <https://openai.com/policies/mar-2023-terms>, last accessed 04.12.2023.

⁵⁰ ‘Google Terms of Service – Privacy & Terms – Google’ <https://policies.google.com/terms>, last accessed 04.12.2023.

⁵¹ ‘Terms of Service | Simplified’ <https://simplified.com/terms-of-service>, last accessed 04.12.2023.

⁵² ‘NAVER CLOUD PLATFORM Terms of Service’ <https://www.ncloud.com>, last accessed 04.12.2023.

⁵³ ‘Wenxin Large Model Documentation – Service Agreement | Baidu Intelligent Cloud Documents’ <https://wenxin.baidu.com/AIDP/wenxin/YI6th25am>, last accessed 04.12.2023.

⁵⁴ 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 of AI-generated content to the users. See translation of the judgment: Xiao Wang, ‘Copyright Protection for ‘AI-Generated’ Images’, GRUR International, 73/4 (2024), 360–368.

⁵⁵ ‘DeepL Terms and Conditions’ <https://www.deepl.com/pro-license>, last accessed 04.12.2023.

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

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

Midjourney's T&C 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 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 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 \$100 US.

In none of these services was there any admission that copyright infringement liability might come from the provider's training of the *model* (eg by using copyright works without consent) rather than any bad actions by the user. Some T&C, eg 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.

2.1.3. 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, including all the T2I services, provided for some sort of mitigation or enforcement measures in relation to copyright infringement liability eg 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 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, Nightcafe's T&C 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; disclosure of some prohibited content to appropriate government authorities.

Notice and takedown (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 Digital Services Act (formerly Electronic Commerce Directive, arts 12-15⁵⁶). 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. Nightcafe 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

⁵⁶ 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').

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 the concluding section below, it is likely that foundation models do not, *per se*, fall under the DSA.

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 (eg 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

2.2. Data protection rights

Takeaways:

- Data protection 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, data protection, was at the early stage we studied, much less mentioned if at all⁵⁷. Arguably it was only the action of the Italian state DP regulator, the Garante, against Open AI and

⁵⁷ Quite possibly this was because of the lack of overarching data protection law akin to the GDPR in the US (see also the reference sometimes to the California DP law rather than the GDPR). One of the authors (Edwards) gave a number of talks on foundation models and the GDPR in early 2023 and at that time, only a few very short blogs could be traced on the topic, eg U. Gal 'ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned', *The Conversation*, February 8, 2023 at <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283>, last accessed 29.04.2024. Since then a few scholarly articles have emerged, primarily on the incompatibility of AI and the "right to be forgotten" – see eg D. Zhang et al "Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions", Sept 2023, arxiv <https://arxiv.org/pdf/2307.03941>, last accessed 29.04.2024; P. Hacker et al. "Regulating ChatGPT and other Large Generative AI Models" FAccT '23, June 12-15, 2023, Chicago, IL, USA, at <https://arxiv.org/abs/2302.02337>, last accessed 04.03.2024; S.B. Vale " Training Large Generative AI Models Based On Publicly Available Personal Data: A GDPR Conundrum That The AI Act Could Solve" DigiCon, <https://digi-con.org/training-large-generative-ai-models-based-on-publicly-available-personal-data-a-gdpr-conundrum-that-the-ai-act-could-solve/>, last accessed 04.03.2024 - but the area is still spectacularly underwritten compared to copyright.

ChatGPT in April 2023⁵⁸ – after our sampling period – that made the world of foundation models wake up to the likelihood of DP infringement⁵⁹. Yet the EU General Data Protection Regulation (GDPR)⁶⁰, 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 subjects⁶¹.

Privacy policies have become effectively universal for digital and platform services, even where not in theory required by local law⁶². 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 making⁶³. If personal data is used in training sets to build large models, then there must be a lawful ground of processing⁶⁴. Purpose limitation, data protection 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 (EDPB) 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.

⁵⁸ N. Lomas ‘Italy orders ChatGPT blocked citing data protection concerns’, TechCrunch 31.3.2023, <https://techcrunch.com/2023/03/31/chatgpt-blocked-italy/>, last accessed at 01.04.2024.

⁵⁹ And regulators to foundation models. The UK's ICO is still consulting on guidance on foundation models – consultation initiated 12 April 2024, see: LNB News, CO launches consultation on accuracy of generative AI models, 12.04.2024, <https://www.lexisnexis.co.uk/legal/news/ico-launches-consultation-on-accuracy-of-generative-ai-models>, last accessed 01.05.2024. EDPB guidance appears to still be a work in progress, although a select handful of regulators have issued advice eg CNIL, Hamburg.

⁶⁰ 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).

⁶¹ GDPR, art 3.

⁶² A. W. Hayne, ‘Online Privacy Policies: Contracting Away Control Over Personal Information?’, Dickinson Law Review, 111/3 (2007), 597. Despite the lack of US Federal level data protection law, the FTC in practice has long taken the role of US privacy commando and effectively required privacy policies.

⁶³ GDPR, Chapter III.

⁶⁴ GDPR, art 6.

Of the 13 models we studied, 9 out of 13 models referenced the GDPR in their privacy policies (if they had one). However, 7 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 Notice⁶⁵.

All of the T2T models we studied 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 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 rights⁶⁶.

11 out of 13 models studied 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” since 2014 in the wake of *Google Spain*⁶⁷ 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. 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 Google⁶⁸, it is extremely difficult to work out what

⁶⁵ LENSEA App California Notice at Collection and Privacy Notice, 10 July 2023, <http://tos.LENSA-ai.com/CCPAPP>, last accessed at 01.04.2024.

⁶⁶ GDPR, art 14 and Chapter III.

⁶⁷ Case C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González Judgment of the Court (Grand Chamber), 13 May 2014.

⁶⁸ 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 papers findings) ownership in outputs belongs to the user. The extra terms at <https://policies.google.com/terms/generative-ai> thus became defunct from that data.

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 (FTC) in the US, 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.⁶⁹ It is fair to say that DP compliance by large models is still very much a work in progress.

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@navercorp.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.

⁶⁹ 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, 29.04.2024 at <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-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.

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

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 hope to examine in more detail a number of points we either missed as significant, or did not have capacity to explore in full, or at all. 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 (eg 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

landgrab 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 outputs⁷⁰. 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 data protection 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 millennium⁷¹. 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 Store⁷² muddies the waters somewhat). 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

⁷⁰ As noted above at 2.1, some but not all providers did at least demand back from users a non-exclusive license over outputs.

⁷¹ For a general history of how this paradigm emerged, see L. Edwards “With Great Power Comes Great Responsibility”: The Rise of Platform Liability’ in: L. Edwards (ed), *Law, Policy and the Internet*, Hart Publishing, 2019; and M. Husovec ‘Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules’, *Berkeley Technology Law Journal*, 38/3 (2023), 883–920.

⁷² See: OpenAI, *Introducing the GPT Store*, January 10, 2024, <https://openai.com/index/introducing-the-gpt-store>, opened 10.1.2024, last accessed 02.05.2024.

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 tools⁷³. 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” (art 6).

This description simply does not match foundation models. The only content the user supplies is the prompt or other input (eg 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 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 content⁷⁴.” [italic added]. Hacker et al agree, pointing out that with LLMs, the relevant content is decidedly not provided by the user, but by the LLM itself⁷⁵.

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 large language models might sneak in to the DSA hierarchy as search engines, since Very Large Online Search Engines (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 Gemini⁷⁶)).

⁷³ See: DSA arts 3-6. From here on we will use the DSA only as the legal paradigm for discussing platformisation.

⁷⁴ B. Botero Arcila “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, (2023), 453-488.

⁷⁵ Hacker et al, supra n 60.

⁷⁶ See: 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 2.5.2024. In fact the Commission warned they would be opening action against Microsoft to require information concerning their use of generative AI models in Bing, see <https://digital->

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 natures⁷⁷. 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)⁷⁸). 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. At present, model providers have their cake and eat it; 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,

strategy.ec.europa.eu/en/news/commission-compels-microsoft-provide-information-under-digital-services-act-generative-ai-risks, last accessed 22.05.2024.

⁷⁷ See: Husovec, *supra* n 74.

⁷⁸ J. P. Quintais, N. Appelman, R. Ó Fathaigh, 'Using Terms and Conditions to apply Fundamental Rights to Content Moderation', *German Law Journal*, 24/5 (2023), 881 – 911.

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 PRC. 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 West⁷⁹. 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 regulator⁸⁰. In order to be approved, developers need to provide documentation of the model, including T&C, to the regulator (CAC). This pre-market 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 pre-market 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 Institute⁸¹. 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 control⁸².

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

⁷⁹ A. H. Zhang, *High Wire: How China Regulates Big Tech and Governs Its Economy*, (OUP, Oxford 2024), 291.

⁸⁰ Announcement of the Cyberspace Administration of China on the release of registered information on generative artificial intelligence services, 02.04.2024, https://www.cac.gov.cn/2024-04/02/c_1713729983803145.htm, last accessed 10.04.2024.

⁸¹ See: Safe before sale: Learnings from the FDA’s model of life sciences oversight for foundation models, Ada Lovelace Institute, 2023 at <https://www.adalovelaceinstitute.org/report/safe-before-sale/>, last accessed 3.5.2024.

⁸² See: I. Szpotakowski, ‘Self-regulation and pre-market licensing as methods of controlling generative artificial intelligence in China’, In: Piotr Grzebyk (Ed.), *New Technologies as a challenge for Polish legal system*, (Scholar Publishing Press, Warszawa 2024).

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.

References

1. Abdoullaev A., 'AI Bible: why Generative AI bubble is to burst and Interactive AI is to rise...', LinkedIn (2023), <https://www.linkedin.com/pulse/why-generative-ai-bubble-burst-2024-world-embeddings-vs-abdoullaev-14ycf/>, last accessed 23.02.2024.
2. Abiri G., Huang Yue, 'A Red Flag? China's Generative AI Dilemma', *Harvard Journal of Law & Technology*, 37(2023), 7.
3. 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 01.03.2024.
4. Announcement of the Cyberspace Administration of China on the release of registered information on generative artificial intelligence services, 02.04.2024, https://www.cac.gov.cn/2024-04/02/c_1713729983803145.htm, last accessed 10.04.2024.
5. Artificial Intelligence Act (P9_TA(2024)0138), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf, last accessed 29.03.2024.
6. Ben-Shahar O. and Porat A., *Personalized Law: Different Rules for Different People* (Oxford University Press, Oxford 2021).
7. Betkier M., *Privacy Online, Law and the Effective Regulation of Online Services* (Intersentia, Bristol 2019).
8. Birhane A., Kasirzadeh A., Leslie D., Wachter S., 'Science in the age of large language models', *Nature Reviews Physics*, 5(2023), 277-280.
9. Blueprint For An AI. Bill Of Rights Making Automated Systems Work For The American People October 2022. <https://www.whitehouse.gov/wp-content/uploads/2022/10/blueprint-for-an-ai-bill-of-rights.pdf>, last accessed 01.03.2024.
10. Botero Arcila B. '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 (2023), 453-488.
11. Brown T. B. et al., 'Language Models are Few-Shot Learners', In: *Advances in Neural Information Processing Systems*, 34th Conference on Neural Information Processing Systems, 33 (NeurIPS, Vancouver, 2020), 1877-1901.
12. Case C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de DaT&C (AEPD) and Mario Costeja González Judgment of the Court (Grand Chamber), 13 May 2014.

13. Case C-252/21 Meta v Bundeskartellamt Judgment of the Court (Grand Chamber) of 4 July 2023.
14. Celeste E., 'Digital punishment: social media exclusion and the constitutionalising role of national courts', *International Review of Law, Computers & Technology* 35/2 (2021), 162–84.
15. Celeste E., 'Terms of service and bills of rights: new mechanisms of constitutionalisation in the social media environment?', *International Review of Law, Computers & Technology*, 33/2 (2019), 122–38.
16. Cerullo M., 'AI-powered "robot" lawyer won't argue in court after jail threats', CBS News, 26.01.2023, <https://www.cbsnews.com/news/robot-lawyer-wont-argue-court-jail-threats-do-not-pay/>, last accessed 01.05.2023.
17. Chalkidis I. and Brandl S., 'Llama meets EU: Investigating the European Political Spectrum through the Lens of LLMs' (2024), arxiv.org/pdf/2403.13592, last accessed 20.03.2024.
18. Chinese 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) 2023.
19. Christ W. Oct 23, 2023, Twitter (X) <https://twitter.com/WolfieChristl/status/1716536697666580658?t=UCGyqLKsuCoP07E1e30tLA&s=03>, last accessed 09.04.2024.
20. CLAUDETTE Machine Learning Powered Analysis of Consumer Contracts and Privacy Policies, claudette.eui.eu/index.html, last accessed 20.03.2024.
21. [Clio Privacy Policy - Updates Effective 24 May, 2023](#). last accessed 24.07.2023.
22. Cobbe J. et al., 'Understanding accountability in algorithmic supply chains', FAccT '23: 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, June 2023 DOI: <https://doi.org/10.1145/3593013.3594073>, last accessed 01.05.2024.
23. Code.europa.eu, code.europa.eu/p2b/contrib-versions, last accessed 01.05.2024.
24. DeepL Terms and Conditions <https://www.deepl.com/pro-license>, last accessed 04.12.2023.
25. Department for Science, Innovation & Technology, *Consultation outcome A pro-innovation approach to AI regulation: government response Updated 6 February 2024*, 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 01.03.2024.

26. 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').
27. Doctorow's C., 'Pluralistic: In defense of bureaucratic competence (23 Oct 2023), <https://pluralistic.net/2023/10/23/getting-stuff-done/>, last accessed 23.02.2024.
28. Edwards L. (Ed.), *Law, Policy and the Internet*, Hart Publishing 2019.
29. Edwards L. and Brown I. 'Privacy, Law, Code and Social Networking Sites' in Brown I. (Ed.) *Research Handbook On Governance Of The Internet* (OUP, Oxford 2013).
30. Efferenn F., '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 24.03.2023.
31. Ferrandis C.M., 'The BigScience OpenRAIL-M License', <https://bigscience.huggingface.co/blog/bigscience-openrail-m>, last accessed 01.05.2023.
32. Filgueiras F., 'The politics of AI: democracy and authoritarianism in developing countries', *Journal of Information Technology & Politics*, 19:4 (2022), 449-451.
33. Firmsy | Terms and Conditions; Privacy Policy - Firmsy. <https://firmsy.com/terms/> last accessed 24.07.2023.
34. Friedmann D., 'Creation and Generation Copyright Standards', *NYU Journal of Intellectual Property & Entertainment Law*, forthcoming 2024, SSRN, https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4770924_code1572004.pdf?abstractid=4770924&mirid=1&type=2, last accessed 01.05.2024.
35. Frosio G., 'Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity?', in: Bonadio E., Sganga C. (Eds.), *A research agenda for EU copyright law*, Edward Elgar Publishing 2023.
36. Gal U. 'ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned', *The Conversation*, February 8, 2023 at <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283>, last accessed 29.04.2024.
37. Goanta C., 'The New Social Media: Contracts, Consumers, and Chaos', *Iowa Law Review Online* 108 (2023). 118-30.
38. Google Terms of Service - Privacy & Terms - Google <https://policies.google.com/terms> last accessed 04.12.2023.

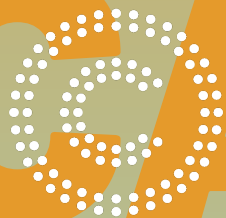
39. Guadamuz A., 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs', *GRUR International*, 73(2), 2024, 111-127.
40. Gunningham N., 'Private Ordering, Self-Regulation and Futures Markets: A Comparative Study of Informal Social Control', *Law & Policy*, 13/4, (1991), 297-326.
41. Hacker P. et al. "Regulating ChatGPT and other Large Generative AI Models" FAccT '23, June 12-15, 2023, Chicago, IL, USA, at <https://arxiv.org/abs/2302.02337>, last accessed 04.03.2024.
42. Harvey AI, Harvey | Generative AI for Elite Law Firms, <https://www.harvey.ai/>, last accessed 04.12.2023.
43. Hayne A. W., 'Online Privacy Policies: Contracting Away Control Over Personal Information?', *Dickinson Law Review*, 111/3(2007), 597-624.
44. Helberger N. and Samuelson P. 'Will the EU's Digital Services Act Become a Global Transparency Regime?', In: *Verfassungsblog: From the DMCA to the DSA—A Transatlantic Dialogue on Online Platform Regulation and Copyright*, forthcoming March 2024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4783341, last accessed 29.04.2024.
45. Helberger N., 'Generative AI in Media & Journalism: Think Big, But Read the Small Print First', Medium, 18.07.2023 <https://generative-ai-newsroom.com/generative-ai-in-media-journalism-think-big-but-read-the-small-print-first-375f2ecb1256>, last accessed 24.07.2023.
46. Husovec M., 'Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules', *Berkeley Technology Law Journal*, 38/3(2023), 883-920.
47. Jensen C. and Potts C., 'Privacy policies as decision-making tools: an evaluation of online privacy notices' In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM*, (2004) 471-478.
48. Land M. K., 'The Problem of Platform Law: Pluralistic Legal Ordering on Social Media', in: P. Schiff Berman (Ed.), *The Oxford Handbook of Global Legal Pluralism*, (Oxford University Press, Oxford 2020), 974-994.
49. LENSEA App California Notice at Collection and Privacy Notice, 10 July 2023, T&C.LENSA-ai.com/CCPAPP, last accessed 24.07.2023.
50. LENSEA Terms of Use' <https://T&C.LENSA-ai.com/terms-dec-2-2022>, last accessed 04.12.2023.

51. LNB News, CO launches consultation on accuracy of generative AI models, 12.04.2024, <https://www.lexisnexis.co.uk/legal/news/ico-launches-consultation-on-accuracy-of-generative-ai-models>, last accessed 01.05.2024.
52. Lomas N., 'Italy orders ChatGPT blocked citing data protection concerns', TechCrunch 31.3.2023, <https://techcrunch.com/2023/03/31/chatgpt-blocked-italy/>, last accessed at 1.04.2024.
53. Loos M. and Luzak J., 'Wanted: a Bigger Stick. On Unfair Terms in Consumer Contracts with Online Service Providers', *Journal of Consumer Policy* 39/ 1(2016), 63–90.
54. Marique E. and Marique Y., 'Sanctions on Digital Platforms: Beyond the Public-Private Divide', *Cambridge International Law Journal* 8/2 (2019), 258–81.
55. McDonald A. M. and Faith Cranor L., 'The Cost of Reading Privacy Policies', *A Journal of Law and Policy for the Information Society*, 4/3 (2008), 543–568.
56. McDonald A. M. and Faith Cranor L., 'The Cost of Reading Privacy Policies', *A Journal of Law and Policy for the Information Society*, 4/3 (2008), 543–568.
57. Micklitz H., 'Unfair Terms in Consumer Contracts' in, Reich N. et al (Eds.), *European Consumer Law*, 2nd, (Intersentia, Antwerp 2014), 125–64.
58. Micklitz H.W., Palka P., Panagis Y., 'The Empire Strikes Back: Digital Control of Unfair Terms of Online Services', *Journal of Consumer Policy* 40/3 (2017), 367–88.
59. Midjourney Terms of Service <https://docs.midjourney.com/docs/terms-of-service>, last accessed 04.12.2023.
60. MUSICSTREAM: Music Culture in the Age of Streaming, <https://musicstreamproject.com>, last accessed 20.03.2024.
61. NAVER CLOUD PLATFORM Terms of Service <https://www.ncloud.com>, last accessed 04.12.2023.
62. 'NightCafe Creator Terms of Service' <https://nightcafe.studio/policies/terms-of-service>, last accessed 04.12.2023.
63. NoYB, 'ChatGPT provides false information about people, and OpenAI can't correct it', 29.04.2024, <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it>, last accessed 01.05.2024.
64. Obar J. A. and Oeldorf-Hirsch A., 'The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services', *Information, Communication & Society* 23/1(2020), 128–47.
65. Pałka P. and Lippi M., 'Big Data Analytics, Online Terms of Service and Privacy Policies', in R. Vogl et al. *Research Handbook on Big Data Law* (Edward Elgar Publishing, Cheltenham 2021).

66. Palka P. J., 'Terms of Service are not Contracts – Beyond Contract Law in the Regulation of Online Platforms', in Grundmann S. (Ed.), *European Contract Law in the Digital Age*, vol. 3, (Intersentia, Bristol 2018), 135–62.
67. Pałka P., 'Terms of Injustice', *West Virginia Law Review*, 126 (2023), 134–184.
68. Pratim Ray P., 'ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope', *Internet of Things and Cyber-Physical Systems*, 3, (2023), 121–154.
69. Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset, <https://privacypolicies.cs.princeton.edu/>, last accessed 24.07.2023.
70. Privacy Policy – Lawgeex, <https://www.lawgeex.com/privacy-policy/>, last accessed 24.07.2023.
71. Privacy Policy & Data Protection Policy | Kira Systems, <https://kirasystems.com/privacy/>, last accessed 24.07.2023.
72. Privacy Policy (donotpay.com), <https://donotpay.com/learn/privacy-policy/>, last accessed 24.07.2023.
73. Privacy Policy | Ontra, <https://www.ontra.ai/privacy-policy/>, last accessed 24.07.2023.
74. Puckett D.A., 'Terms of Service and the Computer Fraud and Abuse Act: A Trap for the Unwary?', *Oklahoma Journal of Law and Technology* 7/1(2011), 1–26.
75. Quintais J. P., Appelman N., Ó Fathaigh R., 'Using Terms and Conditions to apply Fundamental Rights to Content Moderation', *German Law Journal*, 24/5 (2023), 881 – 911.
76. 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).
77. Roberts H. et al., 'Global AI governance: barriers and pathways forward', *International Affairs*(2024), 1–12.
78. Roth E., 'DoNotPay is launching an AI chatbot that can negotiate your bills', The Verge, 13.12.2022, <https://www.theverge.com/2022/12/13/23505873/donotpay-negotiate-bills-ai-chatbot>, last accessed 02.05.2024.
79. Runway Terms of Use <https://runwayml.com/terms-of-use/>, last accessed 04.12.2023.
80. Rustad M. and Onufrio M., 'Reconceptualizing Consumer Terms of Use for a Globalized Knowledge Economy', *University of Pennsylvania Journal of Business Law* 14/4 (2012), 1085–1190.
81. Samuelson P., 'Thinking About Possible Remedies in the Generative AI Copyright Cases', *Communications of the ACM* (2024), SSRN,

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4770671, last accessed 11.04.2024.
82. Stable Diffusion License on Dezgo.Com <https://dezgo.com/>, last accessed 04.12.2023.
83. Suzor N., 'The responsibilities of platforms: A new constitutionalism to promote the legitimacy of decentralized governance', *AoIR Selected Papers of Internet Research*, 6 (2016), 1-4.
84. Szpotakowski I., 'Self-regulation and pre-market licensing as methods of controlling generative artificial intelligence in China', In: Piotr Grzebyk (Ed.), *New Technologies as a challenge for Polish legal system*, (Scholar Publishing Press, Warszawa 2024).
85. Taylor L., *Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector*, *Philosophy & Technology*, 34 (2021), 897-922.
86. Terms & Conditions | Colossyan Creator <<https://www.colossyan.com/terms>> last accessed 04.12.2023.
87. Terms of Service (donotpay.com), <https://donotpay.com/learn/terms-of-service/>, last accessed 24.07.2023.
88. Terms of Service | Kira Systems, <https://kirasystems.com/terms-of-service/>, last accessed 24.07.2023.
89. Terms of Service | Ontra, <https://www.ontra.ai/terms-of-service/>, last accessed 24.07.2023.
90. Terms of Service | Simplified <https://simplified.com/termservice>, last accessed 04.12.2023.
91. Terms of Service | Synthesia, <https://www.synthesia.io/terms/terms-of-service>, last accessed 04.12.2023.
92. Terms of Use - Lawgeex, <https://www.lawgeex.com/terms-of-use/>, last accessed 24.07.2023.
93. Terms of Use - March 2023 <https://openai.com/policies/mar-2023-terms> last accessed 04.12.2023.
94. The OCW-funded Gravitation Program Public Values in the Algorithmic Society (AlgoSoc), algosoc.org/programme, last accessed 20.03.2024.
95. T&CBack. The terms-of-service tracker, <https://T&Cback.org/> , last accessed 24.07.2023.
96. Vale SB., 'Training Large Generative AI Models Based On Publicly Available Personal Data: A GDPR Conundrum That The AI Act Could Solve', DigiCon, <https://digi-con.org/training-large-generative-ai-models-based-on-publicly-available-personal-data-a-gdpr-conundrum-that-the-ai-act-could-solve/>, last accessed 04.03.2024.

97. Veale M., Matus K., Gorwa R., 'AI and Global Governance: Modalities, Rationales, Tensions', *The Annual Review of Law and Social Science*, 19 (2023), 255-275.
98. Vizir - The Terms of Service and Privacy Agreement used to be available on the website <https://www.vizir.co>, last accessed 20.03.2024.
99. Wang Xiao, 'Copyright Protection for 'AI-Generated' Images', *GRUR International*, 73/4 (2024), 360-368.
100. Wauters E., Lievens E., and Valcke P., '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 (2014): 254-94.
101. Weidinger L. et al., 'Ethical and social risks of harm from Language Models', *DeepMind* (2021), arxiv.org/abs/2112.04359, last accessed 29.03.2024.
102. Wenxin Large Model Documentation - Service Agreement | Baidu Intelligent Cloud Documents' <https://wenxin.baidu.com/AIDP/wenxin/YI6th25am>, last accessed 04.12.2023.
103. Zhang A. H., *High Wire : How China Regulates Big Tech and Governs Its Economy*, (OUP, Oxford 2024).
104. Zhang D. et al., 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions', Sept 2023, arqiv, <https://arxiv.org/pdf/2307.03941>, last accessed 04.03.2024.



CREATE

Centre for Regulation of the Creative Economy

School of Law / University of Glasgow

10 The Square

Glasgow G12 8QQ

www.create.ac.uk

2024/05 DOI: 10.5281/zenodo.11276105

CC BY-SA 4.0

In collaboration with:

