
Standardization of Behavioral Use Clauses is Necessary for the Adoption of Responsible Licensing of AI

Openness is a tenet of scientific research and plays an important role in the development of new technologies. By making assets available to third parties, scientific results can be verified, and systems can be interrogated, tested and audited (Resnick, 2006; von Krogh and Spaeth, 2007). In AI, significant advances have been made possible thanks to the open sharing of data, code, models, and applications (Gokaslan et al., 2019; 2023; Workshop et al., 2022; Raffel et al., 2020; Rombach et al., 2022; Touvron et al., 2023). The fact that researchers and developers can use, modify or extend what others have built enables an important form of AI decentralization and supports accessibility.

However, openness has also come with significant tensions, particularly for pretrained (so-called *base* or *foundation*) models. On the one hand, these models can be used in a variety of domains, often with little or no finetuning (Radford et al., 2019; Brown et al., 2020). On the other, this versatility means that they can be used by different actors in ways that are not aligned with the applications intended by their creators (Lee et al., 2023). Some of these uses may be overtly harmful (e.g., generating content to deceive or harass a person/people) and others may present unintended higher risks (e.g., generating diagnoses that may sometimes be inaccurate or biased, extracting PII from training data) (Fergusson et al., 2023; Nasr et al., 2023; Brundage et al., 2022). Not only is there potential for harm, in decentralized systems downstream uses can present challenges for accountability and recourse as there is no clear defacto or centralized authority making oversight decisions (Cooper et al., 2022; 2023).

Usage restrictions on contractual agreements. The release of assets by private organizations is usually accompanied by a provider-user contractual agreement. Users must agree to provider specified terms for unhindered and continued access to the service. While traditionally such terms were geared towards compliance with laws, mitigating legal risks and protecting intellectual property, some AI providers have begun to include additional clauses that govern usage. For example, Open AI’s policies disallow generation of content for dissemination in electoral campaigns (OpenAI, 2024), Microsoft’s FaceAPI services are subject to “specific license terms” and are only available in limited access to “customers managed by Microsoft” (Microsoft, 2022).

To release or not to release. Not all researchers or research teams have the resources to create customized legal agreements for the AI models or source code they would like to release. The choice often becomes to release the code with no restrictions or to not release it at all. This forces the creator to decide between openness and democratization on one hand, and responsible use on the other.

While tools for software licensing do exist (Bretthauer, 2001), historically the most frequently adopted licenses (e.g., MIT, Apache) do not contain restrictions on how code, models or applications are to be *used*. Adapting licenses as a tool for responsibly releasing AI software was proposed by Contractor et al. (Contractor et al., 2022b). These licenses contain *behavioral use clauses* that enable software and models to be released with restrictions around how they are *used*. The paper, proposed that these licenses could be implemented to complement existing responsible AI guidelines.

Licenses with Behavioral-use Clauses. Over the past five years, licenses with behavioral use clauses (BUC) have been gaining adoption at an increasingly rapid rate (see Fig. 1). In this paper, we refer to responsible AI licenses as a broad category of licenses that incorporate BUC.¹ Licenses that allow adaptation and reuse of software, models, data, or applications can be made dependent on behavioral use conditions. However, each different type of asset has their own idiosyncracies.

Most licenses with BUCs are based on copyright and contractual agreements that are enforced by the licensor using a given dispute resolution mechanism (e.g., a court system, arbitration, mediation). The penalties for violation of a license agreement fall into two categories: (1) injunction preventing the licensee to use the licensor’s intellectual property, and (2) monetary damages, which will depend on the facts underlying each given violation. Even without aggressive enforcement of license terms, the existence of the contractual restrictions serve as a deterrent. Corporations have legal teams in place to ensure that they comply with the agreements that they sign or adopt, including any restrictions embedded in open source software terms.

A study by OpenFutures of 39,000 repositories found a clear

¹Licenses using the acronym ‘RAIL’ are specific variant of the broader class of responsible AI licenses.

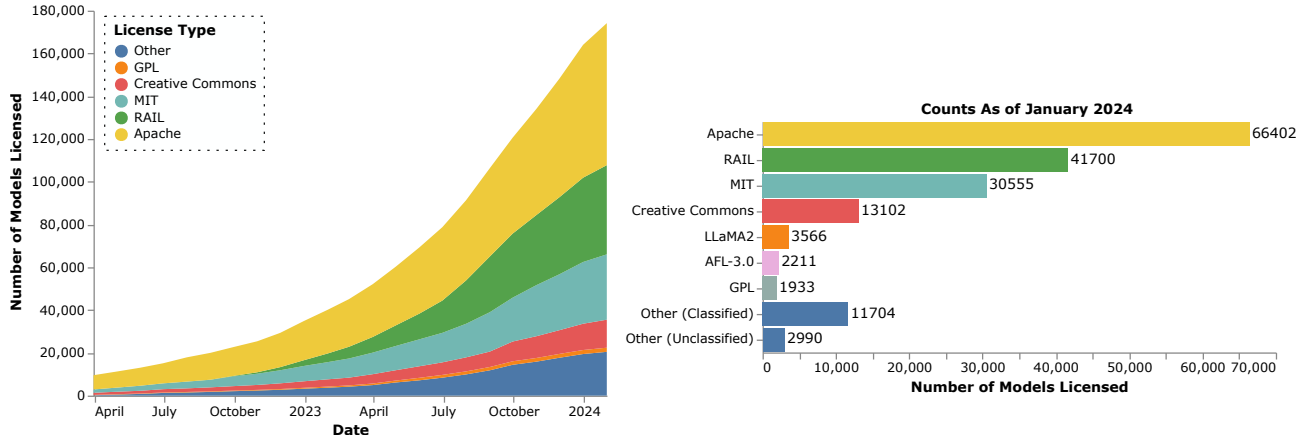


Figure 1. Adoption of Licenses with Behavioral Use Clauses. The number of repositories by license type on the HuggingFace model hub. As of January 2024, 41,700 RAIL licensed repositories and 3,566 LLaMA2 licensed repositories existed, both of these licenses include behavioral-use clauses.

trend towards the adoption of responsible AI licenses (Keller and Bonato, 2023). OpenRAIL licenses (Contractor et al., 2022a), a specific variant of RAIL licenses, were the second most used license category. To date, such licenses have primarily been applied to AI models.

For example, BLOOM (Workshop et al., 2022) is a large parameter multilingual language model, and accompanying BigScience OpenRAIL license enables derivative uses but restricts applications amongst others that violate laws, generate or disseminate verifiably false information, or predict if an individual will commit fraud/crime. Stable diffusion (Rombach et al., 2022) “provided weights under a license to prevent misuse and harm as informed by the model card (Mitchell et al., 2019), but otherwise remains permissive.” Subsequently the LLaMA 2 (Touvron et al., 2023) and FALCON (Almazrouei et al., 2023) models were released with intersecting behavioral use restrictions, yet FALCON has a smaller set of clauses than the others.

The use of responsible AI licenses is not restricted only to foundation models, Robotics platforms (GRID) (Vemprala et al., 2023), edge IoT systems (Janapa Reddi et al., 2023) and medical sensors (Liu et al., 2023) that use AI components have also adopted similar clauses. In the realm of training data, similar approaches to licensing have also been the subject of discussion (Li et al., 2023) and experimentation. For instance, AI2 created the ImpACT license to apply broadly to ML artifacts that include both models and data (Allen Institute for AI, 2024) and used an ImpACT license for the DOLMA dataset release (Soldaini et al., 2023).

Contributions. While the growth of interest in, and adoption of, these licenses is very apparent, there is no single standard license. In this paper we explore the reasons for the trend towards behavioral use licensing, the proliferation of different licenses and choice of clauses, and the need

for standardization. We perform a quantitative analysis of the licenses used in over 170,000 model repositories and highlight the growing trend toward adopting responsible AI licenses. We also qualitatively evaluate the similarities and differences between the specific license clauses included in these artifacts. We then report on semi-structured interviews conducted with researchers who have released high-profile AI models and software with responsible AI licenses (across computer vision, natural language, and robotics). **We take the position that responsible AI licenses need standardization to avoid confusing users or diluting their impact. At the same time, customization of behavioral restrictions is necessary and appropriate in some contexts (e.g., medical domains) and can be supported by tooling.** In the full version of our paper we:

- Provide a synthesized review of regulation and licensing for AI as distinct yet related tools for AI governance.
- Detail our mixed-methods study of AI licenses, which includes interviews with license adopters, and quantitative and qualitative analysis of licenses and their clauses.
- Discuss our findings concerning how responsible AI licensed are being *adopted* and *adapted*, which ground our recommendations for license standardization, suggested community-oriented license generator, and additional tooling to ease the adoption and adaptation of responsible AI licenses.

References

Allen Institute for AI. AI2 ImpACT Licenses, 2024. URL <https://allenai.org/impact-license>.

- Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Merouane Debbah, Etienne Goffinet, Daniel Heslow, Julien Launay, Quentin Malartic, et al. Falcon-40B: an open large language model with state-of-the-art performance. *Findings of the Association for Computational Linguistics: ACL*, 2023:10755–10773, 2023.
- David Bretthauer. Open Source Software: A History. *Published Works*, 7, 2001.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Nee-lakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*, 2020.
- Miles Brundage, Katie Mayer, Tyna Eloundou, Sandhini Agarwal, Steven Adler, Gretchen Krueger, Jan Leike, and Pamela Mishkin. Lessons learned on language model safety and misuse, 2022. URL <https://openai.com/research/language-model-safety-and-misuse>.
- Danish Contractor, Carlos Muñoz Ferrandis, Jenny Lee, and Daniel McDuff. From RAIL to Open RAIL: Topologies of RAIL Licenses, 2022a. URL <https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses>.
- Danish Contractor, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 778–788, 2022b.
- A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’22*, page 864–876, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522. doi: 10.1145/3531146.3533150.
- A. Feder Cooper, Katherine Lee, James Grimmelmman, Daphne Ippolito, Christopher Callison-Burch, Christopher A. Choquette-Choo, Niloofar Mireshghallah, Miles Brundage, David Mimno, Madiha Zahrah Choksi, Jack M. Balkin, Nicholas Carlini, Christopher De Sa, Jonathan Frankle, Deep Ganguli, Bryant Gipson, Andres Guadamuz, Swee Leng Harris, Abigail Z. Jacobs, Elizabeth Joh, Gautam Kamath, Mark Lemley, Cass Matthews, Christine McLeavey, Corynne McSherry, Milad Nasr, Paul Ohm, Adam Roberts, Tom Rubin, Pamela Samuelson, Ludwig Schubert, Kristen Vaccaro, Luis Villa, Felix Wu, and Elana Zeide. Report of the 1st Workshop on Generative AI and Law. *arXiv preprint arXiv:2311.06477*, 2023.
- Grant Fergusson, Caitriona Fitzgerald, Chris Frascella, Megan Iorio, Tom McBrien, Calli Schroeder, Ben Winters, and Enid Zhou. Generating Harms: Generative AI’s Impact & Paths Forward. Technical report, Electronic Privacy Information Center, 2023.
- Aaron Gokaslan, Vanya Cohen, Ellie Pavlick, and Stefanie Tellex. Openwebtext corpus, 2019.
- Aaron Gokaslan, A. Feder Cooper, Jasmine Collins, Landan Seguin, Austin Jacobson, Mihir Patel, Jonathan Frankle, Cory Stephenson, and Volodymyr Kuleshov. CommonCanvas: An Open Diffusion Model Trained with Creative-Commons Images. *arXiv preprint arXiv:2310.16825*, 2023.
- Vijay Janapa Reddi, Alexander Elum, Shawn Hymel, David Tischler, Daniel Situnayake, Carl Ward, Louis Moreau, Jenny Plunkett, Matthew Kelcey, Mathijs Baaijens, et al. Edge impulse: An mlps platform for tiny machine learning. *Proceedings of Machine Learning and Systems*, 5, 2023.
- Paul Keller and Nicolò Bonato. Growth of responsible AI licensing. Analysis of license use for ML models published on. *Open Future*, feb 7 2023. <https://openfuture.pubpub.org/pub/growth-of-responsible-ai-licensing>.
- Katherine Lee, A. Feder Cooper, and James Grimmelmman. Talkin’ ’Bout AI Generation: Copyright and the Generative-AI Supply Chain. *arXiv preprint arXiv:2309.08133*, 2023.
- Hanlin Li, Nicholas Vincent, Yacine Jernite, Nick Merrill, Jesse Josua Benjamin, and Alek Tarkowski. Can licensing mitigate the negative implications of commercial web scraping? In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing, CSCW ’23 Companion*, page 553–555, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701290. doi: 10.1145/3584931.3611276. URL <https://doi.org/10.1145/3584931.3611276>.

- Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Soumyadip Sengupta, Shwetak Patel, Yuntao Wang, and Daniel McDuff. rPPG-Toolbox: Deep Remote PPG Toolbox. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.
- Microsoft. Responsible AI investments and safeguards for Facial Recognition, 2022. URL <https://azure.microsoft.com/en-us/updates/facelimitedaccess/>.
- Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.
- Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. Scalable Extraction of Training Data from (Production) Language Models. *arXiv preprint arXiv:2311.17035*, 2023.
- OpenAI. Usage policies, 2024. URL <https://openai.com/policies/usage-policies>.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21:1, 2020.
- David B. Resnick. Openness versus Secrecy in Scientific Research Abstract. *Episteme*, pages 135 – 147, February 2006.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- Luca Soldaini, Rodney Kinney, Akshita Bhagia, Dustin Schwenk, David Atkinson, Russell Authur, Khyathi Chandu, Jennifer Dumas, Li Lucy, Xinxu Lyu, et al. Dolma: An Open Corpus of 3 Trillion Tokens for Language Model Pretraining Research. *Allen Institute for AI, Tech. Rep*, 2023.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- Sai Vemprala, Shuhang Chen, Abhinav Shukla, Dinesh Narayanan, and Ashish Kapoor. Grid: A platform for general robot intelligence development. *arXiv preprint arXiv:2310.00887*, 2023.
- Georg von Krogh and Sebastian Spaeth. The open source software phenomenon: Characteristics that promote research. *The Journal of Strategic Information Systems*, 16(3):236–253, 2007. ISSN 0963-8687. doi: <https://doi.org/10.1016/j.jsis.2007.06.001>. URL <https://www.sciencedirect.com/science/article/pii/S096386870700025X>.
- BigScience Workshop, Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, et al. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.