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

Daniel McDuff [1 2] **Tim Korjakow** [1 3] **Scott Cambo** **Jesse Josua Benjamin** [1 4] **Jenny Lee** [1] **Yacine Jernite** [5] **Carlos Muñoz Ferrandis** [6] **Aaron Gokaslan** [7] **Alek Tarkowski** [8] **Joseph Lindley** [1 4] **A. Feder Cooper** [9 7] **Danish Contractor**

Abstract

Growing concerns over negligent or malicious uses of AI have increased the appetite for tools that help manage the risks of the technology. In 2018, licenses with behaviorial-use clauses (commonly referred to as Responsible AI Licenses) were proposed to give developers a framework for releasing AI assets while specifying their users to mitigate negative applications. As of the end of 2023, on the order of 40,000 software and model repositories have adopted responsible AI licenses licenses. Notable models licensed with behavioral use clauses include BLOOM (language) and LLaMA2 (language), Stable Diffusion (image), and GRID (robotics). This paper explores why and how these licenses have been *adopted*, and why and how they have been *adapted* to fit particular use cases. We use a mixed-methods methodology of qualitative interviews, clustering of license clauses, and quantitative analysis of license adoption. Based on this evidence 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 also appropriate in some contexts (e.g., medical domains). We advocate for "standardized customization" that can meet users' needs and can be supported via tooling.

[*]Equal contribution [1]Responsible AI Licenses [2]University of Washington, USA [3]Technical University of Berlin [4]Lancaster University [5]Hugging Face [6]Alinia AI [7]Cornell University [8]Open Future Foundation [9]The Center for Generative AI, Law, and Policy Research. Correspondence to: Daniel McDuff <dmcduff@uw.edu>, Danish Contractor <danishcontractor@outlook.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

1. Introduction

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 & 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 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

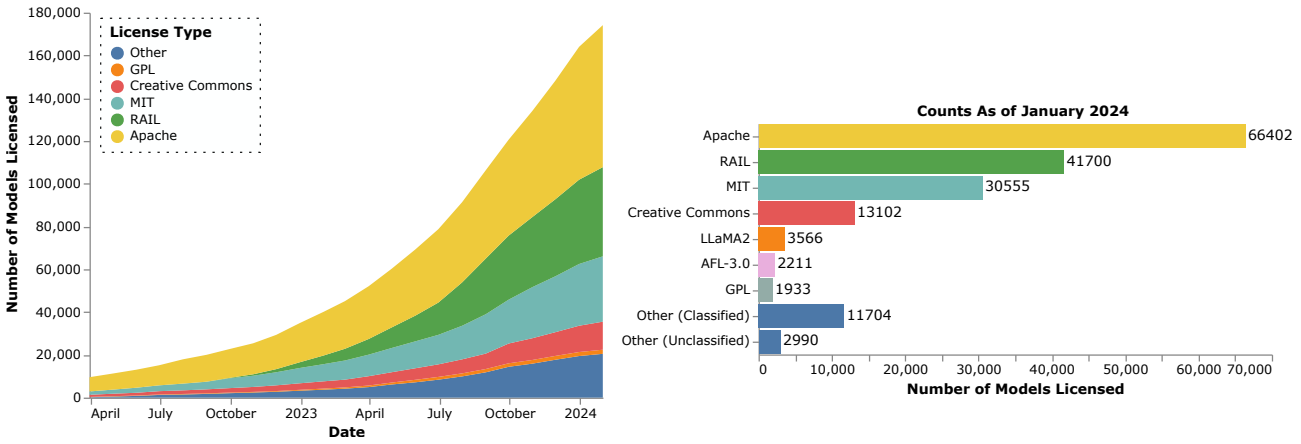

Figure 1. **Adoption of Licenses with Behavioral Use Clauses.** The number of repositories by license type on the HuggingFace model hub. As of January 2024, 41,700 RAIL licensed repositories and 3,566 LLaMA2 licensed repositories existed, both of these licenses include behavioral-use clauses.

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 Behaviorial-use Clauses. Over the past five years, licenses with behavioral use clauses (BUC) have been gaining adoption at an increasely 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 ideosyncracies.

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 trend towards the adoption of responsible AI licenses (Keller & 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., 2019b), 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)

[1]Licenses using the acronym 'RAIL' are specific variant of the broader class of responsible AI licenses.

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., 2023a) 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 apparant, 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 qualtitatively 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.**

2. Regulation and Licensing for AI

Regulation and licensing of AI systems is difficult because of their complex supply chains and the uncertainty about how copyright rules apply to AI systems (Lee et al., 2023). One must consider the licensing of the data, the code, the models, and the machine learning libraries used to train the models. New considerations are being raised as the technology is developed; for example, courts are now examining the relationship between the licensing of the data and the models where they have not in the past. Outcome-based regulation of relatively simple classifiers has proven difficult (Cooper et al., 2022), let alone regulation for large foundation models. Some prominent academics believe that training AI models on copyrighted data is fair use under current US copyright law (Lemley & Casey, 2021; Klosek & Blumenthal, 2024), while others argue that there is likely not a blanket fair-use rationale for all possible systems and their possible outputs (Lee et al., 2023; Samuelson, 2023; Sag, 2023). Courts and legislators, if asked existential questions about whether large-scale AI models should exist, may realistically decide that using copyrighted data to train AI models does not constitute copyright infringement, and that the fair use doctrine may be invoked to shield such use.[2]

How can regulation help?: Regardless of the currently-unresolved intellectual property regime governing AI systems, the aim of regulation is to provide an enforceable mechanism to govern how AI is or is not used – for example, by focusing on data privacy. Many countries, states, and cities are using regulations to avoid undesired uses of AI. For instance, the city of San Francisco banned the use facial regulation in CCTV cameras, and imposed a series of certifications for the use of self-driving cars in its streets. In Europe, the General Data Protection Regulation GDPR[3] requires data collectors to state the purpose for which data is being collected, to obtain consent, and then to collect only the minimum amount of data required to accomplish the task for which consent was received. Additionally, it imposes requirements on data processors to be able to verifiably remove data when consent is revoked.

The recently proposed EU AI Act,[4] bans certain uses of AI (like the use of biometric identification in public spaces by governments) and imposes conditions (such as conformity assessment, trustworthy AI properties, and human oversight) for the use of AI in certain domains considered high-risk, such as hiring and education. However, solely relying on regulations to prevent undesired applications of AI presents several difficulties. In contrast to other fields of science, AI is moving very rapidly and there is a lack of specific regulation at the present time. More advanced models and systems are being released every month. The introduction of transformers (Vaswani et al., 2017), coupled with the increased availability of computing resources, has resulted in rapid advancement in the areas of natural language generation and multi-modal reasoning, in under five years. We argue that while government regulation processes are best equipped to handle the effect of a new technology in the long term, they leave a gap in the short term.

Second, often a responsible approach to releasing AI systems needs to consider the capabilities and limitations of a specific AI model. For instance, if a machine translation system has a high word-error rate, it may not be a problem if used for recreational purposes, but could be dangerous when employed in other contexts (e.g., translating documents with medical or legal implications). Such case-specific determinations can be difficult to encode in a legislation that aims to specify general rules to balance different broad categories of risks, and even harder to operationalize in the absence of meaningful and generally applicable thresholds. Further, different parties, developers or researchers may want to im-

[2]However, there are many copyright issues beyond fair use and copyrighted training data. See Lee et al. (2023).

[3]https://gdpr.eu/

[4]https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

pose specific restrictions on the use of the AI models they build in cases where they feel government regulation is too weak or does not exist.

Can licenses help? Recognizing these challenges, the use of contractual and copyright law have been recently suggested as a means of controlling end-use (Contractor et al., 2022b). This approach holds promise, after all much of the software in use today is accompanied by a license governing its terms-of-use. Such terms-of-use can govern commercial use, compliance with local laws and regulation, dispute resolution mechanisms, disclaimers of warranties and risks and intended-use. This is not only true for software sold by companies, source code that is open-sourced is accompanied by a license which grants the permissions for licensee to distribute, re-use, modify the accompanying software and source code. In some cases, terms referred to as "copy-left" clauses are included in some open-source licenses, which require that any downstream software or code reusing any component of the licensed artifact is subject to the same requirements as the original license. We argue that licensing will become increasingly important as AI continues to develop at a rapid pace.

Related Tools for AI Governance. A number of other mechanisms aimed at promoting the ethical and responsible use of AI have been proposed that are relevant to mention.

Promoting Informed-use: A number of tools have been proposed as a way to organize information about machine learning models in a consistent manner. AI factsheets (Arnold et al., 2019) and AI model cards (Mitchell et al., 2019a) both help to achieve these ends. However, they themselves do not have an explicit mechanism for enforcement about how the models are used.

Ethical Guidelines: Ethical guidelines[5] [6] and Responsible AI initiatives[7] [8] help to communicate best practices and standards that can help practioners act responsibly.

3. A Study of AI Licenses

To motivate our position on how responsible AI licenses should evolve, we conducted a study of the current state of the licensing in AI. Given the complex socio-technical nature of of the domain we pursue a mixed-methods approach combining insights from interviews of licenses users, quantitative analysis of license adoption and a review of existing licenses including clustering of their behavioral use clauses. The interviews involved practitioners and subject matter experts who have adopted licenses for high profile AI projects. We analyzed over 170,000 AI model repositories. To cluster the licenses we manually categorized behavioral use clauses in the most heavily utilized license types.

Q1. Describe your background and the project that you chose to license with behavioral use clauses.
Q2. Why did you choose to release the model with behavioral use clauses?
Q3. Did you feel it was necessary to modify an existing responsible AI license? Why/why not?
Q4. What considerations drove your choices of clauses to include in the license?
Q5. What other restrictions did you include in the license?
Q6. What other restrictions did you consider including in the license, but ultimately excluded?
Q7. What do you think the restrictions will achieve?
Q8. How can use restriction based licenses have impact in the AI sector?
Q9. Imagine: What tools, platforms or processes could accompany use restriction based licenses to increase their efficacy, adoption or robustness?

Figure 2. Questions posed to our interviewees regarding the adoption of responsible AI licenses.

3.1. Interviews with License Adopters

Participants. We performed semi-structured 30-minute interviews with four AI researchers and software developers. The interviewees has a mean of 18 years of post-graduate experience in the field of AI (8, 30, 10, 24 years). The interviews explored the questions listed in Fig. 2. The interviewees had all release projects in the past two years under a responsible AI license. The participants had the following roles: CEO AI startup (2), Research Scientist (1), AI Director (1). The participants were based in the US (3) and France (1). These included: large language and code models, and robotics and computer vision software.

Analysis. A common theme the interviewees discussed was the tension between the "willingness to release assets on an open basis" on one hand and the "concerns within the community and internal values about uses of the model." (P1). This was a motivating factor for the choice to select a responsible AI license rather than a traditional OS license. With the evolving nature of AI there is a "need for people to think about how to use [models] responsibly" (P2).

The interviewees spoke about the evolving nature of AI technology as one of the reasons they chose to adapt use clauses in the licenses. One interviewer summarized a common theme describing that "existing OS licenses didn't perfectly match our mission" (P3). One roboticist, commented on the fact that physical instantiations of AI required terms

[5] https://www.microsoft.com/en-us/ai/responsible-ai-resources
[6] https://ai.google/responsibilities/responsible-ai-practices/
[7] https://pwc.com/rai
[8] https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf

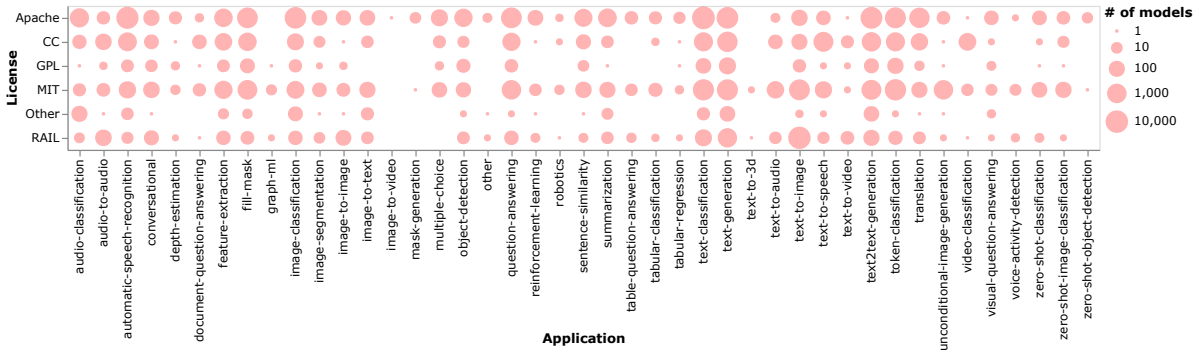

Figure 3. **Applications of Models by License Type.** The number of repos in each application domain that are under each license.

that restricted "destruction of physical people of property" that other licenses did not include. A second point was that existing OS licenses were "not suitable for [other assets]" (e.g., datasets) (P3).

In considering their options, a practice that multiple interviewees used was to review existing OS licenses to see if any satisfied their needs. However, licenses like Apache 2.0 were not "ML or LLM specific" (P1) and needed to be adapted. They noted that adaptation was not trivial. One participant, who was an ML researcher and entrepenur, said licensing was "not our expertise, and would have required resources that we didn't have" (P2). Another commented on the ambiguity introduced by changing licenses ad hoc - "[It is difficult] for someone to know the difference between license A and B, that is why we didn't modify an OS license" (P4). When asked about tooling, the participants identified that there were other situations in which agreements could be "generated via questionnaire ... by ticking boxes" and that having a template would be easier.

3.2. Analysis of Licenses and their Clauses

Quantitative Analysis. We performed an analysis of the HuggingFace Model Hub. The Model Hub API provides information about the models, licenses and domains/applications, when the repository was created and how many times the model has been downloaded. At the time of the analysis (January 2024) 174,163 model repositories had licenses associated with them.

Of these, the most common licenses used were Apache[9] (38.1%), RAIL[10] (24.0%), MIT[11] (17.5%) and Creative Commons[12] (7.5%) (see Fig. 1). RAIL licenses are the most common type of license *with* behavioral use clauses and were the second most adopted license type overall (see Fig. 1). The share of RAIL licenses has grown from 1% (N=209) in September 2022 to 10% (N=4,035) in January 2023 to 24.0% (N=41,700) in January 2024. In comparison to counting the number of repositories licensed under RAIL licenses, RAIL licenses measured by the number of downloads make up the fourth biggest category behind Apache, MIT and CC licenses (see Fig. 5 in the appendix). These results support the conclusion that RAIL adoption is considerable and growing; however, some of the most popular models are not using RAIL licenses.

RAIL licenses have been used with repositories across many different application domains, including text, image, speech processing and many methods, including supervised and unsupervised learning, reinforcement learning, graph learning, classification and generative methods. Given the adoption by StableDiffusion (Rombach et al., 2022), text-to-image translation was the most common model type to adopt a behavioral use license (see Fig. 3).

Qualitative Analysis. To better understand the types of behavioral uses that have been restricted in the first wave of responsible AI licenses we conducted a qualitative analysis of 77 different legal clauses from 7 different responsible AI licenses that are currently in use: AI Pubs RAIL[13] – which is also available for use by researchers publishing at AAAI conference venues,[14] BigScience OpenRAIL (Workshop et al., 2022), CodeML OpenRAIL (Li et al., 2023b), Llama-2 (Touvron et al., 2023), FALCON (Almazrouei et al., 2023), ImpACT Licenses (Allen Institute for AI, 2024), GRID (Vemprala et al., 2023) and the DBRX model [15] by Databricks.

We guided our analysis with the following questions: Q1: What artifacts are being restricted?, Q2: What are the uses being restricted?, and Q3: How are they being restricted?

[9] https://www.apache.org/licenses/LICENSE-2.0
[10] https://www.licenses.ai
[11] https://opensource.org/license/mit/
[12] https://creativecommons.org/licenses/list.en
[13] https://www.licenses.ai/blog/2023/3/3/ai-pubs-rail-licenses
[14] https://aaai.org/aaai-publications/
[15] https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm

	Behavioral Restrictions	AIPubs RAIL	BigSci. OpenRAIL	CodeML OpenRAIL	LLaMA 2	FALCON	ImpACT L/M/H	GRID	DBRX
Discrim.	(1) Discriminate against people based on legally protected characteristics	✓	✓	✓	✓			✓	✓
	(2) Administrate justice, law enforcement, immigration, or asylum processes, such as predicting that a natural person will commit a crime.		✓	✓			✓	✓	✓
	(3) Defame, disparage or otherwise harass others.	✓	✓	✓	✓	✓		✓	✓
	(4) Exploiting, harming or attempting to exploit or harm minors in any way	✓	✓	✓	✓	✓		✓	✓
Disinformation	(5) Create, present or disseminate verifiably false or misleading information for economic gain, harm people, or to intentionally deceive the public.	✓	✓	✓	✓			✓	✓
	(6) Synthesize or modify a natural person's appearance, voice, or other individual characteristics, unless prior informed consent of said person is obtained.	✓	✓	✓	✓			✓	
	(7) Generate/disseminate information and place the information in any public context without expressly disclaiming that the content is machine generated.	✓	✓	✓	✓		✓	✓	✓
Legal	(8) Fully automate decision-making that creates, modifies or terminates a binding, enforceable obligation between entities.	✓	✓	✓			✓	✓	✓
	(9) Violate any national, federal, state, local or international law or regulations.	✓	✓	✓	✓	✓		✓	✓
Privacy	(10) Utilize personal information to infer additional personal information, including legally protected characteristics, vulnerabilities or categories; unless informed consent from the data subject is received.				✓				
	(11) Generate/disseminate personal identifiable information that can be used to harm an individual or to invade the personal privacy of an individual.	✓	✓	✓	✓			✓	✓
	(12) Engage in, promote, incite, or facilitate the harassment, abuse, threatening, or bullying of individuals or a group.	✓	✓	✓	✓	✓		✓	
Health	(13) Provide medical advice or clinical decisions without accreditation of the system; unless the use is (i) in an internal research context with independent and accountable oversight and/or (ii) with medical professional oversight.		✓	✓		✓			✓
	(14) In connection with any activities that present a risk of death or bodily harm to individuals, including self-harm or harm to others, or in connection with regulated or controlled substances.				✓				
	(15) In connection with activities that present a risk of death or bodily harm to individuals, including inciting or promoting violence, abuse, or any infliction of bodily harm.				✓				
	(16) Exploitaion of vulnerabilities of a group/person in a manner that is likely to cause physical or psychological harm	✓	✓	✓					✓
Military	(17) Use for active deployment of weaponry.							✓	
	(18) Use for purposes of building or optimizing military weapons or in the service of nuclear proliferation or nuclear weapons technology.				✓		✓		
	(19) Use for purposes of military surveillance, including any research related to military surveillance.				✓		✓		
	(20) 'Real time' remote biometric processing or identification systems in publicly accessible spaces for the purpose of law enforcement.						✓	✓	
Other	(21) Generate/disseminate malware/ransomware or other content for the purpose of harming electronic systems.			✓					✓
	(22) Fail to appropriately disclose to end users any known dangers of your AI system				✓				

Table 1. **Summary of Behavioral-Use Clauses.** Clauses included in popular responsible AI licenses.

For Q1, we coded for the specific language referencing the artifact whose use is being restricted in the clause. For Q2, two of the authors independently coded broadly for any concepts that could help us understand the values and social norms of the developer community. Each author independently conducted an axial coding phase (Braun & Clarke, 2006) to refine the granularity of our analysis. Next, we mapped the resulting codes to existing Responsible AI frameworks like the NIST AI RMF (Tabassi, 2023) and the CSET Harm Taxonomy (Hoffman & Frase, 2023). Each of these frameworks are designed to facilitate a more standardized and informed, yet flexible and adaptable approach for measuring, monitoring, and mitigating AI harm. For Q3, we learned early on that nearly all restrictions in the clauses we analyzed restrict a particular use of an artifact by simply prohibiting it. While the results of the question as initially posed would not be interesting with our data, we did observe that some use restrictions had caveats.

The result was 22 clauses (see Table 1). We find that almost all licenses include clauses that govern discrimination, disinformation and violations of the law. Most contain privacy related restriction. The other two main groups of clauses are heathcare and military applications of which some licenses include restrictions.

Additional Restrictions on Distribution. Apart from behavioral-use clauses applicable to the use and distribution of the artifacts, licenses sometimes also include additional requirements (See table 2). For instance, the AI Pubs RAIL License permits the use of the licensed code or model only for "Research-use" while the LLaMA-2 license explicitly prohibits the use of the model for anything except creating a derivative of LLaMA-2 with the same terms. It also stipulates additional licensing requirements if the monthly active users of the model exceed 700 million. Similar provisions are included in the license associated with the DBRX model. The ImpACT licenses permit the creation of derivatives

	AI Pubs RAIL	BigScience OpenRAIL	CodeML OpenRAIL	Llama 2	FALCON	ImpACT L/M/H Licenses	GRID	DBRX
Behaviorial-Use Restrictions	✓	✓	✓	✓	✓	✓	✓	✓
Conditions on Derivatives				✓		For H only		✓
Conditions on Distribution	✓					For M & H only		

Table 2. **Requirements for Derivatives and Distributions.** All licenses require inclusion of the original behavioiral-use restrictions; this table reports the presence of additional conditions imposed on derivatives and distributions.

and distribution for low-risk (L) use-cases but do not allow derivatives for high-risk (H) use-cases and they also do not permit derivatives for both, medium (M) and high-risk (H) use-cases.

The flavor of RAIL licenses, known as the OpenRAIL family of licenses, avoid non-behavioral use restrictions (such as "research only" requirements). Currently, the "OpenRAIL" moniker is given to license terms that (1) incorporate behavioral-use clauses, and (2) allow for the otherwise unlimited use, modification, and distribution of applicable artifacts, as long as the behavioral-use restrictions are included. The "open" designation has been important to certain actors within the AI eco-system who wish to emphasize that the IP they are sharing can be treated much like other open source code (freely share-able and modifiable), albeit with behavioral-use restrictions. Both the BigScience and CodeML RAIL Licenses are OpenRAIL Licenses.

4. Discussion

Responsible AI Licenses Are Being Adopted. Our quantitative analysis of the repositories shows that RAIL licenses are being adopted at a relatively large scale. This has been driven by a number of high profile and widely adopted projects and the associated derivative projects. The adoption is large enough to suggest that it is necessary to seriously study these types of licenses and their implications for the AI and open source communities, the latter of whom have been grappling with the definition of Open Source AI.[16] The rapid rise of licenses with behavioral use clauses, particularly amongst projects involving large parameter models, signals a perception that existing open source licenses do not contain sufficient restrictions for responsibly distributing capable machine learning models. All of our interviewees highlighted that conundrum.

Responsible AI Licenses Are Being Adapted. Our analysis of behavioral use clauses included in responsible AI licenses shows that there is a large amount of overlap in restrictions. However, they differ in important ways. Some licenses (e.g., FALCON) are more permissive and include a smaller subset of clauses while others (e.g., OpenRAIL, ImpACT) include a much larger list. Some licenses (e.g. CodeML OpenRAIL) include domain specific restrictions.

Standardization of Responsible AI Licenses is Necessary. Based on our analysis we take the position that the proiferation of responsible AI licensing is good, as it provides options to developers who are worried about misuse. However, there is a need for standardization to avoid confusing users or diluting impact while enabling some customization. We believe that there are tools that could support customizing licenses that still satisfy a set of core criteria. This is pertinent given that AI developers who aim to mitigate harmful uses of the artifacts they share will need to tailor licenses for the specific domain, context of use, and type of AI system. Below, we gather implications for license tooling development, considering how possible tools may support people (most likely without exhaustive legal expertise) in their practice of licensing the AI artifacts they are creating.

A Community-Oriented License Generator. An actionable route to standardized and customizable licenses is a generator purpose-built for AI licenses. Existing RAIL licenses already follow a topology ameanable to structured generation (e.g., Open-/ vs. RAIL, sub-specifications for Application/Model/Source Code artifacts).[17] Such generators are an established means to allow people without commensurate legal experise to choose among potential licenses for their artifacts. For instance, the Creative Commons (CC) initative has published a tool[18] which takes users through the process of deciding how people may or may not use their creative work. In the case of AI licenses, standardized customization along use restrictions may allow for more ML-specialized tailoring (as called for by the interviewees); which in turn places specific demands on a license generator.

Responsible AI licenses, unlike CC-type creative works, are *not* broadly artifact-agnostic: in practice it matters whether one is licensing a ready made application, a set of source code, or a model, andthere may be specific restrictions and conditions that need to apply to each. For instance, one could choose to release a model with behaviorial-use restrictions but release the accompying source code under an open source license. In terms of a license generator, a design priority would be the ability to specify the artifact(s) to be

[16] https://deepdive.opensource.org/wp-content/uploads/2023/02/Deep-Dive-AI-final-report.pdf

[17] See e.g. https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses, accessed 12/01/2024.

[18] https://chooser-beta.creativecommons.org/, accessed 15/01/2023.

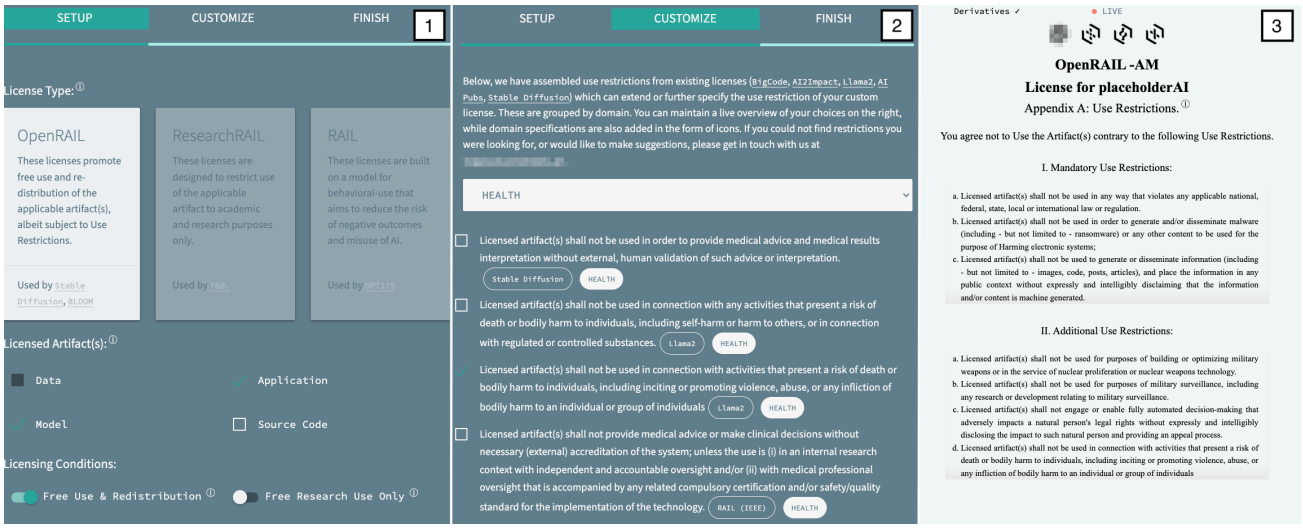

Figure 4. **Interface Designs for a Responsible AI License Generator.** *1.* Setup panel for choosing a license for "placeholderAI." On the left hand side, the type of artifact(s) specification and derivation handling can be specified by the user. *2.* Customization panel for adding further use restrictions. On the left hand side, a dropdown menu allows selection of use restrictions by domain. *3.* On the right hand side, ticked selections are added to the preview in a separate block at the bottom.

licensed (e.g., *D*ata, *A*pplication, *M*odel, *S*ource Code) with behavioral-use clauses.

Specifying use restrictions gives AI developers further means to appropriately customize licenses. However, the large number of possible use restrictions risks leading to many different licenses that could be generated could confuse 1) the licensor who may not be able to easily judge how restrictive a license should be, 2) the licensee who may find it difficult to know what the assets are, and are not, permitted to be used for. Thus, aggregating potential use restrictions by domain could help (e.g., privacy, disinformation, health).

We developed an interface design for such a license generator (see Fig. 4), that incorporates the above considerations while also providing a "live" preview of the RAIL-type license that is currently being generated. The protoype shown includes a base set of mandatory use-restrictions, that were selected based on our analysis of existing licenses. Icons related to the domains of selected restrictions are added to the top of the license being generated.We have implemented the license generator prototype [19] and the code for the front and back ends is available. [20] [21]

Code Scans for AI Licenses. Similar to a license generator, it is feasible to construct tools which act as aids for AI developers when sharing and maintaining their created artifacts on platforms such as Huggingface and GitHub. An open question as far as licenses are concerned is the compability of dependencies (e.g., other repositories, third-party code, etc.) with the requirements of a responsible AI license. When dependencies exist that are licensed under a GPL license, for instance, a developer may not be able to select an OpenRAIL license for their artifacts due to requirements imposed by the GPL license. Alternatively, even when there are dependencies licensed under different responsible AI licenses, the arising artifact, will most likely need to be compliant with the stricter set of usage restrictions.

In practice, AI systems—just like any contemporary software system—may integrate many dependencies, complicating this process further. To this end, a tool that scans existing dependencies for their licenses may aid AI developers in figuring out whether and which clauses are actually applicable. Such a tool could crawl the repositories of a given platform, and be integrated, for instance, as a service in GitHub Marketplace or in a Huggingface Space; or be offered as a command line tool for local use. In either form, such support for the workflow of AI developers in the ways they manage and share their artifact(s) may encourage a more reflective licensing practice that reflects the actual constraints of given dependencies.

Additional Considerations in Licensing. There is a risk that licenses could become overly restrictive, lead to concentration of power (Widder et al., 2023), lack representation from diverse positions, or simply be ineffective (Cui & Araujo, 2024). Our analysis suggests a level of consensus amongst the AI development community about the use of licenses and certain licenses clauses; however, it is important that there is broad participation in the design process. There

[19] https://www.licenses.ai/rail-license-generator

[20] https://github.com/RAIL-Initiative/RAIL-license-generator_frontend

[21] https://github.com/RAIL-Initiative/RAIL-license-generator_backend

are example of participatory AI design that can be engaged (e.g., standardization of protocols and the growth of the Internet; participatory AI) (Delgado et al., 2023; Sloane et al., 2022). Participatory development processes have repeatedly played a role in standards-setting processes (Russell, 2014; Cooper & Vidan, 2022). As adoption of licenses grows, opportunities for greater involvement of people with different backgrounds and experiences should be prioritized.

5. Conclusion

The growth in large parameter AI models has increased the need for tools to help responsibly release assests. Adoption of responsible AI licenses has increased significantly with several high-profile releases (e.g., BLOOM, LLaMA2, StarCoder, Stable Diffusion). Open Source licensing has been influential thanks to the clear definition and standardization of licenses such as MIT and Apache 2.0. The proliferation of different licenses risks becoming confusing and we take the position that standarized forms of customization, along with appropriate tooling could make them more user-friendly.

Impact Statement

As is increasingly well understood by the community, machine learning and AI projects can have significant, broad, and uncertain impacts. The development of new forms of licensing provides researchers with the ability and resources to release assets while restricting uses that they anticipate will lead to negative consequences. Standardization of licenses does involve making some value judgements about what types of behaviors are higher risk than others; however, providing the ability for customization can add flexibility. For licenses directed at responsible use such as RAILs, this can be achieved practically by allowing for revision and extension of mandatory and additional use restrictions in individual license instantiations.

References

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

Almazrouei, E., Alobeidli, H., Alshamsi, A., Cappelli, A., Cojocaru, R., Debbah, M., Goffinet, E., Heslow, D., Launay, J., Malartic, Q., 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.

Arnold, M., Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K. N., Olteanu, A., Piorkowski, D., Reimer, D., Richards, J., Tsay, J., and Varshney, K. R. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development*, 63(4/5):6:1–6:13, 2019.

Braun, V. and Clarke, V. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77–101, 2006. doi: 10.1191/1478088706qp063oa. URL https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

Bretthauer, D. Open Source Software: A History. *Published Works*, 7, 2001.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*, 2020.

Brundage, M., Mayer, K., Eloundou, T., Agarwal, S., Adler, S., Krueger, G., Leike, J., and Mishkin, P. Lessons learned on language model safety and misuse, 2022. URL https://openai.com/research/language-model-safety-and-misuse.

Contractor, D., Ferrandis, C. M., Lee, J., and McDuff, D. 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.

Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., Vincent, N., and Li, H. Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 778–788, 2022b.

Cooper, A. F. and Vidan, G. Making the unaccountable internet: The changing meaning of accounting in the early arpanet. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 726–742, 2022.

Cooper, A. F., Moss, E., Laufer, B., and Nissenbaum, H. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pp. 864–876, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522. doi: 10.1145/3531146.3533150.

Cooper, A. F., Lee, K., Grimmelmann, J., Ippolito, D., Callison-Burch, C., Choquette-Choo, C. A., Mireshghallah, N., Brundage, M., Mimno, D., Choksi, M. Z., Balkin,

J. M., Carlini, N., Sa, C. D., Frankle, J., Ganguli, D., Gipson, B., Guadamuz, A., Harris, S. L., Jacobs, A. Z., Joh, E., Kamath, G., Lemley, M., Matthews, C., McLeavey, C., McSherry, C., Nasr, M., Ohm, P., Roberts, A., Rubin, T., Samuelson, P., Schubert, L., Vaccaro, K., Villa, L., Wu, F., and Zeide, E. Report of the 1st Workshop on Generative AI and Law. *arXiv preprint arXiv:2311.06477*, 2023.

Cui, J. and Araujo, D. A. Rethinking use-restricted open-source licenses for regulating abuse of generative models. *Big Data & Society*, 11(1):20539517241229699, 2024.

Delgado, F., Yang, S., Madaio, M., and Yang, Q. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–23, 2023.

Fergusson, G., Fitzgerald, C., Frascella, C., Iorio, M., McBrien, T., Schroeder, C., Winters, B., and Zhou, E. Generating Harms: Generative AI's Impact & Paths Forward. Technical report, Electronic Privacy Information Center, 2023.

Gokaslan, A., Cohen, V., Pavlick, E., and Tellex, S. Openwebtext corpus, 2019.

Gokaslan, A., Cooper, A. F., Collins, J., Seguin, L., Jacobson, A., Patel, M., Frankle, J., Stephenson, C., and Kuleshov, V. CommonCanvas: An Open Diffusion Model Trained with Creative-Commons Images. *arXiv preprint arXiv:2310.16825*, 2023.

Hoffman, M. and Frase, H. Adding structure to ai harm: An introduction to cset's ai harm framework, July 2023. URL https://cset.georgetown.edu/publication/adding-structure-to-ai-harm/.

Janapa Reddi, V., Elium, A., Hymel, S., Tischler, D., Situnayake, D., Ward, C., Moreau, L., Plunkett, J., Kelcey, M., Baaijens, M., et al. Edge impulse: An mlops platform for tiny machine learning. *Proceedings of Machine Learning and Systems*, 5, 2023.

Keller, P. and Bonato, N. 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.

Klosek, K. and Blumenthal, M. S. Training Generative AI Models on Copyrighted Works Is Fair Use, 2024. URL https://www.arl.org/blog/training-generative-ai-models-on-copyrighted-works-is-fair-use/.

Lee, K., Cooper, A. F., and Grimmelmann, J. Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain. *arXiv preprint arXiv:2309.08133*, 2023.

Lemley, M. and Casey, B. Fair Learning. *Texas Law Review*, 99:743, 2021.

Li, H., Vincent, N., Jernite, Y., Merrill, N., Benjamin, J. J., and Tarkowski, A. 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, pp. 553–555, New York, NY, USA, 2023a. Association for Computing Machinery. ISBN 9798400701290. doi: 10.1145/3584931.3611276. URL https://doi.org/10.1145/3584931.3611276.

Li, R., Allal, L. B., Zi, Y., Muennighoff, N., Kocetkov, D., Mou, C., Marone, M., Akiki, C., Li, J., Chim, J., et al. Starcoder: may the source be with you! *arXiv preprint arXiv:2305.06161*, 2023b.

Liu, X., Narayanswamy, G., Paruchuri, A., Zhang, X., Tang, J., Zhang, Y., Sengupta, S., Patel, S., Wang, Y., and McDuff, D. 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/.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pp. 220–229, New York, NY, USA, 2019a. Association for Computing Machinery. ISBN 9781450361255. doi: 10.1145/3287560.3287596. URL https://doi.org/10.1145/3287560.3287596.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pp. 220–229, 2019b.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., and Lee, K. 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.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21: 1, 2020.

Resnick, D. B. Openness versus Secrecy in Scientific Research Abstract. *Episteme*, pp. 135 – 147, February 2006.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Russell, A. L. *Open standards and the digital age*. Cambridge University Press, 2014.

Sag, M. Copyright Safety for Generative AI. *Houston Law Review*, 2023. Forthcoming.

Samuelson, P. Generative AI meets copyright. *Science*, 381(6654):158–161, 2023. doi: 10.1126/science.adi0656. URL https://www.science.org/doi/abs/10.1126/science.adi0656.

Sloane, M., Moss, E., Awomolo, O., and Forlano, L. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–6, 2022.

Soldaini, L., Kinney, R., Bhagia, A., Schwenk, D., Atkinson, D., Authur, R., Chandu, K., Dumas, J., Lucy, L., Lyu, X., et al. Dolma: An Open Corpus of 3 Trillion Tokens for Language Model Pretraining Research. *Allen Institute for AI, Tech. Rep*, 2023.

Tabassi, E. Artificial intelligence risk management framework (ai rmf 1.0), 2023-01-26 05:01:00 2023. URL https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=936225.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

Vemprala, S., Chen, S., Shukla, A., Narayanan, D., and Kapoor, A. Grid: A platform for general robot intelligence development. *arXiv preprint arXiv:2310.00887*, 2023.

von Krogh, G. and Spaeth, S. 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.

Widder, D. G., West, S., and Whittaker, M. Open (for business): Big tech, concentrated power, and the political economy of open ai. *Concentrated Power, and the Political Economy of Open AI (August 17, 2023)*, 2023.

Workshop, B., Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., et al. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.

A. Tree Map of Models by License

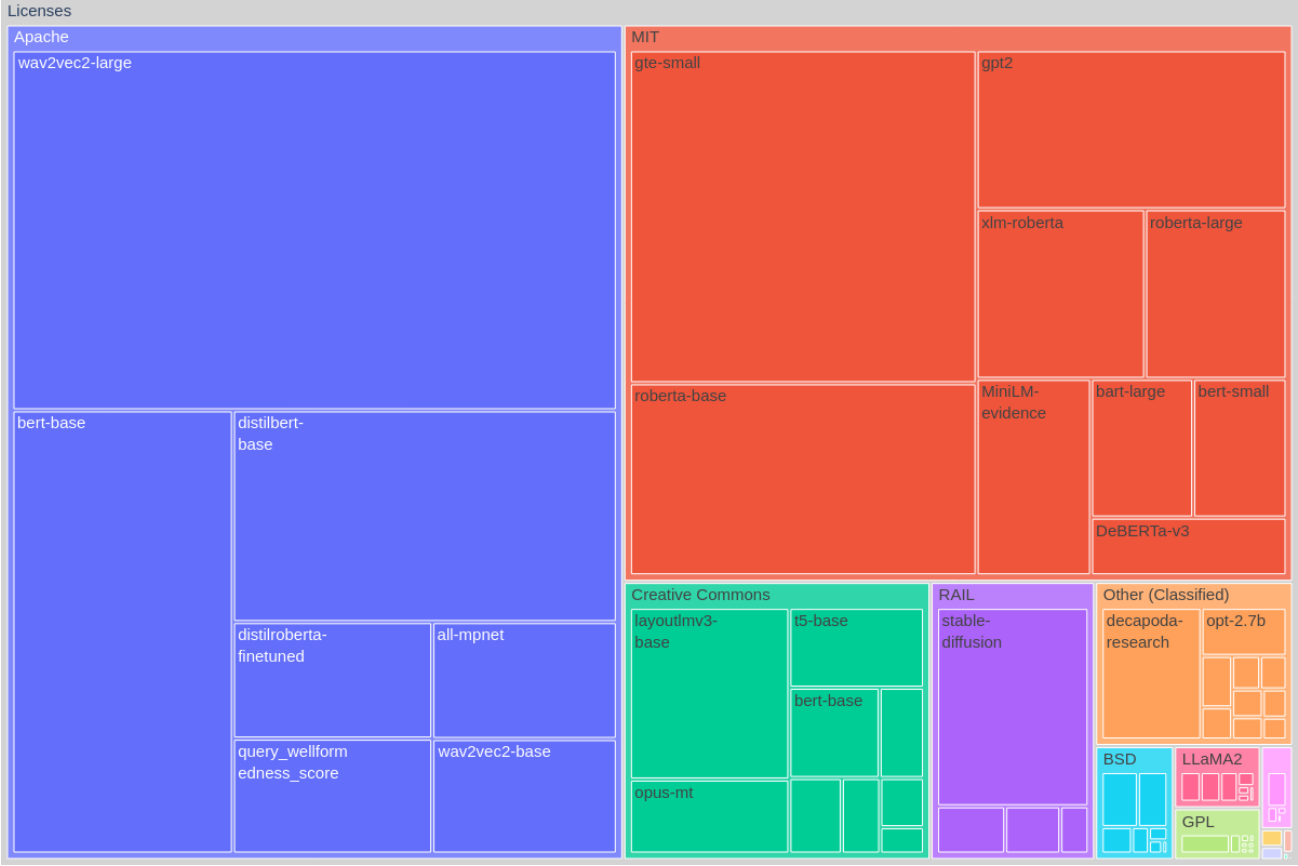

Figure 5. **Model Treemap by Downloads.** Treemap showing the top 10 models for each license weighed by the number of downloads