

Between Commons and Property: Individual Image-Generation Model Creators’ Perceived Accountability for Derivative AI Model on Open Model Marketplaces

Eun Jeong Kang¹, Fengyang Lin¹, Angel Hsing-Chi Hwang²

¹Cornell University,

²University of Southern California

eunkang@infosci.cornell.edu, fl354@cornell.edu, angel.hwang@usc.edu

Abstract

Within the open model ecosystem, AI model marketplaces enable non-AI experts to create and distribute derivative AI models. These platforms establish RAI approaches that creators are expected to follow to prevent negative downstream effects from model deployment, yet such governance approaches often rely on creators’ intrinsic motivation and integrity. Little is known about how individual creators conceptualize their accountability for derivative models and how these conceptualizations shape their RAI practices throughout the development cycle.

Through semi-structured interviews with 19 creators of derivative image-generation models on open model marketplaces, we identify two distinct accountability logics underlying creators’ development practices: commons-based peer production and personal creative property. Guided by these logics, creators engage with platform RAI approaches in ways that diverge from their intended purposes: they strategically repurpose attribution tools for model visibility rather than provenance transparency, displace platform authority in favor of community-based governance, restrict model access to manage downstream accountability costs, and treat RAI-related work as undervalued labor to be minimized. We discuss how self-regulatory use of RAI tools, shaped by individuals’ perceptions, produces shortcomings, and how open model marketplaces might mitigate these limitations by structurally supporting creators’ engagement with RAI practices.

1 Introduction

Open AI model marketplaces¹, such as Hugging Face and Civit AI, act as critical intermediate ML/AI model service providers in the open model ecosystem, contributing to innovating AI development pipelines (Widder, West, and Whittaker 2023; Raffel 2023; Gorwa and Veale 2024). These platforms enable users to easily access user-uploaded models as well as create derivatives from those models (Choksi, Mandel, and Benthall 2025). Individuals – including hobbyists and independent developers – can build customized derivative models without training from scratch or a lot of

resources, through fine-tuning open-source AI models, custom dataset training, or other modification methods (Anisuzaman et al. 2025; Canet Sola and Guljajeva 2026).

As the number of models hosted on the platforms grows, concerns regarding downstream effects of user-created custom models have intensified (Schneider and Hagendorff 2025; Wei et al. 2024). Pre-trained models can be fine-tuned on copyrighted content or pornographic imagery to serve individuals’ personal interests. In 2025, over 5,000 models generating non-consensual sexual imagery of real individuals were reportedly migrated to Hugging Face from another major marketplace (Maiberg 2025a). Once these models are downloaded by downstream users, creators can no longer trace how users interact with their models or how those models spread (Widder et al. 2022). These concerns are particularly pronounced for image generation models, where the aesthetic visual content amplifies both public attention and potential harms (Marchal et al. 2024; Lima, Grgić-Hlača, and Redmiles 2025). Because such harms often remain latent until downstream deployment, platforms increasingly need governance approaches that extend beyond moderation of within-platform trust and safety issues.

Platforms have begun to promote responsible open model development guidelines and Responsible AI (RAI) tools for mitigating harms that creators may embed in their models (Gorwa and Veale 2024; Berman, Goyal, and Madaio 2024; Gebru et al. 2021). Beyond centralized moderation, these approaches are intended to support creators in voluntarily upholding the safety and transparency of their models (Mitchell et al. 2019; He, Houde, and Weisz 2025).

However, these existing self-regulatory approaches yield limited effect, particularly when creators lack intrinsic motivation or relevant expertise. Creators apply ad-hoc, self-constructed ethical norms drawn from personal experience when developing models without a concrete understanding of the actors involved in the AI pipeline (Widder et al. 2022; Widder and Nafus 2023). Without a standardized legal framework enforcing RAI practices (Hacker and Holweg 2026), creators tend to enact RAI approaches according to their own sense of accountability for the models they produce. For instance, model card completion relies on creators’ integrity and accountability (Mitchell et al. 2019); documents may be left blank or filled minimally when creators anticipate that detailed documentation could enable

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹In this paper, we use “open AI” as an umbrella term for open-source and open-weight models, following Widder et al. (Widder, West, and Whittaker 2023).

malicious actors to exploit their models. However, we have limited understanding of how individual creators conceptualize their accountability for their models – open derivative models customized through model modification methods – and how this conceptualization, together with their broader values and community relationships, shapes their RAI practices throughout the development cycle.

The goal of this work is therefore twofold: (1) to examine how existing platform governance approaches become misaligned with creators' experiences, and (2) to envision new approaches to platform governance grounded in individual creators' practices and their experiences on these platforms, in ways that better support responsible AI. Through semi-structured interviews with 19 individual model creators who produce and distribute image-generation derivative models on open model marketplaces, we examine their current model creation and distribution practices and how platform governance approaches influence such processes. To generate concrete feedback and envision alternative governance approaches, we gather input from model creators on how they respond to existing Responsible AI (RAI) approaches (such as responsible model development guidelines and attribution tools) used in other AI-infused creative contexts. Together, our interviews address two key research questions:

1. **RQ1.** How do individual model creators' perspectives on image-based derivative models shape their sense of accountability in model development and sharing?
2. **RQ2.** How do individual model creators engage with RAI approaches on open model marketplaces, and in what ways do their practices diverge from the intended purposes of these approaches?

Our findings reveal two distinct logics underlying how model creators conceptualize accountability for their derivative models, each producing a different orientation toward responsible model development. The first logic draws on commons-based peer production, where creators view their work as a byproduct of community engagement and shared resources (Benkler and Nissenbaum 2006). The second draws on personal creative property, where creators treat their derivatives as personal proprietary artifacts (Birch and Muniesa 2020). These logics interact, jointly guiding creators' derivative model production cycle.

Within these logics, RAI approaches are practiced based on creators' understanding of how responsible practices fit their circumstances on platforms, within their communities, and in relation to themselves. When engaging with RAI approaches provided by platforms, creators' practices diverge from the intended purposes of these tools in several ways: they strategically repurpose attribution tools for model visibility rather than provenance transparency, rely on community-based governance that displaces platform authority, restrict model access to manage downstream accountability costs, and treat RAI-related work as undervalued labor to be minimized. These practices, while locally rational for creators, can distort the transparency of model dependencies and complicate attribution across the open model ecosystem.

Together, our findings show that platform governance approaches that rely on individuals' decisions self-regulating their workflow are insufficient, given individual creators' divergent motivations, situated perspectives, and constantly changing perceived accountability. We argue that addressing these risks requires recognizing individual creators as a structural locus of risk, and suggest mechanisms by which platforms can support individuals in performing meaningful RAI practices.

2 Background and Related Work

The Role of Open Model Marketplaces in the Open Model Ecosystem

Aligned with the open-source culture – in which software is freely distributed, technologically neutral, and granted free sub-licensing rights (Aksulu and Wade 2010, p. 577) – that has historically fostered innovation and democratized software development (Meneely and Williams 2009; Google Public Policy Blog 2009), openness in AI has been strongly endorsed by scholars and industry actors as a means to stimulate innovation and experimentation while strengthening accountability (Tan et al. 2025; Paris, Moon, and Guo 2025; Langenkamp and Yue 2022; Kilamo et al. 2012). This openness also facilitates community engagement through peer-production – such as collaboratively auditing code and involving stakeholders in value-centered development, which has been viewed as a pathway toward more affordable, transparent, and accountable AI systems (Eiras et al. 2024; Wider, Whittaker, and West 2024).

In line with these goals, open model ecosystems have emerged with multiple actors operating across the AI supply chain. Major AI corporations and organizations (e.g., Meta, Stability AI ²) publicly released pre-trained models under open licenses for non-commercial purposes (Meta 2023; Eiras et al. 2024; Liesenfeld and Dingemanse 2024). Building on these open models, derivatives – models customized through post-training modification methods such as fine-tuning – can be created and hosted on dedicated platforms, which we refer to as “*open model marketplaces*” (e.g., Hugging Face, CivitAI) (Choksi, Mandel, and Benthall 2025). Within this ecosystem, our research focuses on open model marketplaces and the users who create and distribute derivative models on these platforms.

Governing Open Model Marketplaces for Responsible AI

Open model marketplaces are known for affording users freedom to experiment with hosted models, but researchers have raised concerns that this openness can also facilitate the distribution of models that produce harmful outputs or operate in unresolved legal areas (Harris 2023; Wei et al. 2024; Maiberg 2025b). First, models fine-tuned on non-consensual intimate images (NCII) that violate publicity rights can be widely distributed via the platforms (Wei et al. 2024; WIRED 2025). Researchers found that Text-to-Image

²<https://stability.ai/>

models hosted on these marketplaces can present representational harms and biases when downstream users generate images, calling for strategies to mitigate downstream harms (Schneider and Hagendorff 2025). Additionally, these models carry varied licenses that are not always complied with across derivative chains (Duan et al. 2025, 2024).

Under legal pressure to ensure internet trust and safety, platforms have attempted to regulate models that generate child sexual abuse material (CSAM)-related content or that may amplify harms when used by downstream users (Kilcher 2022; Gorwa and Veale 2024). For instance, CivitAI removed all user-uploaded models designed to generate CSAM or NCII of real individuals (Maiberg 2025c,a). However, platforms encounter significant governance challenges in moderating these models, in part because accountability is difficult to allocate across chains of derivative models (Laufer, Oderinwale, and Kleinberg 2025). The dual-use nature of these models further complicates moderation, as platforms must distinguish legitimate creative uses from unintended harmful applications (Gorwa and Veale 2024; François et al. 2025). Centralized moderation at the platform level is thus likely to incur substantial operational costs while still overlooking latent risks that only manifest in downstream contexts.

As a complementary approach, platforms promote user-driven, Responsible AI (RAI) practices through which users self-regulate their model development. These approaches typically rely on individuals' own sense of moral responsibility, assuming that actors will act in good faith. One mechanism is the *responsible model development guideline*, which describes what data should be filtered during fine-tuning and which kinds of outputs creators should be cautious about when deploying their models. Another approach is *attribution disclosure*, through which creators use templates or tools to document their evaluation results, data sources, and licenses (Mitchell et al. 2019; Gebru et al. 2021; Pepe et al. 2024). These provisions are intended to enhance the transparency, reproducibility, and accountability of derivative models, helping downstream users understand how to apply them.

However, prior work has demonstrated that “Responsible AI (RAI)” is not universally defined or practiced. Depending on the purpose and activities of AI development, people hold different priorities and norms about what to prioritize as their responsibility (Jakesch et al. 2022; Widder et al. 2022; Widder and Nafus 2023; Sadek et al. 2025). Moreover, prior research highlights that RAI approaches (e.g., Model Cards, Data Cards) may not be fully implemented by AI practitioners when other concerns, such as reputational and political risks, take precedence (Kawakami, Wilkinson, and Chouldechova 2024). Indeed, model documentation practices on one of the major model marketplaces (Hugging Face) remain insufficient, and cooperative commitments to transparency and reproducibility across open models are weakly upheld (Choksi, Mandel, and Benthall 2025; Pepe et al. 2024).

Bovens argues that accountability can be assigned through specific mechanisms or social relations that involve an obligation to explain conduct (Bovens 2010). That is,

how actors are rewarded for their creations, how they interact with other actors (e.g., community members), and to whom they are held accountable may affect how they use RAI approaches. The users of open model marketplaces include not only AI practitioners affiliated with companies or institutions, but also individuals who create AI models recreationally and independent freelancers. Widder et al. show that these creators enact ethical norms shaped by personal beliefs and cultural and labor-related experiences, often relying on informal community guidelines in open-source AI projects (Widder et al. 2022). Personal motivations such as reputational concerns further drive how they scrutinize their work (Bartsch et al. 2024). Tang et al. further show that developers experience epistemic, methodological, and ecosystem-level uncertainties when documenting their open-source models, pointing to the importance of infrastructural support across the AI supply chain (Tang et al. 2025). Yet little is known about how platform governance approaches—which require creators to self-regulate their workflows through RAI practices—function for individuals accessing, using, or distributing derivative models on these platforms.

We therefore examine how these creators conceptualize their accountability from their model development practices, and then how they engage with RAI approaches in their workflows.

3 Methodology

We conducted semi-structured interviews with 19 individual creators who create and distribute derivative image-generation models to understand how they conceptualize their accountability over their workflows and how they engage with RAI approaches in their practices. The study was approved by our University Institutional Review Board (IRB), and participants received a \$35 gift card for their 75–90 minute sessions.

Participants

We define individual model creators as users of online platforms who create and distribute open models independently. We focus on creators of image-generation models, since this type of model has attracted significant public attention as a form of entertainment, alongside concerns about its use to produce sexually abusive content (Marchal et al. 2024; Lima, Grgić-Hlača, and Redmiles 2025). Criteria also included being over 18 years of age and residing in the United States.

Participants were recruited via social media channels (LinkedIn, Reddit, Discord). Due to the limited population, we also reached out on model marketplaces (Hugging Face, CivitAI, and Tensor.Art) using their public nomination lists and snowball sampling (Parker, Scott, and Geddes 2019; Choksi, Mandel, and Benthall 2025).

Demographic Information Nineteen individual model creators participated in the study (see Table 1). Although we did not restrict recruitment to creators of derivative models specifically, all participants developed *derivative models*, open models customized from pre-trained models through

PID	Gender	Professional Context	Concept of Image-generation models	Platforms Where Participants Deploy Derivative Models
P1	Male	Hair designer	Afro hair style texture	Github
P2	Female	Creative technologist	Graphic visual	Hugging Face, CivitAI
P3	Male	UI/UX designer	Japanese anime character	Discord (Private)
P4	Male	UI/UX designer	Sports player meme	CivitAI
P5	Male	Graphic designer	Afro-artistic image	CivitAI
P6	Male	Independent model creator	Realistic portrait	CivitAI
P7	Male	Graduate (CS)	Japanese anime character	CivitAI
P8	Male	Graduate (CS)	Anime character fashion	Discord (Private)
P9	Male	Engineer	Superhero anime character	Discord (Private)
P10	Male	Undergraduate (CS)	Asian culture	Hugging Face, CivitAI
P11	Female	Graduate (CS)	Graphic visual	Discord (Private)
P12	Male	Independent model creator	Photorealistic character	Hugging Face
P13	Male	UI/UX designer	Photorealistic portrait	Hugging Face, CivitAI
P14	Male	Engineer	80s fashion style	CivitAI
P15	Male	Graphic designer	Animal character	CivitAI, Github
P16	Female	Physiotherapist	Education materials	Discord (Private)
P17	Male	UI/UX designer	Japanese anime character	Discord (Private)
P18	Male	Computer scientist	Japanese anime character, photorealistic portraits	Hugging Face, CivitAI
P19	Prefer not to say	Engineer	Japanese anime character	Hugging Face , CivitAI

Table 1: Demographic Information of Participants. Creative concept of model is reported based on the concept of image output generated from a model.

modification methods such as fine-tuning. Most participants distributed their models on Hugging Face and CivitAI, while six participants used private Discord channels for sharing models.

Seventeen participants were hobbyists, while two (P6, P12) identified themselves as independent freelance engineers, who operated independently and received commissions for building custom models. P2 has developed only image-to-image models, while P18 has created both image-to-image and text-to-image models. All others have only produced text-to-image models. Participants developed derivative models by re-using pre-trained model weights using either node-based interfaces like Comfy-UI or code-based environments like Google Colab. All participants’ models are derivatives of Stable Diffusion or Flux open weight models.

We also acknowledge the gender imbalance among participants. We suspect that the predominance of anime character models on model marketplaces (Song et al. 2024; Wei et al. 2024) contributes to gender disparities within these communities (Vasilescu, Capiluppi, and Serebrenik 2014).

Study Design

Our interview protocol was periodically revised as the study evolved, reflecting the iterative nature of qualitative research and the need to explore the meaning of participants’ contexts in depth (Creswell and Creswell 2017; Creswell and Poth 2016; Tan and Subramonyam 2024). This iterative process was essential to ensure that the overall protocol and the spe-

cific questions remained aligned with our research goals and effectively surfaced participants’ needs and challenges.

First, participants were asked to share their current model development workflow, and where and why they usually deploy their creations. This provided a foundational understanding of how they build and distribute their AI models, as well as what they prioritize in their workflows. Next, participants shared their experiences and opinions regarding how they practice RAI approaches in their workflows, based on the existing approaches proposed by model marketplaces and RAI researchers. We focused on two approaches: (1) a safety guideline for responsible model development (Civit AI 2025) and (2) a model attribution tool combined with an AI-generated content (AIGC) indicator.

- **Responsible Model Development Guideline:** It describes what model creators should consider during the training and testing process of their models. It introduces prohibited image output types (e.g., CSAM, NCII, publicity violation) and prohibited custom data used for their creations (e.g., Do not create resources with the purpose of depicting extreme fetishes or illegal conduct, such as rape, sexual abuse, scat or bestiality.) We used an existing guideline introduced by current model marketplaces as ‘Responsible Resource Development’ (Civit AI 2025).
- **Model Attribution tool & AIGC indicator:** Since static model cards can be complicated for non-AI practitioners to engage with (Mitchell et al. 2019; Turri et al. 2024; Tang et al. 2025), we used an improved design inspired by He et al. (He, Houde, and Weisz 2025), originally in-

tended to disclose AI attribution of AIGC. In this design, users can select options to disclose upstream dependencies, such as the pre-trained models and datasets they used, and the attribution of their models' creative concepts. We also presented an AIGC indicator illustrating what downstream users might see when images are generated using participants' models. Following design conventions from the Coalition for Content Provenance and Authenticity (C2PA) (Content Authenticity Initiative 2024), the indicator displays the original pre-trained model, the creator's creative concept, the creator and model names, and a link to access the model.

We presented each RAI approach and asked which, if any, they currently use in their workflows and how they apply them; for those who did not currently adopt these approaches, we walked through how the approaches work. Then, we asked in detail how they would adopt them and what they would expect to achieve through the approaches.

Additionally, beyond evaluating RAI approaches presented, we asked participants to envision a platform regime that prioritizes their personal values. This helped us understand emerging regulatory needs and the potential benefits such a regime would offer their workflows. Interviews were conducted online and an online visual board was used to present probes (Miro 2025). The detailed content is available at <https://osf.io/e6a3p>.

Data Analysis

Interviews were transcribed via Otter.AI (Otter.ai 2025); Korean interviews (P2, P19) were translated using DeepL (DeepL SE 2025). We employed reflexive thematic analysis (Braun and Clarke 2006, 2019) to identify patterns of meaning. In an initial code-alignment workshop, three authors collaboratively sorted and interpreted quotes from three transcripts to develop preliminary themes, guided by the research questions. Based on these initial themes, two authors independently coded the remaining transcripts. When new themes emerged, three authors engaged in discussion to refine them. Finally, the first author reviewed the full thematic structure and reconciled codes to ensure consistency in level of abstraction.

Ethical Considerations To create a comfortable environment where participants could share their perspectives candidly and to minimize social desirability bias, all interviews were conducted online by the first author alone. Before beginning each interview, the researcher explained the purpose of the study and assured participants that there were no appropriate answers, so as not to steer responses toward any predetermined direction. Transcripts were subsequently shared with the research team for collaborative analysis. Throughout the process, we took care not to place participants in ethically uncomfortable situations.

4 Finding 1: Individual Model Creators' Two Coexisting Logics of Accountability

Individual model creators conceptualized their accountability for their derivatives under two distinct logics — (1)

commons-based peer production and (2) personal creative property. These two accountability approaches coexisted but could lead to different accountability practices, including how creators identified the downstream impact and harm of their derivative models and how they addressed these concerns in ways that might diverge from platform policies.

Commons-Based Accountability: Reciprocity as the Basis of Responsibility

Echoing peer production of open-source software and remix cultures (Monroy-Hernández et al. 2011; Bartsch et al. 2024), derivative models are perceived as **artifacts socially produced through the participation of communities**. P14 framed his work as part of a collective innovation effort with his peers: *"It (building a model) is, at the moment, a small step, whatever we're doing. We will maybe be the best available in our field."* (P14) In their communities, composed of small groups of creators apart from model marketplaces, creators' work-in-progress and resources (e.g., ComfyUI workflow templates for fine-tuning models) are voluntarily shared, mostly produced based on members' intellectual curiosity. Additionally, model modification requires different technical parameters (e.g., datasets, hyperparameters) for different visual outcomes, making peer feedback essential for creators to refine their workflows. Peers' practical feedback, such as resolving error messages and testing reports with their checkpoint files, helps them in the iterative refinement process. Thus, being responsible is a way to consistently remain within their communities and build their models, which depend on others' voluntary contributions. P6 mentioned: *"It (Crediting models I used) helps (increasing their) visibility and also builds trust in the community."*

Considering the impact of their derivative models, creators viewed their accountability as limited to fellow creators within their communities, rather than extending to the broader upstream and downstream stakeholders (e.g., users). Echoing Widder et al.'s findings (Widder et al. 2022), participants considered unexpected use of models by downstream users to be uncontrollable, leaving responsibility to users to behave in good faith. For upstream data owners, whose image data may have been scraped by creators for the fine-tuning process, participants framed their data collection process as legally compliant under *fair use*, by which the scraped images were used as guides rather than resources to be directly taken. P19 mentioned, *"These (derivative) models are transformative in such a way. [...] I think the laws are still developing, and I generally don't feel I am stealing anything, because these models very rarely can reproduce the original work (images) in the first place."* Within this norm, their approaches in their workflow are limited to follow other community members' data practices, not clearly recognizing the impact of their derivatives in open model ecosystem.

Instead, giving credit when participants distribute their models is a *virtue* that retains motivation for their volunteer work within their community (Benkler and Nissenbaum 2006). 'Credit' is attributed to appreciation of other fellow creators' assistance to their models through resources, rather than accurate provenance disclosure when they used others' pipeline templates or checkpoint files (P3, P16). For

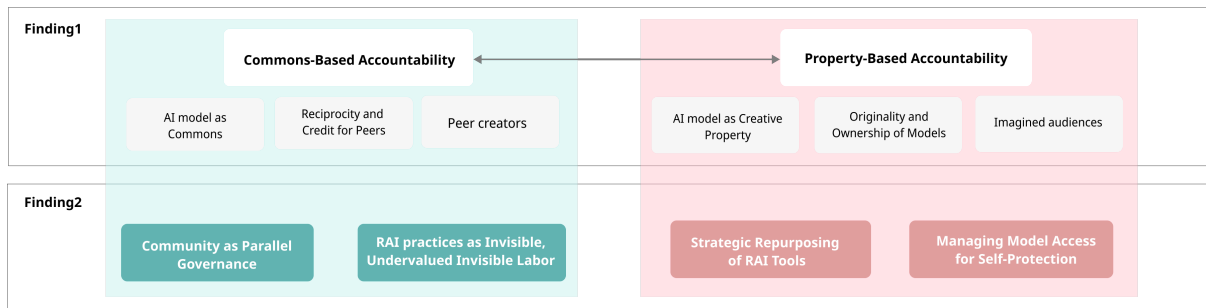


Figure 1: We summarize our findings. Finding 1 examines how creators conceptualize accountability based on their values, while Finding 2 traces how each accountability logic shapes their engagement with self-regulatory Responsible AI approaches.

instance, they encouraged other downstream users to introduce their usernames or unique numbers to signal how their content is attributed. However, these actions are reciprocal implicit expectations across model creators with no rules or guidelines indicating what and how model creators should disclose as attribution (P1, P6, P19). Thus, **these actions mostly depend on model creators’ personal judgment** in calculating how their derivative models are sourced from other model creators’ work and datasets shared in communities.

Property-Based Accountability: Originality and Ownership as the Basis of Responsibility

Users of derivative models were often framed as ‘engineers’ contributing to community-based models or company products (Widder et al. 2022; Gorwa and Veale 2024). By contrast, in the image-generation model context, **participants identified themselves as independent “artists” or “creators”** who build artifacts that generate experimental images using existing ingredients—pre-trained models. P16 said, “A good model creator is a creative, decisive one.” Collecting images, pre-trained models to use, and adapters to transfer subjects or styles (e.g., IPAdaptor) requires multiple refinements to gain expected image outputs and these ideas are obtained through multiple individuals’ trial and error. Thus, participants considered their derivatives as original artifacts sourced from their ideas. P13 mentioned, “As time goes by, you have to practice (your) originality, where you can come up with something unique, something original for you.”

Herein, ‘originality’ was used to express the novelty of their models, compared to other fine-tuned models and pre-trained models they used. The idea for originality began with examining weaknesses of pre-trained models and brainstorming how to modify the models (P2, P5, P10, P16). P10 mentioned, “when I wanted to generate images of ‘Hanbok [Korean traditional costume]’ (from Stable Diffusion XL), ‘Hanbok’ prompt either produced something entirely unrelated or brought up images of Kimono [Japanese traditional costume]. I collected images of Hanbok and manually tagged each of them with text labels, and then used that dataset to generate images.” As previous researchers noted (Suresh and Gutttag 2021), a series of choices and practices driven from their social identities, experiences were embedded in collecting data and choosing concepts.

This made them feel as though the models represent their identity. P5, who creates Afro style models, said, “*Most creatives lean towards the Eurocentric side. I wanted something different. I wanted to portray myself in the model.*”

This perception created strong psychological ownership of derivative models, leading them to treat their derivatives as **potential future income sources**. They wanted to gain full accountability for their derivatives throughout their workflow. Currently, opportunities to earn revenue on model marketplaces are limited (Duan et al. 2024), but creators believe that their models would eventually function as licensed assets, for which they would be compensated once legal frameworks and technical infrastructure more clearly recognize their technical efforts. P13 mentioned, “*In the long run, it should be mine, and you can’t be proud of something if you use [...] someone else’s property. That is why I always try to ensure I don’t use anyone’s licensed property.*” Likewise, their practices were self-protective, rather than an ethical obligation to upstream creators. Similarly, P15 mentioned “*I own the right to use every image that I collected and avoid any copyright issues that may later affect my model when it becomes officially public.*” They engaged in data practices they believed to be ‘ethical’ in their workflow, which were not validated and were mostly gained from community knowledge-sharing. These include collecting images on their own (e.g., taking pictures by visiting a hair salon — P1), blurring facial identities through AI, generating images using AI image generators (e.g., Leonardo), or choosing images that have no person or creative commons (P15). Given this perception, they described other creators who replicated their creative concepts without attribution as ‘thieves’ and ‘pirates’ who took their benefits (P1, P11, P17, P18).

Downstream users who might use their models for image generation or re-use them for another derivative were a central consideration when participants decided their creative concepts, echoing creative workers who aspire to high visibility on other social media platforms (e.g., YouTubers, Instagrammers) (P10, P14, P17) (Simpson and Semaan 2023; Tafesse and Dayan 2023). Their norms are centered on imagined audiences’ satisfaction (Duffy, Pruchniewska, and Scolere 2017), negotiating between normative ethics and the users’ preferences. P9 mentioned, “*when things are ethical, they (derivative models) are boring.*” He attributed the continued creation of sexually offensive

image-generation models to audience demand, noting that audiences often want to generate sexually or politically inciting images. Hence, their focus sometimes ignores IP rights holders' values to maximize audiences' attention. P4 created models that enabled people to generate memes of a specific soccer player: *"Everyone was waiting for him [the Soccer player] to not do it [score], and I created a meme (using my custom model) in case he failed to do it [score]. It really went viral."*

5 Finding 2: Divergent Engagements with Platform Responsible AI Practices

Having identified the two logics through which creators conceptualize accountability for their derivative models, we now turn to how they engage with Responsible AI approaches — safety guidelines and attribution tools — provided by platforms.

Creator Community as Parallel Governance to Displace Platform Authority

Rather than model marketplaces' official policy, **participants heavily relied on guidance and resources of communities for developing their models**, where model creators are committed to exchanging ideas, sharing their outputs, and collectively gaining feedback outside of those public platforms. P14 mentioned, *"I have a small community on Discord, mostly for testing. I have people there, so when I make a small change, they just test it and tell me, 'Okay, now there's a problem here.'"* As found in prior literature (Sadek et al. 2025), the static guideline for responsible model development does not provide actionable ways to work for their unique and diverse workflows. P6 mentioned, *"I often feel constrained by vague or overly strict content policies... the platform should clearly define what's allowed and what's not."* Since they often create models that generate experimental images, they needed to devise how to fit the rules to their practices. P15 mentioned, *"Learning from other models is really important for me, since it makes me understand and get better prompts for my own model. I can have something to twist to fit in on my model."* Thus, from the communities, participants obtained more approachable tips and tricks that they could directly adopt to their models (P1, P3, P13, P16, P17). For instance, P1 shared his Google Colab link in his Discord community channel, then other community members tested them and advised him on how to revise the code.

Such reliance on community is also grounded in an overall distrust of model marketplace platforms. P2 mentioned, *"If more users come in and there's more traffic, the platform succeeds. So I'm somewhat skeptical that platforms can actually solve ethical problems... at first there was no censorship at all (in CivitAI). It didn't matter whether minors came in or not."* The safety guideline was believed to be a performative and symbolic action that displaces accountability onto individual users. Even if they follow the guideline, risks may persist if downstream users are willing to tweak to generate explicit images; since the solution is imperfect, they viewed this guideline as political (Gillespie 2010), pri-

oritizing platforms' interests over user support rather than enabling open sharing. Participants thought consistent and comprehensive communication with users regarding platforms' policy changes is required to enable creators to adopt the changes quickly. P15 mentioned, *"You see the policies normally change. (When) you wake up, there's this policy that has been introduced by the government that requires everybody."* From that sense, they believed that their creator community is more reliable than platforms (P10, P11, P14, P17). P17 said *"I always say that it's good we keep the communities as small as possible. It promotes a sense of responsibility."*

However, their reliance on community commitment might reflect a skewed viewpoint of community-prioritized norms. Participants followed 'the same as what others do', focusing on improving 'quality' of image outputs when developing their derivatives (P3, P6, P13, P18). In this sense, they lacked an inclusive perspective in representing cultures, gender, and other dimensions of diversity, leaving some participants unable to find adequate resources within their communities (P2, P16). For instance, most derivative models shared in communities almost exclusively represent women as sexual symbols, making it difficult for P2 to apply community resources to her workflow. This indicates that model creators' responsible development workflows can be established or restructured by what the majority of model creators in their associated community endorse.

RAI Practices as Invisible, Undervalued Labor in Communities

Although participants recognized the need to reduce potential risks of their models in development process, they perceived practicing RAI approaches as those requiring massive invisible labor that goes unrecognized by platforms and audiences. P3 noted, *"it would be a lot of work on my side to work on such a big thing. But if I successfully execute such a thing, it would be an added advantage to whoever is going to interact with my model."* In open-source software development, code contributors often perform activities (e.g., version control) to operate their projects as active communities, which are often invisible and emotionally intensive (Geiger, Howard, and Irani 2021). Similarly, the effects of RAI practices are rarely visible in their model outputs, meaning that safer models depend on creators' willingness to perform labor whose results others cannot easily see. P10 described, *"I don't really compare prompts one by one. Because if I did that too, I would truly have no time left and nothing would get produced."*

Examining upstream models through audits fell solely on individual creators when those pre-trained models did not provide enough information for participants to judge whether they contain bias (P2, P6). *"When using multiple downloaded checkpoints, it's hard to confirm if the underlying models introduce biases or hidden content that violate platform policies."* (P6). However, their practices were often limited to testing models with prompts that downstream users would likely use (P10). Participants mentioned that *"Risks are always sleeping"* (P16), questioning whether risks could be fully removed even through rigorous regula-

tion of their development process. Because these practices also required substantial time and money for GPU usage, they considered robust evaluation as “*not ideal*.” As a result, this labor is perceived as undervalued and overlooked, and practices to mitigate risks were often minimized, left to individual creators to carry out according to their own criteria.

Strategic Repurposing of RAI Tools for Creator Visibility

Given the characteristics of derivative models, participants felt that the attribution tool would not create clear attribution of their upstream models. P2 noted, “*Practically all existing models are scraped from somewhere, so there are no authentic models*.” Because derivative models emerge through complex iterations — involving pipeline templates, custom data scraping, and other creators’ resources — participants found it difficult to determine the appropriate scope of attribution for their models. P3 said, “*That person (upstream creators) also brought in someone else’s model anyway, so it feels like the open-source spirit. But if I just directly used the thing they’re selling as-is, then I think I should credit that person*.” Therefore, there are mixed opinions across participants in deciding who and what should be credited. P18 mentioned, “*I think it’s also important to include the original data owners. [...] I’ve seen AI [tools] do this, so other creators that the content has been borrowed from — if it could be acknowledged, because it would have respected the original content owner*.”

This uncertainty led creators to reinterpret attribution tools and AIGC indicators not as transparency mechanisms, but as marketing features that could enhance the visibility of their derivative models (P6, P9). “*This (tool – AIGC indicator) would help in discoverability and trust, because when many users mention my page and people generate images using my model, it will increase my [model] discoverability*” (P6). They viewed the indicator as a branding signature that could authenticate their AI-generated images and signal credibility to audiences (P9). Thus, although accurate and clear data provenance is important, participants wished to re-organize content to disclose based on how downstream users would be attentive to their models. They were cautious about whether AIGC indicators and their attribution disclosure would negatively impact their model visibility (P2, P3, P7, P17, P18): “*It is important that every piece of information you put as an attribution has a target [audience] and needs to be there. When you put the attribution, it is a good idea to make sure that any information you think is not necessary should not be included*.” (P17) These reactions suggest that creators selectively appropriate attribution tools — disclosing information strategically to maximize visibility rather than transparently conveying model provenance as intended by platform designers.

Managing Model Access for Self-Protection

Participants were concerned that their models might be exploited or targeted by harassment from unidentified downstream users. Open model marketplaces have the advantage

of reaching larger audiences, as P12 noted, “*it’s way easier for me to access more people there who have the same vision of creating models*.” However, they worried that audiences might criticize them if their models produced negative results. P14 described, “*you can’t control everyone’s movement, and then it can be traced back to you. I thought the creator is a different entity from the creation, but now it is just coming to me that it can be used against me*.” Once models were uploaded, unexpected downstream use or negative reactions became largely uncontrollable. To mitigate these risks, participants established their own gatekeeping strategies, utilizing platform features such as access controls and disclaimers to delineate accountability boundaries. P1 described, “*Other model creators cannot replicate my model because I put guidelines — my licensing, watermarks on my work, and also disable downloading my content. It’s for read-only*.” On the other hand, P10 deliberately obscured model information to reduce discoverability. He scraped social media images of ‘Hanbok’ (traditional Korean attire) for his models, although he recognized this practice verged on infringement: “*Even though the model I made isn’t one that could really get me into trouble, I still worried that if I uploaded it there, people might later question how I built my dataset. It wasn’t a concrete fear, but more like a lingering concern [...] I just make sure the name isn’t associated with ‘Hanbok’ so it doesn’t show up in searches. I avoid providing as much information as possible*.”

Five participants shared their models exclusively within private communities through Google Colab or Drive links. These arrangements were grounded in trust built through sustained social interactions — with close school friends, colleagues in similar fields (e.g., fashion, anime fandom), and online acquaintances. P11 described, “*They [people who I share a model with] are my friend[s]. We have been through a lot together. So, yeah, I really trust them*.” Through these relationships, participants could anticipate how members would behave and how they should respond if problems arose from model use. P9 noted, “*I can get direct messages from these people, and understand different individuals according to their characteristics*.”

6 Discussion

Open model users—people who download, modify, and distribute open models—have received growing attention from researchers seeking to understand how harms are produced in the open model ecosystem (Widder et al. 2022; Tang et al. 2025; Lee et al. 2026; Choksi, Mandel, and Benthall 2025). We focused particularly on individual creators creating image generation models. Our findings showed that contributing to open model development—especially creating derivatives of upstream models—can be motivated not only by enjoyment-based collaboration but also by the personal as-sertization of one’s creations. These varied motivations can lead to different responsible AI practices that are considered acceptable only within particular community norms. In Lee et al.’s work, open model users adopt open models for entrepreneurial benefit, valuing less restrictive licenses and reduced guardrails (Lee et al. 2026). However, if creators release their modified models as open models without clear

data disclosures or attribution practices, such models are no longer truly “open” from downstream users’ perspectives, since auditing them requires substantial resources.

Individual Creators as a Locus of Risk in the Open Model Ecosystem

Our findings show that individuals without formal organizational processes, resources, or goals exhibit distinctive behaviors in how they develop derivative models and how their perceived accountability shifts depending on their motivations and creative goals. However, these characteristics may in turn undermine the transparency of their models, introducing risks for downstream dependencies.

First, **individuals’ workflows are largely self-directed**, and continuously adjusted based on their creative visions and peer feedback. In organizational contexts, AI practitioners’ workflows are structured, collaborative, and divided (Widder and Nafus 2023). They discuss the legitimacy of their practices with internal legal teams or AI compliance experts (Bogucka et al. 2024; Balayn et al. 2025). By contrast, individual creators manage nearly all aspects of their workflow independently, drawing on techniques and resources from community knowledge and cross-platform channels. Their decisions about model development are largely grounded in self-judgment and operate within legal ambiguity (e.g., fair use). Thus, unless creators proactively seek ethical guidance for their workflow from communities, unintended dependencies may emerge that obscure critical characteristics of their models, introducing vulnerabilities for downstream users (Hopkins et al. 2025).

Moreover, **individuals’ derivative models reflect their personal experiences, value orientations, and domain knowledge**. The choices creators make in collecting, curating, and annotating data, fine-tuning pre-trained models to achieve desired image outputs, and evaluating model performance are all shaped by their individual subjectivity. These situated standpoints can enable derivative models to address socio-cultural needs that general-purpose models often fail to adequately capture — such as fine-tuning for low-resource cultural image representation (Zhang, Doughty, and Snoek 2024). However, at the same time, this subjectivity can reproduce dominant community norms when creators lack reflexivity about their own positionality, risking the unintentional embedding of biases through data collection practices and decisions about target image outputs (Winner 2017; Scheuerman, Hanna, and Denton 2021). As our findings showed (Section 5), community norms can reproduce skewed representation, reflecting the demographic composition of creator communities. Such biases can be further compounded by upstream pre-trained models that may disproportionately misrepresent or underrepresent particular cultures, identities, and social groups (Shelby et al. 2024; Katzman et al. 2023; Bird, Ungless, and Kasirzadeh 2023).

Additionally, **individuals’ motivations for creating and distributing derivative models differ from those of organizations that adopt open models within specific domains**. Some individuals position themselves as solo entrepreneurs who generate financial value through their creations, embracing the monetization affordances of model

marketplace platforms (Bhargava 2022). Others view their work as part of remix and ‘participatory culture’ (Jenkins 2006), in which creators produce amateur media artifacts and contribute to peer communities. This orientation leads them to treat their derivatives as communal objects (Benkler and Nissenbaum 2006), while the broader social consequences of their work often go unexamined, as they frame their activities as recreational and harmless within their peer communities (Jenkins 2007).

Finally, **individuals’ perceived accountability is a moving target**; it is continually negotiated in relation to internal and external factors, including the degree of psychological ownership they hold over their creations (Polimetta, Gero, and Glassman 2026), the imagined audiences whose appreciation they anticipate (Duffy, Pruchniewska, and Scolere 2017), and how they interpret prevailing norms through platform governance (Jakesch et al. 2022). When actors using AI are situated outside the technology companies that govern its deployment, uncertainty around AI governance and insufficient information about upstream models hinder their ability to enact RAI practices (Kawakami, Wilkinson, and Chouldechova 2024). This challenge emerges for individual creators; the untraceability of downstream use, combined with the high accessibility of model marketplace platforms, generates considerable uncertainty, leading them to prioritize self-protection in their engagement with RAI practices.

Taken together, our findings indicate that individual creators may constitute a significant bottleneck for responsible AI in the open model ecosystem. The risks of open models can compound as models are re-used and re-derived by individuals who are themselves shaped by platforms, communities, and end-users.

Policy Implications

Open model marketplaces face a contested question of regulatory intensity (Gorwa and Veale 2024), but given the risks that individual creators may introduce to the open model ecosystem, we argue that open model marketplaces should adopt a more proactive moderation stance, moving beyond their position as neutral intermediaries — particularly for derivative models. Encouragement for practicing RAI approaches alone is insufficient to ensure that individuals exercise them adequately.

Rewarding Responsible Practice, Not Just Prohibiting Harm Not engaging with responsible AI practices is not technically a case of ‘legitimately wronging’ someone. Rather, it relies on individuals’ intrinsic motivation (Sadek et al. 2025). Thus, it should be treated differently from technology-enabled sexual abuse (synthetic NCII, CSAM), which is strongly prohibited on current open model marketplace platforms. Platforms should orient their moderation strategies toward incentivizing creators who contribute to safer model ecosystems by demonstrating their RAI practices, complementing guidelines that focus solely on prohibitions. Positive incentives, such as visibility boosts and attribution recognition, can provide strong motivation for creators seeking broader audience engagement to invest meaningfully in RAI practices (Gillespie 2022; Törhönen et al.

2020). Concretely, rather than surfacing only quantitative indicators of platform activity (e.g., download counts, follower numbers) on creator profiles, platforms could foreground qualitative indicators of responsible practice — such as creators’ documented evaluation approaches and dataset curation methods — as visible signals of accountability work (Marlow and Dabbish 2013). However, platforms should also consider the emergence of creators who game the moderation systems to maximize these incentives, or struggle to interpret moderation criteria when they remain opaque (Ma and Kou 2021; Choi et al. 2023). To mitigate these risks, platforms should clearly communicate moderation criteria and mechanisms to creators through transparent, externally-facing policies.

Removing Barriers to Creators’ RAI Engagement Platforms should commit not only to protecting public safety by regulating models that produce harmful outputs, but also to safeguarding creators against the perceived risks that discourage meaningful engagement with RAI practices enhancing model transparency — such as reputational harms from anonymous audiences, unauthorized use of their models, and plagiarism by other creators. Enhanced model transparency can expose creators to personal attacks, particularly historically marginalized creators who have long been disproportionately subjected to audience harassment, ranging from identity-based attacks to coordinated negative reviews (Thomas et al. 2022). Such harassment renders creators vulnerable, degrades their platform experiences, and may ultimately drive them off the platforms entirely. Open model marketplaces, as an emerging context for cultural production, may give rise to distinct forms of these harms, as creators have limited capacity to trace how their derivative models are used or misused after deployment (Widder et al. 2022).

Cultivating Stewardship Culture Among Creators for Responsible Practice To prevent unintended harms from individual creators’ misjudgment, platforms should cultivate a culture of stewardship in which creators take responsibility for guiding one another toward RAI practices. Stewardship has well-established precedent in open source software culture, where it has been shown to lower maintenance costs, increase flexibility, and improve operational efficiency (Tridgell and Singh 2025; Priest 1997; O’Mahony 2007). While OSS stewardship has traditionally focused on codebase maintenance, the underlying governance structures — distributed authority, peer accountability, and community-defined norms — could be productively adapted to encompass ethical guidance and harm prevention in the open model marketplace context. However, as our findings (Finding 5.1) indicate, community-based governance can also reproduce dominant norms and exclude marginalized perspectives. Stewardship structures therefore require accompanying support for inclusive participation and oversight, such as cross-community auditing, in which creators from different communities periodically review one another’s practices to surface blind spots that may go unnoticed within a single community.

Recognizing Derivative Models in Attribution Framework Beyond governance focused on internal stakeholders, platforms should also engage external stakeholders in the open model ecosystem. Existing frameworks for standardizing AI attribution and content authenticity (Content Authenticity Initiative 2024; He, Houde, and Weisz 2025) have largely overlooked individual model creators as relevant stakeholders, as well as the distinctive characteristics of their models. At present, AI companies log digital fingerprints and edit histories as metadata when end-users generate images with AI tools, while social media platforms display AIGC indicators (e.g., “AI info”) to signal AI-generated content. As these frameworks extend to derivative model contexts, the granularity of modification methods and fine-tuning sources will likely become important dimensions of attribution. When downstream users rely on derivative models to generate images, who should be credited as the primary contributor to the resulting content? This question becomes especially complicated when creators substantially customize pre-trained models — by altering stylistic character or fine-tuning with training data from unverified sources (Hacker and Holweg 2026). Moreover, if derivative models are documented within AIGC attribution systems, AIGC indicators would shift in function — no longer serving solely as disclosure mechanisms for AI use, but also operating as forms of credit recognizing creators’ contributions.

In light of these complexities, future attribution framework for derivative models should engage model creators directly in defining what modifications warrant disclosure, how their contributions should be represented, and what ethical concerns emerge from their own perspectives.

Limitations

We acknowledge several limitations. First, we focused exclusively on the viewpoints of creators of image generation models, particularly Text-to-Image models. The needs and challenges they encounter to responsibly develop models may vary for different modalities (e.g., text and speech). Whereas images convey creative ideas visually, making originality easier to recognize, creators of language models may feel a less pressing need for originality. We also acknowledge that small samples may not fully represent individual creators’ voices. Also, our research participants were based in the U.S., so findings may not be globally generalizable. Owing to differing jurisdictions, users in other countries may have differing perspectives on our results.

Conclusion

Our study surfaces how creators engage with current governance approaches and remedies to support them for Responsible AI, based on their perceptions and experiences. We believe our findings offer insights into how individuals may act as hidden actors who affect the trust and safety of the model marketplace, and broadly the open model ecosystem, by developing and sharing derivative models on platforms. We also shed light on what factors shape how these individuals construct their perceived accountability within this emerging yet still niche platform ecosystem.

Positionality Statement

Our research team consists of scholars with backgrounds in cultural production, human-AI interaction, and privacy. We have prior experience conducting qualitative research on how AI governance shapes the behavior of social media users, freelance creators, and creative workers. This combined expertise directs our analytical attention to both individual creators' situated experiences and the broader platform structures that shape them. We therefore approached this project with a focus on creators' lived experiences on emerging platforms and on how platform governance operates in relation to those experiences.

We acknowledge that our positioning—as researchers based in U.S. institutions—may shape what we attend to and what we overlook, and that creators in other geographic or cultural contexts may navigate platform governance differently. Throughout the analysis, we worked reflexively to surface tensions between individual creators' practices and the structural conditions imposed by platforms, while remaining alert to how our own disciplinary lenses framed what counted as a meaningful finding.

Acknowledgements

We thank our participants for generously sharing their thoughts and experiences with us. We also thank the anonymous reviewers, whose feedback helped us strengthen this work. The first author deeply appreciates Dr. danah boyd, Dr. Karen Levy, Dr. Ned Cooper, and Daniel Mwesigwa in providing their advices to improve this work. We utilized ChatGPT to correct grammar and create Tables in the manuscript. We did not use any AI-driven content.

References

- Aksulu, A.; and Wade, M. R. 2010. A comprehensive review and synthesis of open source research. *Journal of the Association for Information Systems*, 11(11): 6.
- Anisuzzaman, D.; Malins, J. G.; Friedman, P. A.; and Attia, Z. I. 2025. Fine-tuning large language models for specialized use cases. *Mayo Clinic Proceedings: Digital Health*, 3(1): 100184.
- Balayn, A.; Yurrita, M.; Rancourt, F.; Casati, F.; and Gadi-raj, U. 2025. Unpacking Trust Dynamics in the LLM Supply Chain: An Empirical Exploration to Foster Trustworthy LLM Production & Use. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–20.
- Bartsch, S. C.; Lothar, M.; Schmidt, J.-H.; Adam, M.; and Benlian, A. 2024. The origin and opportunities of developers' perceived code accountability in open source ai software development. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 94–106.
- Benkler, Y.; and Nissenbaum, H. 2006. Commons-based peer production and virtue. *Journal of political philosophy*, 14(4).
- Berman, G.; Goyal, N.; and Madaio, M. 2024. A Scoping Study of Evaluation Practices for Responsible AI Tools: Steps Towards Effectiveness Evaluations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–24.
- Bhargava, H. K. 2022. The creator economy: Managing ecosystem supply, revenue sharing, and platform design. *Management Science*, 68(7): 5233–5251.
- Birch, K.; and Muniesa, F. 2020. *Assetization: Turning things into assets in technoscientific capitalism*. MIT press.
- Bird, C.; Ungless, E.; and Kasirzadeh, A. 2023. Typology of risks of generative text-to-image models. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 396–410.
- Bogucka, E.; Constantinides, M.; Šćepanović, S.; and Quercia, D. 2024. Co-designing an AI impact assessment report template with AI practitioners and AI compliance experts. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 168–180.
- Bovens, M. 2010. Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics*, 33(5): 946–967.
- Braun, V.; and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2): 77–101.
- Braun, V.; and Clarke, V. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4): 589–597.
- Canet Sola, M.; and Guljajeva, V. 2026. Why Open Small AI Models Matter for Interactive Art. In *Proceedings of the 2025 12th International Conference on Digital and Interactive Arts: Media Art Cultures, Communities & Territories*, ARTECH '25. New York, NY, USA: Association for Computing Machinery. ISBN 9798400720017.
- Choi, Y.; Kang, E. J.; Lee, M. K.; and Kim, J. 2023. Creator-friendly algorithms: Behaviors, challenges, and design opportunities in algorithmic platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–22.
- Choksi, M. Z.; Mandel, I.; and Benthall, S. 2025. The Brief and Wondrous Life of Open Models. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 3224–3240.
- Civit AI. 2025. Safety Center: Policies and Guidelines. <https://civitai.com/safety>.
- Content Authenticity Initiative. 2024. Content Credentials. <https://www.contentcredentials.org/>. Accessed: 2025-09-03.
- Creswell, J. W.; and Creswell, J. D. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Creswell, J. W.; and Poth, C. N. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- DeepL SE. 2025. DeepL Translate: The World's Most Accurate Translator. <https://www.deepl.com/en/translator>.
- Duan, M.; Du, M.; Zhao, R.; Wang, M.; Wu, Y.; Shadbolt, N.; and He, B. 2025. Position: Current model licensing practices are dragging us into a quagmire of legal noncompliance. In *Forty-second International Conference on Machine Learning Position Paper Track*.

- Duan, M.; Zhao, R.; Jiang, L.; Shadbolt, N.; and He, B. 2024. "They've Stolen My GPL-Licensed Model!": Toward Standardized and Transparent Model Licensing. *arXiv preprint arXiv:2412.11483*.
- Duffy, B. E.; Pruchniewska, U.; and Scolere, L. 2017. Platform-specific self-branding: Imagined affordances of the social media ecology. In *Proceedings of the 8th international conference on social media & society*, 1–9.
- Eiras, F.; Petrov, A.; Vidgen, B.; Schroeder de Witt, C.; Pizati, F.; Elkins, K.; Mukhopadhyay, S.; Bibi, A.; Botos, C.; Steibel, F.; et al. 2024. Position: Near- to Mid-term Risks and Opportunities of Open-source Generative AI. *Journal of Machine Learning Research*.
- François, C.; Péran, L.; Bdeir, A.; Dziri, N.; Hawkins, W.; Jernite, Y.; Kapoor, S.; Shen, J.; Khlaaf, H.; Klyman, K.; et al. 2025. A Different Approach to AI Safety: Proceedings from the Columbia Convening on Openness in Artificial Intelligence and AI Safety. *arXiv preprint arXiv:2506.22183*.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Geiger, R. S.; Howard, D.; and Irani, L. 2021. The labor of maintaining and scaling free and open-source software projects. *Proceedings of the ACM on human-computer interaction*, 5(CSCW1): 1–28.
- Gillespie, T. 2010. The politics of 'platforms'. *New media & society*, 12(3): 347–364.
- Gillespie, T. 2022. Do not recommend? Reduction as a form of content moderation. *Social Media+ Society*, 8(3): 20563051221117552.
- Google Public Policy Blog. 2009. The Meaning of Open. <https://publicpolicy.googleblog.com/2009/12/meaning-of-open.html>.
- Gorwa, R.; and Veale, M. 2024. Moderating model marketplaces: Platform governance puzzles for AI intermediaries. *Law, Innovation and Technology*, 16(2): 341–391.
- Hacker, P.; and Holweg, M. 2026. The regulation of fine-tuning: Federated compliance for modified general-purpose AI models. *Computer Law & Security Review*, 60: 106234.
- Harris, D. E. 2023. How to regulate unsecured "Open-Source" AI: No exemptions. *Tech Policy Press*.
- He, J.; Houde, S.; and Weisz, J. D. 2025. Which contributions deserve credit? perceptions of attribution in human-ai co-creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18.
- Hopkins, A.; Cen, S. H.; Struckman, I.; Ilyas, A.; Videgaray, L.; and Mađry, A. 2025. Ai supply chains: An emerging ecosystem of ai actors, products, and services. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, 1266–1277.
- Jakesch, M.; Buçinca, Z.; Amershi, S.; and Olteanu, A. 2022. How different groups prioritize ethical values for responsible AI. In *proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 310–323.
- Jenkins, H. 2006. *Fans, bloggers, and gamers: Exploring participatory culture*. nyu Press.
- Jenkins, H. 2007. Confronting the challenges of participatory culture—media education for the 21st century (Part Two). *Nordic Journal of Digital Literacy*, 2(2): 97–113.
- Katzman, J.; Wang, A.; Scheuerman, M.; Blodgett, S. L.; Laird, K.; Wallach, H.; and Barocas, S. 2023. Taxonomizing and measuring representational harms: A look at image tagging. In *Proceedings of the AAAI Conference on artificial intelligence*, volume 37, 14277–14285.
- Kawakami, A.; Wilkinson, D.; and Chouldechova, A. 2024. Do responsible AI artifacts advance stakeholder goals? four key barriers perceived by legal and civil stakeholders. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 670–682.
- Kilamo, T.; Hammouda, I.; Mikkonen, T.; and Aaltonen, T. 2012. From proprietary to open source—Growing an open source ecosystem. *Journal of Systems and Software*, 85(7): 1467–1478.
- Kilcher, Y. 2022. ykilcher/gpt-4chan. <https://huggingface.co/ykilcher/gpt-4chan>. Accessed: 2025-09-03.
- Langenkamp, M.; and Yue, D. N. 2022. How open source machine learning software shapes ai. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 385–395.
- Laufer, B.; Oderinwale, H.; and Kleinberg, J. 2025. Anatomy of a Machine Learning Ecosystem: 2 Million Models on Hugging Face. *arXiv preprint arXiv:2508.06811*.
- Lee, W.; Howison, J.; Lee, M. K.; and Li, H. 2026. Open AI in the Wild: Adoption and Adaptation of Open Models on r/LocalLLaMA. In *The 2026 ACM Conference on Fairness, Accountability, and Transparency*, 7203–7236.
- Liesenfeld, A.; and Dingemans, M. 2024. Rethinking open source generative AI: open-washing and the EU AI Act. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1774–1787.
- Lima, G.; Grgić-Hlača, N.; and Redmiles, E. M. 2025. Public Opinions About Copyright for AI-Generated Art: The Role of Egocentricity, Competition, and Experience. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–32.
- Ma, R.; and Kou, Y. 2021. "How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–25.
- Maiberg, E. 2025a. Hugging Face Is Hosting 5,000 Nonconsensual AI Models of Real People. <https://www.404media.co/hugging-face-is-hosting-5-000-nonconsensual-ai-models-of-real-people/>. 404 Media. Accessed: 2026-05-21.
- Maiberg, E. 2025b. Hugging Face Is Hosting 5,000 Nonconsensual AI Models of Real People. *404 Media*. Accessed: 2026-05-18.
- Maiberg, E. 2025c. Payment Processors Are Pushing AI Porn Off Its Biggest Platforms. <https://www.404media.co/payment-processors-are-pushing-ai-porn-off-its-biggest-platforms/>. 404 Media.

- Marchal, N.; Xu, R.; Elasmr, R.; Gabriel, I.; Goldberg, B.; and Isaac, W. 2024. Generative AI misuse: A taxonomy of tactics and insights from real-world data. *arXiv preprint arXiv:2406.13843*.
- Marlow, J.; and Dabbish, L. 2013. Activity traces and signals in software developer recruitment and hiring. In *Proceedings of the 2013 conference on Computer supported co-operative work*, 145–156.
- Meneely, A.; and Williams, L. 2009. Secure open source collaboration: an empirical study of linus’ law. In *Proceedings of the 16th ACM conference on Computer and communications security*, 453–462.
- Meta. 2023. Industry Leading, Open-Source AI | Llama. <https://www.llama.com/>.
- Miro. 2025. Boards in Education Team - Miro. <https://miro.com/app/dashboard/>.
- Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, 220–229.
- Monroy-Hernández, A.; Hill, B. M.; Gonzalez-Rivero, J.; and Boyd, D. 2011. Computers can’t give credit: How automatic attribution falls short in an online remixing community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3421–3430.
- Otter.ai. 2025. Otter Voice Meeting Notes. <https://otter.ai/home>.
- O’Mahony, S. 2007. The governance of open source initiatives: what does it mean to be community managed? *Journal of Management & Governance*, 11(2): 139–150.
- Paris, T.; Moon, A.; and Guo, J. L. 2025. Opening the Scope of Openness in AI. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1293–1311.
- Parker, C.; Scott, S.; and Geddes, A. 2019. Snowball sampling. *SAGE research methods foundations*.
- Pepe, F.; Nardone, V.; Mastropaolo, A.; Bavota, G.; Canfora, G.; and Di Penta, M. 2024. How do hugging face models document datasets, bias, and licenses? an empirical study. In *Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension*, 370–381.
- Polimetla, T.; Gero, K. I.; and Glassman, E. L. 2026. A Paradigm for Creative Ownership. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 1–16.
- Priest, M. 1997. The privatization of regulation: Five models of self-regulation. *Ottawa L. Rev.*, 29: 233.
- Raffel, C. 2023. Building machine learning models like open source software. *Communications of the ACM*, 66(2): 38–40.
- Sadek, M.; Kallina, E.; Bohné, T.; Mougnot, C.; Calvo, R. A.; and Cave, S. 2025. Challenges of responsible AI in practice: scoping review and recommended actions. *AI & society*, 40(1): 199–215.
- Scheuerman, M. K.; Hanna, A.; and Denton, R. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on human-computer interaction*, 5(CSCW2): 1–37.
- Schneider, M.; and Hagendorff, T. 2025. Investigating toxicity and Bias in stable diffusion text-to-image models. *Scientific Reports*, 15(1): 31401.
- Shelby, R.; Srinivasan, R.; Burgdorf, K.; Lena, J. C.; and Rostamzadeh, N. 2024. Creative ML assemblages: the interactive politics of people, processes, and products. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1): 1–30.
- Simpson, E.; and Semaan, B. 2023. Rethinking creative labor: A sociotechnical examination of creativity & creative work on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16.
- Song, S.; Song, J.; Lee, J.; Kang, Y.; and Moon, H. 2024. Exploring the Potential of Novel Image-to-Text Generators as Prompt Engineers for CivitAI Models. In *2024 16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 626–631. IEEE.
- Suresh, H.; and Gutttag, J. 2021. A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9.
- Tafesse, W.; and Dayan, M. 2023. Content creators’ participation in the creator economy: Examining the effect of creators’ content sharing frequency on user engagement behavior on digital platforms. *Journal of Retailing and Consumer Services*, 73: 103357.
- Tan, J.; Vincent, N.; Elkins, K.; and Sahlgren, M. 2025. If open source is to win, it must go public. *arXiv preprint arXiv:2507.09296*.
- Tan, M.; and Subramonyam, H. 2024. More than model documentation: uncovering teachers’ bespoke information needs for informed classroom integration of ChatGPT. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–19.
- Tang, N.; Li, M.; Winecoff, A.; Madaio, M.; Heidari, H.; and Shen, H. 2025. Navigating Uncertainties: Understanding How GenAI Developers Document Their Models on Open-Source Platforms. *arXiv preprint arXiv:2503.23574*.
- Thomas, K.; Kelley, P. G.; Consolvo, S.; Samermit, P.; and Bursztein, E. 2022. “It’s common and a part of being a content creator”: Understanding How Creators Experience and Cope with Hate and Harassment Online. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–15.
- Törhönen, M.; Sjöblom, M.; Hassan, L.; and Hamari, J. 2020. Fame and fortune, or just fun? A study on why people create content on video platforms. *Internet Research*, 30(1): 165–190.
- Tridgell, J.; and Singh, J. 2025. ‘Stewardship’ as a Fair, Accountable and Transparent Model for Free and Open-Source

- Software Governance? Looking Beyond the EU’s Cyber Resilience Act. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 473–484.
- Turri, V.; Morrison, K.; Robinson, K.-M.; Abidi, C.; Perer, A.; Forlizzi, J.; and Dzombak, R. 2024. Transparency in the Wild: Navigating Transparency in a Deployed AI System to Broaden Need-Finding Approaches. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1494–1514.
- Vasilescu, B.; Capiluppi, A.; and Serebrenik, A. 2014. Gender, representation and online participation: A quantitative study. *Interacting with Computers*, 26(5): 488–511.
- Wei, Y.; Zhu, Y.; Hui, P.; and Tyson, G. 2024. Exploring the use of abusive generative AI models on Civitai. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6949–6958.
- Widder, D. G.; and Nafus, D. 2023. Dislocated accountabilities in the “AI supply chain”: Modularity and developers’ notions of responsibility. *Big Data & Society*, 10(1): 20539517231177620.
- Widder, D. G.; Nafus, D.; Dabbish, L.; and Herbsleb, J. 2022. Limits and possibilities for “ethical ai” in open source: A study of deepfakes. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2035–2046.
- Widder, D. G.; West, S.; and Whittaker, M. 2023. 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).
- Widder, D. G.; Whittaker, M.; and West, S. M. 2024. Why ‘open’ AI systems are actually closed, and why this matters. *Nature*, 635(8040): 827–833.
- Winner, L. 2017. Do artifacts have politics? In *Computer ethics*, 177–192. Routledge.
- WIRED. 2025. The Dark Side of Open Source AI Image Generators. <https://www.wired.com/story/dark-side-open-source-ai-image-generators/>.
- Zhang, Y.; Doughty, H.; and Snoek, C. G. 2024. Low-resource vision challenges for foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21956–21966.