Intellectual Property in the Age of Generative AI: How Social Group Priorities Surrounding Copyright are Changing in Response to Novel AI Tools

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Madelyn Khoury

Spring, 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Bryn E. Seabrook, Department of Engineering and Society

Intellectual Property in the Age of Generative AI: How Social Group Priorities Surrounding Copyright are Changing in Response to Novel AI Tools

Discussions on the Impacts of Generative AI

The past two years have seen the invention of many tools that employ generative artificial intelligence (AI), or computer programs that train on large datasets and learn how to create novel content from a prompt. These tools have exploded in popularity, with the top few tools having an average increase of over 200 million monthly visits. Such tools have been used for applications ranging from text generation to photo editing to the creation of chatbots with particular personalities (*AI Industry Analysis*, n.d.). Along with this rise in the usage of generative AI, conversations and conflicts surrounding generative AI usage have emerged. A recent study by Miyazaki et al. (2024) found that Twitter users talked about generative AI more than they talked about most other topics. Meanwhile, labor historians speculate that generative AI may change the workforce by allowing a single employee to do a job that previously required many, or conversely, by causing the emergence of new jobs that depend less on rote work (Malinsky, 2023). In a survey by Jobs for the Future, 58% of respondents felt that they needed to acquire new skills due to AI (*Artificial Intelligence & the Future of Work Survey*, n.d.).

The potential of generative AI to impact job security raises questions about the ethics of using it, particularly when it uses training data produced by humans to outcompete those humans in their own industry. For example, at the 2022 Colorado State Fair, Jason M. Allen won the top prize for digital art with a piece he used generative AI to create, provoking a negative response among artists (Ho, 2024). The same tool that Allen used is now facing a class action lawsuit from artists who claim that it "infringed on copyright by scraping artists' work from the internet without permission" (UNESCO, 2023).

Clearly, different social groups view generative AI in different ways, a phenomenon defined by Pinch and Bijker (1987) as "interpretive flexibility," a key aspect of their Social Construction of Technology (SCOT) framework. Using SCOT as the basis of its analysis, this paper asks: how have the priorities of key social groups regarding copyright changed in response to the advent of three popular generative AI tools?

Methods

To answer this question, I employed two strategies. First, I analyzed social media posts – specifically from Twitter – to determine if members of three social groups discussed different topics after the release of specific AI tools. Similarly to Miyazaki et al., I grouped Tweets by occupation of their author (using the poster's profile information to identify occupation) before further analysis. However, I further subdivided the data into eras before and after the release of each of the three relevant generative AI tools, which allowed for a comparison of how often groups discussed certain topics before and after the release of these AI tools. See Appendix A for details on the chosen tools and eras. I downloaded four months' worth of tweets from the Twitter Stream Grab project on the Internet Archive, a project meant to save tweets "for the purposes of research, history, testing and memory" (Scott, 2012). Downloading and extracting the data from these months resulted in a dataset of over 30 million tweets, which I then filtered into a set of about 8 million usable tweets. Figure 1 shows the quantity of tweets filtered out at each step in the process, as well as the number of tweets remaining from each of the four months. Finally, I grouped the tweets by the social group of their authors and conducted a thematic analysis of the

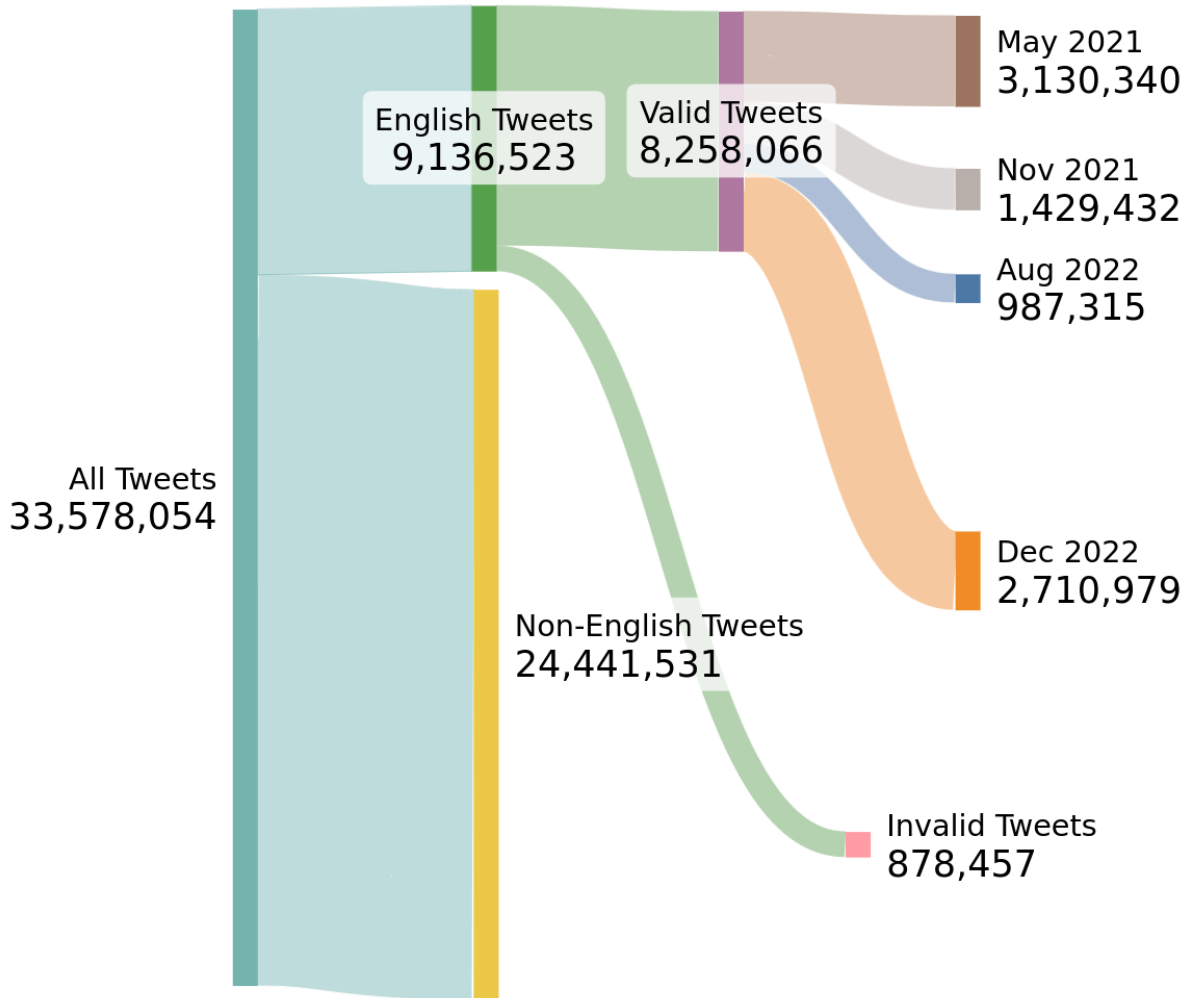tweets, specifically focusing on the occurrence of keywords relating to copyright.

Figure 1. Quantities of Tweets Processed. Created with SankeyMATIC (Bogart, n.d.).

For my second method, I sourced information on filed civil suits relating to both AI and copyright to determine which social groups have taken action to change copyright law as a response to AI. I gathered court data from two databases, Justia (*Justia Law*, 2024) and CourtListener, a product of the Free Law Project (Free Law Project, 2024). I compiled

information about the number of cases, the date of all cases, whether each case involved generative AI or other forms of AI, and the social group of all entities involved in each case.

Relevant Generative AI Tools

It is common to categorize generative AI tools based on the type of output they produce: chat conversations, text, images, graphic designs, videos, music, or other miscellaneous outputs (*AI Industry Analysis*, n.d.). This paper focuses on three specific tools that generate text, images, and computer programs, respectively. These tools are ChatGPT, Midjourney, and Github Copilot, and I selected them in order to observe the effect of content type on a generative AI tool's impact on discussions about copyright.

ChatGPT, a creation of AI research company OpenAI, is an AI chat bot capable of maintaining a conversation with users (*ChatGPT*, n.d.). Although its output is intended to be conversational, users can ask it to generate virtually any text-based content, such as movie scripts, short computer programs, or travel itineraries. Released on November 22nd, 2022, ChatGPT achieved "viral" status on social media, accumulating over one million users in just five days (Marr, 2023). The second generative AI tool of interest, Midjourney, was released a few months before, on July 13th, 2022 (Midjourney [@midjourney], 2022). Rather than producing text, Midjourney generates images ranging from photorealistic pieces to stylistic digital artworks, sometimes even emulating a particular artist (*Midjourney Feed*, n.d.). Finally, Github, a site for hosting software, released Github Copilot in 2021, a tool which it markets as an "AI pair programmer," or an AI assistant to help developers write code. Perhaps due to this marketing strategy, 55% of software engineers express support for the tool, seemingly unworried about being displaced by it (*GitHub Copilot · Your AI Pair Programmer*, n.d.). Still, some

computer scientists believe that programming will become obsolete due to similar tools in the future (Welsh, 2022).

Although all three of these models produce different outputs, they share a commonality in that their training datasets consist of works made by – and representing the intellectual property of – humans. ChatGPT was trained on multiple datasets, including the Stanford Natural Language Inference Corpus (Bowman et al., 2015) and the RACE dataset, which consist of passages written by authors and teachers (Lai et al., 2017). Copilot's developers, meanwhile, trained it on publicly available code on Github, which may not have been licensed for such use. Finally, evidence has surfaced that Midjourney was trained on the LAION dataset, which contains over 5 billion pre-processed images, each of which was created by a human (Schuhmann et al., 2022).

With all three of these tools, the human creators of the training data may express concern that their intellectual property is being stolen. Members of varying occupations have already taken efforts to voice — and mitigate — these concerns. For example, a group of artists filed a lawsuit against Midjourney, claiming that the company used the LAION database (which is intended only for academic use) in its commercial products and that Midjourney specifically included works from 16,000 artists in order to allow its model to emulate their styles (*Leaked*, 2024). The artists claim that models like Midjourney can only produce "derivative works" and thus infringe upon their copyright, while the generative AI companies refute this claim on the basis that models are trained on billions of images and are unlikely to reflect any single artist's work (Solomon, 2023). Meanwhile, in order to combat the unfair use of code for the training of generative AI models, Kocetkov et al. (2022) produced The Stack, a dataset consisting only of permissively-licensed code snippets.

Analysis of Generative AI via the Social Construction of Technology

The previous examples illustrate the breadth of opinions surrounding generative AI and intellectual property, as well as the fact that opinions tend to be similar within a group of people who share an occupation or role. This phenomenon makes sense when viewed through the lens of Pinch and Bijker's Social Construction of Technology framework (SCOT), which defines social groups as groups of people who share similar problems. According to Pinch and Bijker, technological artifacts are not inherently good or bad, but rather are perceived to be so based on whether they solve – or fail to solve – the problems facing a particular group (Pinch & Bijker, 1987). SCOT evaluates technological artifacts through four phases: interpretive flexibility, in which researchers observe the varying interpretations of the artifact, relevant social groups, in which they identify groups with a vested interest in how the artifact will be adopted; closure, in which they observe how each social group begins to share a common interpretation of an artifact; and wider context, in which they analyze how different social groups experience conflict due to their varying interpretations of the artifact.

I have already shared several examples of the interpretive flexibility of generative AI tools. For the purposes of this paper, the relevant social groups are artists, authors, and computer programmers, as each of these groups produce content that a novel generative AI tool is also capable of producing. Throughout the rest of this paper, I explore the closure within these groups and the conflicts among them. Of particular interest is how each group's interpretations of AI tools vary based on the type of content that the tool produces. For example, artists face the challenge of producing and marketing their artwork in order to make a living, and thus are likely to share a negative view of image-generative AI like Midjourney, as it does not solve their

problem, but rather contributes to it. Authors, however, may not see any issue with using Midjourney to generate art of their characters. In other words, it is possible that people may not protest AI tools that disrupt other industries, but would protest tools that disrupt their own. Meanwhile, people in unaffected industries (i.e., not belonging to any of the three relevant social groups) may not oppose the use of AI at all.

In fact, Miyazaki et al. (2024) examined how sentiments around generative AI varied by occupation in a recent study. They analyzed posts from Twitter to identify how individuals interacted with generative AI, which topics and tools individuals tended to talk about, and how sentiment changed as a result of greater interaction with generative AI, grouping and analyzing data by occupation. Most Twitter users expressed positive sentiment toward generative AI, and said sentiment tended to increase upon greater interaction with generative AI tools. Illustrators, however, held negative sentiments toward generative AI, except that which was used solely for producing computer programs. This paper builds upon Miyazaki et al.'s work, similarly analyzing social groups' opinions. However, this paper adds to the body of knowledge by 1) focusing explicitly on topics of interest relating to copyright and ownership, 2) evaluating how interests change after the advent of the three relevant generative AI tools, and 3) examining actions taken by social groups in an attempt to change copyright policies.

Miyazaki et al.'s work did not subdivide occupations into more specific social groups or consider other roles that might define a social group, such as "parents of young children" or "young adults entering the workforce." Although outside the scope of this paper, it may be useful to consider additional social groups, such as users who consider their use of generative AI as creative and thus want copyright over the prompts they used to produce outputs (Hayes, 2023).

Other previous research has focused on social group involvement in more insidious generative AI. Zhou and Ferguson (2022) use SCOT to analyze a particularly infamous technology: deepfakes, or generative AI that intends to create realistic photos and videos with the purpose of passing this media off as real. Zhou and Ferguson identify social groups invested in the development of deepfake technology as well as the conflicts among these groups, which include profit-driven organizations, end users, independent developers, researchers, policymakers, and advocates (those who raise public awareness about generative AI). Zhou and Ferguson also discuss how deepfake technology may achieve the SCOT principle of closure – in other words, how it may become no longer controversial. They argue that deepfake technology can be regulated – and countermeasures against it can be developed – so as to prevent it from doing harm, allowing it to be used only for beneficial purposes. However, achieving consensus about less obviously harmful generative AI tools, such as the three discussed in this paper, will likely require a more nuanced discussion and more complex policy changes.

Results and Discussion

All three social groups appear to have a growing interest in copyright and ownership due to the advent of popular generative AI tools. However, there does not appear to be a correlation between the type of content that generative AI produces and whether a social group is likely to change priorities due to its release. Still, the three social groups do not necessarily hold the same priorities surrounding generative AI. Members of all three groups have expressed their priorities by way of filing civil suits, and they frequently do so against members of a different social group, which indicates conflict among these groups.

Analysis of Twitter Posts Relating to Copyright

Individuals interpret any technological artifact through the lens of their own experience, grading the usefulness and morality of the artifact based on whether it addresses their needs (Pinch & Bijker, 1987). On a broader scale, groups of people with similar needs come to a shared interpretation of an artifact through a process called closure, which usually consists of public discussion (ibid). Thus, looking at public posts on social media such as Twitter allows researchers to observe the process of closure in real time. When researchers analyze trends through time, it may even be possible to make conclusions about the interpretations that different social groups have settled upon.

First, it is useful to examine the social composition of the social media content. I processed over 30 million tweets, though only about 8 million of these were valid for analysis. I screened each valid tweet for the social group of the tweet's author, specifically by searching for keywords in the author's description. Words such as "developer" or "programmer" indicated that a tweet was written by a computer programmer, while words such as "artist" or "painter" indicated that the tweet was written by an artist (see Appendix B for the full list of keywords). In addition to the three key social groups – artists, authors, and programmers – I also examined trends among all the other tweets, which were grouped in a catch-all "other" category. Figure 2 shows the social group composition of each of the four months of tweets that I analyzed, along with the composition of the total collection of tweets. As might be expected, the "other" category accounted for the majority of the tweets. Interestingly, the relative composition of the social groups does not fluctuate much over time, indicating that there is no influx or emigration of users of any social group to/from Twitter. In other words, key social groups did not begin posting more frequently as a result of generative AI tools being released, nor did they stop posting for fear of their content being scraped.
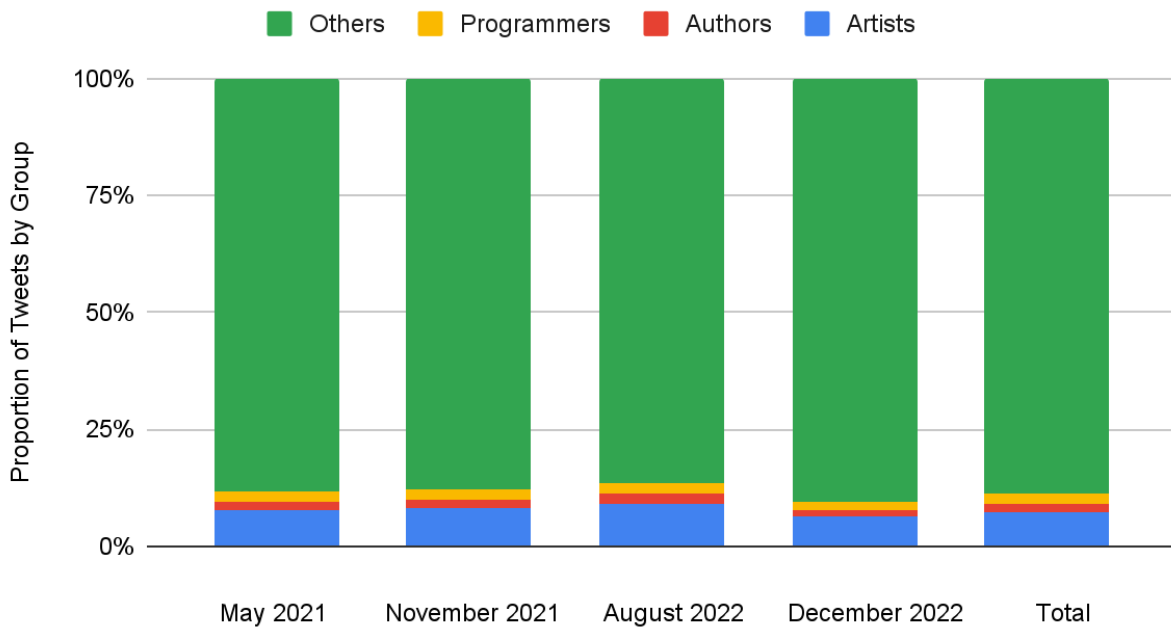
Tweet Social Group Composition

Figure 2. Social Composition of Tweets Over Time

Figure 3 shows a more detailed view of the social composition of all analyzed tweets. Interestingly, artists accounted for about 7% of all tweets, while both authors and programmers accounted for only about 2%. This composition difference may be due to the fact that artists' work – as it is visual – may be more easily spread on social media. Thus, social media may be a more beneficial form of exposure and income to artists than it is to the other key social groups.
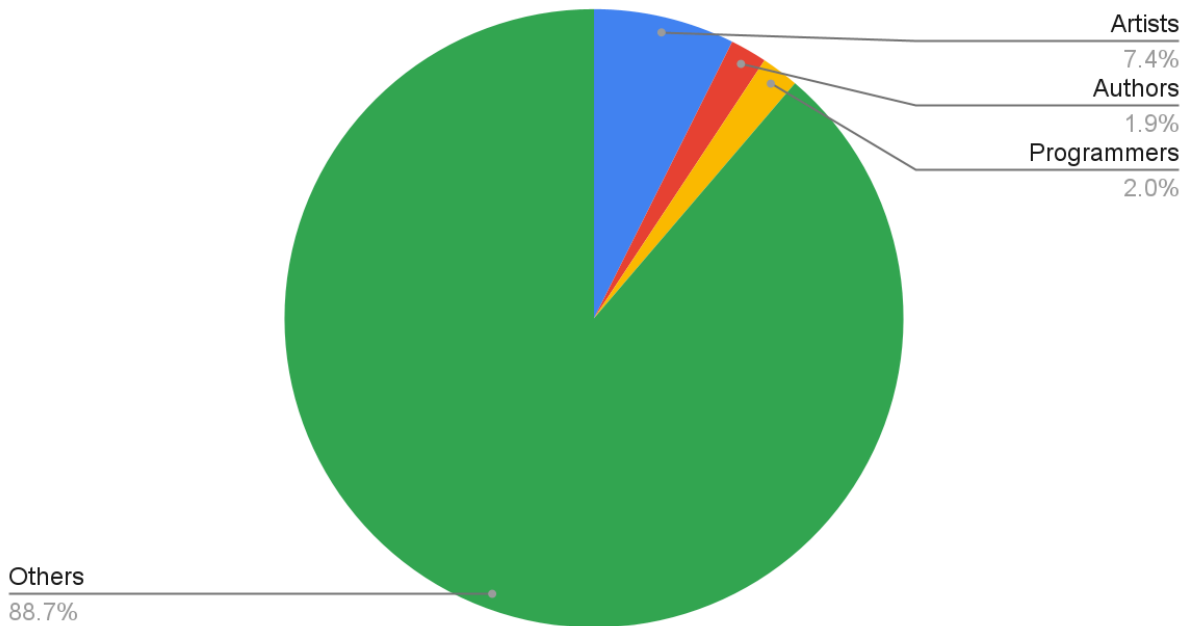
Social Group Composition of Tweets

Figure 3. Social Composition of all Tweets

To determine how each of these social groups has expressed changing priorities about copyright over time, we must look in depth at which themes the groups discuss. I elected to study twelve keywords relating to copyright and artificial intelligence, most derived from Stonybrook University's copyright research guide glossary (Pilato, 2021). See Appendix C for tables with the raw keyword counts for all social groups in each of the four eras. To establish a baseline, I first examined the proportion of all tweets that involved these keywords, which is shown in Figure 4.
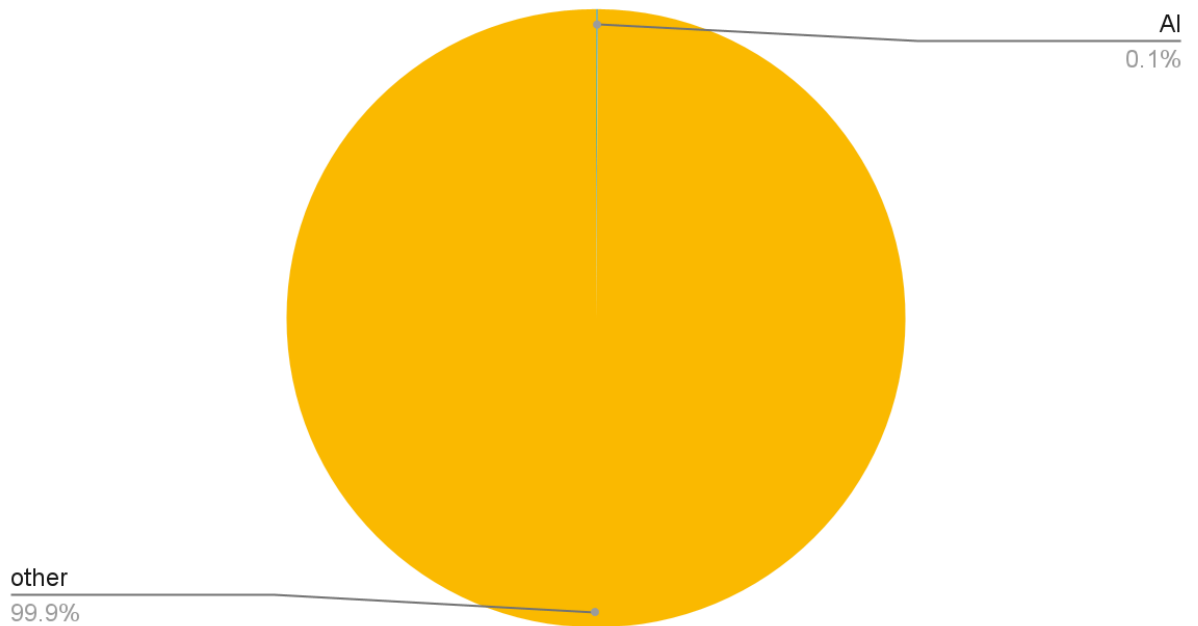
Theme Composition of All Tweets

Figure 4. Theme Composition of All Tweets

Most tweets do not relate to copyright or artificial intelligence. In fact, 11 of the 12 keywords constituted less than a hundredth of a percent of the total tweet composition. The last keyword, "AI," occurred in roughly a tenth of a percent of all analyzed tweets. Figure 4 indicates that Twitter users tend not to discuss – and therefore, may not prioritize addressing – copyright. Meanwhile, Figure 5 shows the relative occurrence of the keywords in the set of all tweets that contained at least one keyword. As seen in Figure 5, the keyword "AI" occurred in over half the tweets about AI and copyright, while "ownership" and "artificial intelligence" were the next most frequent terms.

Theme Composition of Relevant Tweets

patent
8.4%

ownership
12.1%

AI
52.4%

copyright
7.0%

intellectual property
1.3%

IP
5.8%

public domain
1.1%

open access
0.8%
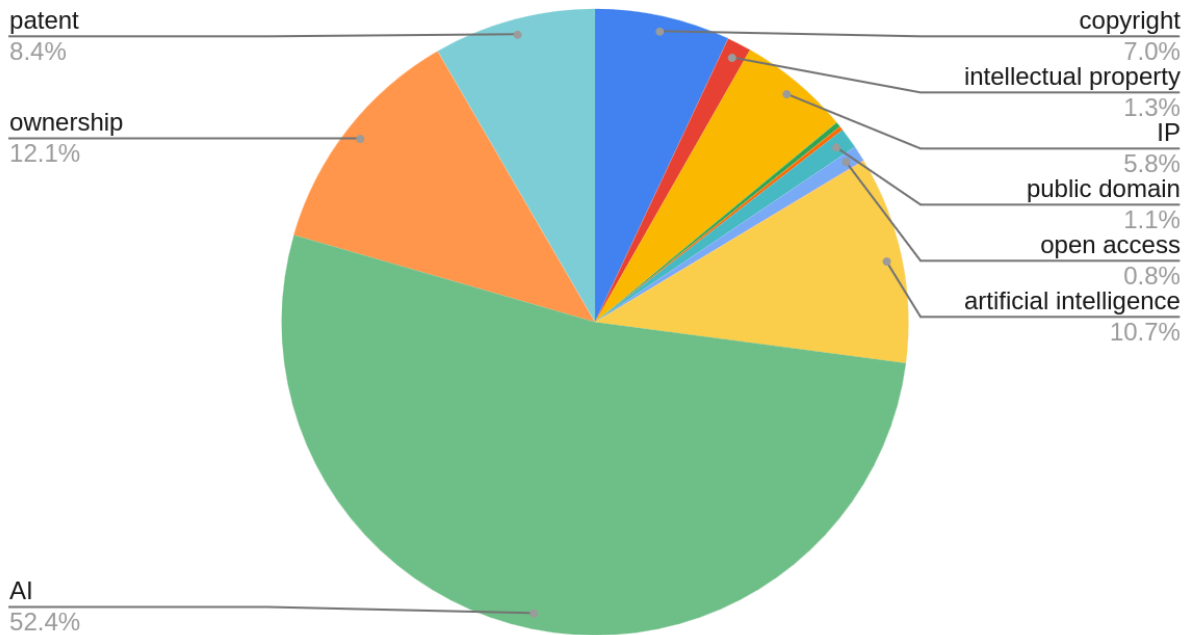
artificial intelligence
10.7%

Figure 5. Theme Composition of Tweets Related to AI and Copyright

People tend to discuss topics that they are interested in or worried about. Thus, the frequency with which they tweet about a topic indicates that they hold some urgency or have a priority related to said topic. For example, tweets about artificial intelligence occurred roughly twice as much as tweets about copyright and intellectual property, indicating that Twitter users are more interested in AI than copyright, and that users discuss topics related to AI without bringing up copyright. To examine how the three key social groups' priorities have changed over time – likely due to the release of new generative AI tools and subsequent online discussion – I quantified how the proportion of tweets that each social group made relating to a topic changed over time. These proportions are shown in Figures 6 through 11. Note that the proportion in each

figure relates the proportion of tweets *only in the specified social group*.
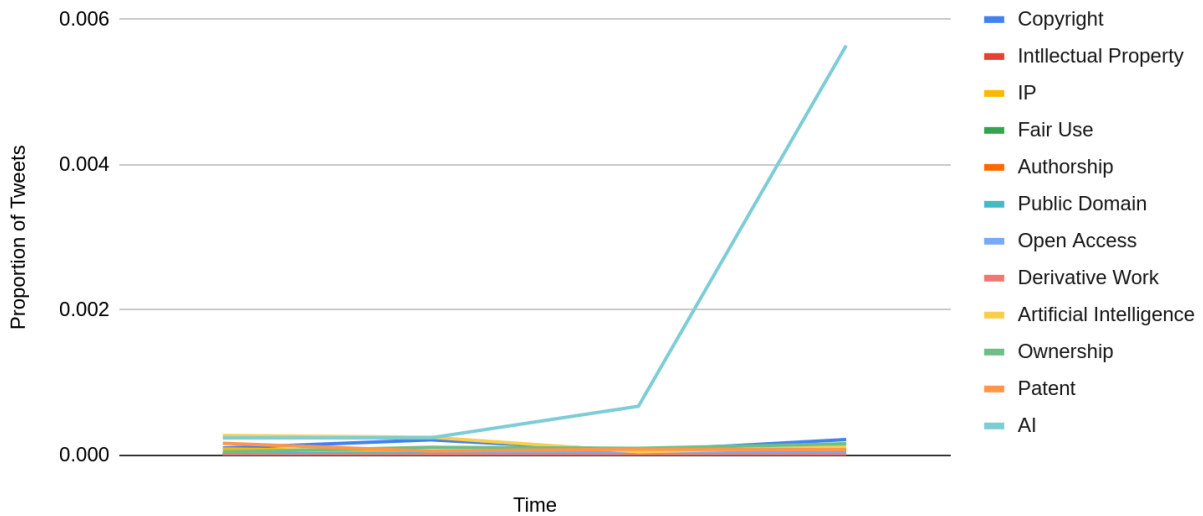
Artist Tweet Theme Frequencies

Figure 6. Frequencies of Themes Discussed by Artists Over Time

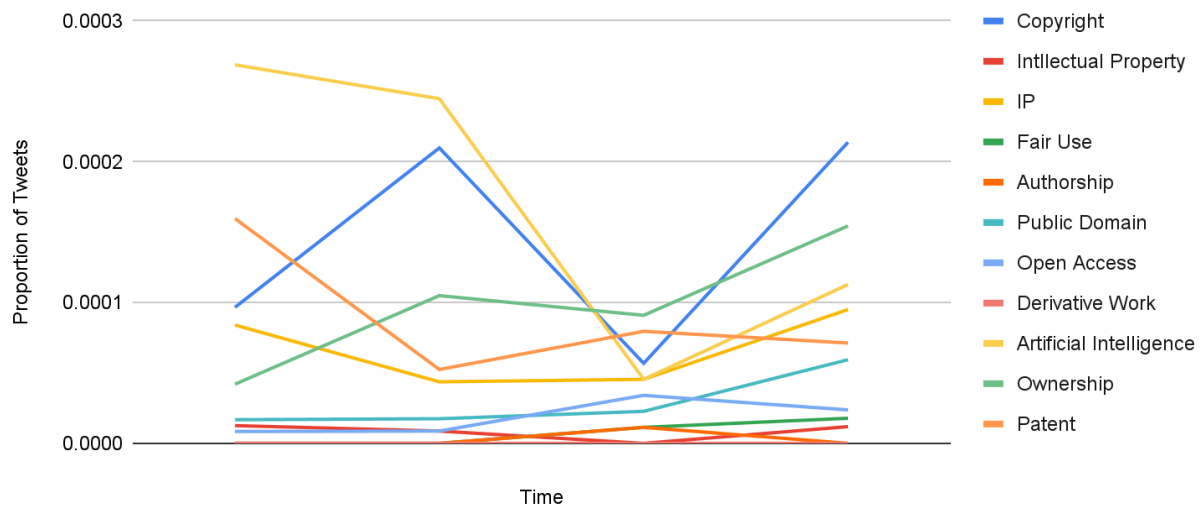## Artist Tweet Theme Frequencies - No Outliers

Figure 7. Frequencies of Themes Discussed by Artists Over Time (Outliers Removed)

Artists began speaking about AI more and more frequently over time, with a subtle increase just after the release of Midjourney, an AI system that can generate artworks. Initially, artists talked about artificial intelligence with high frequency, though this frequency dropped off

before resurging slightly in December 2022. As the number of occurrences of "AI" increased dramatically in the later half of 2022, it is possible that the decrease in occurrences of "artificial intelligence" is due to artists beginning to use "AI" as a shorthand for the topic. Most topics remained relatively stable and low in frequency, though artists tended to talk about copyright and ownership more frequently over time. Interestingly, discussions of artificial intelligence, ownership, and copyright all temporarily decreased in frequency after the release of Midjourney.
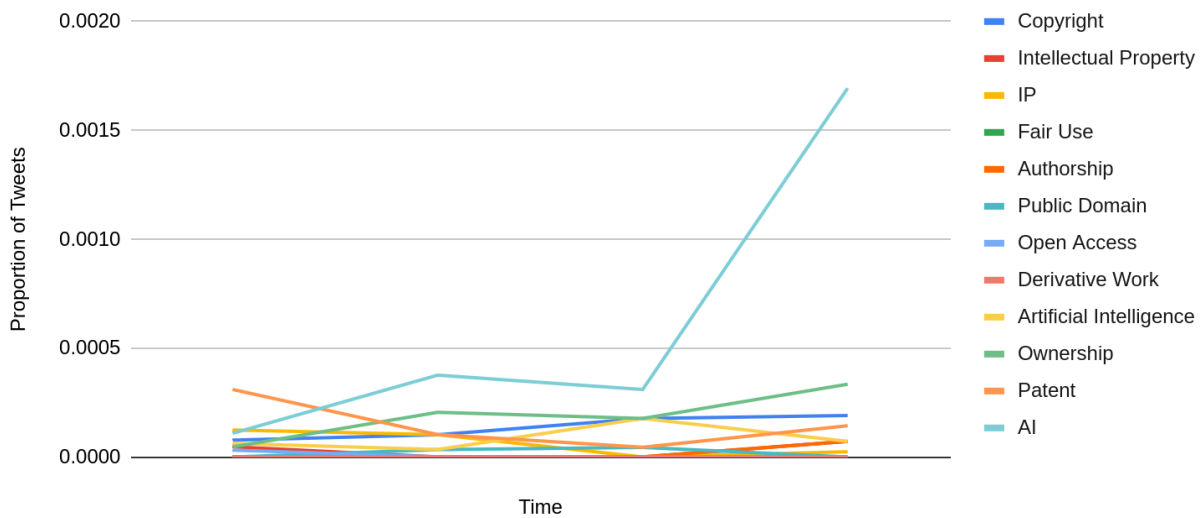

Figure 8. Frequencies of Themes Discussed by Authors Over Time

Author Tweet Theme Frequencies - No Outliers
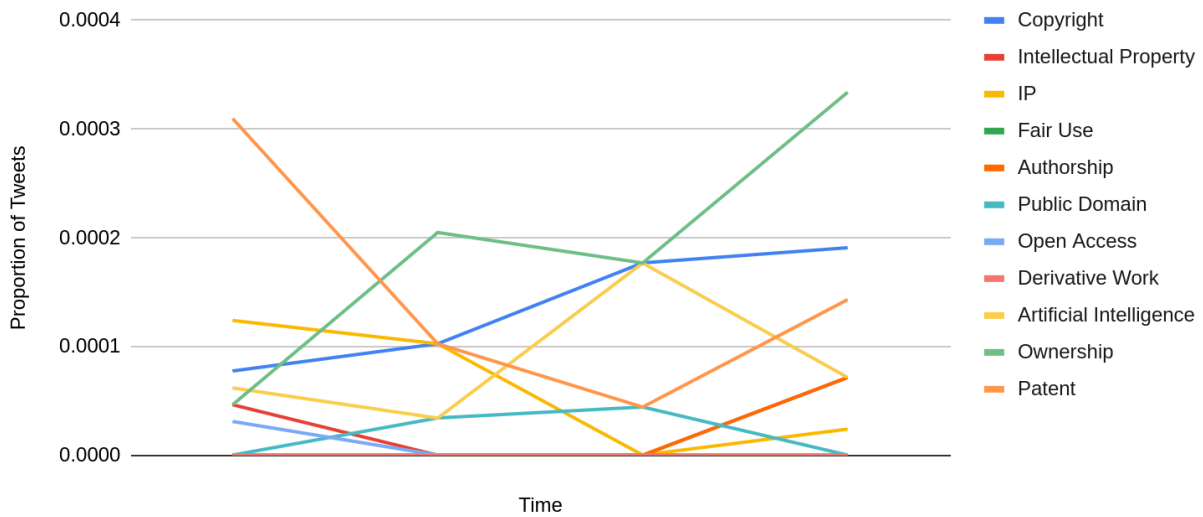
Figure 9. Frequencies of Themes Discussed by Authors Over Time (Outliers Removed)

Authors, too, tended to discuss ownership and copyright more over time and exhibited a large increase in the number of times that they discussed AI. However, they tended to use the phrase "artificial intelligence" less (and with less consistency). Authors initially discussed the topic of authorship relatively frequently, then showed less of an interest in the topic over time. Note that I was unable to study each individual tweet, a limitation of this study that means that discussions of "authorship" could be referencing general authorship, not authorship as it relates to copyright. The release of ChatGPT, which can produce text, was followed by an increase in

the frequencies of "ownership," "authorship," and "copyright."

Programmer Tweet Frequencies
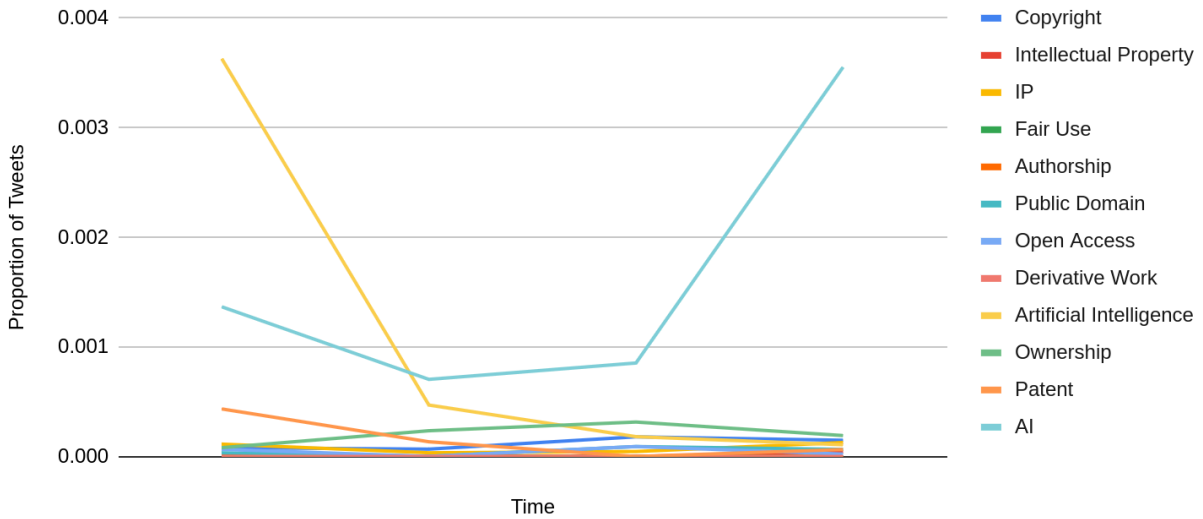
Figure 10. Frequencies of Themes Discussed by Programmers Over Time

Programmer Tweet Frequencies - No Outliers
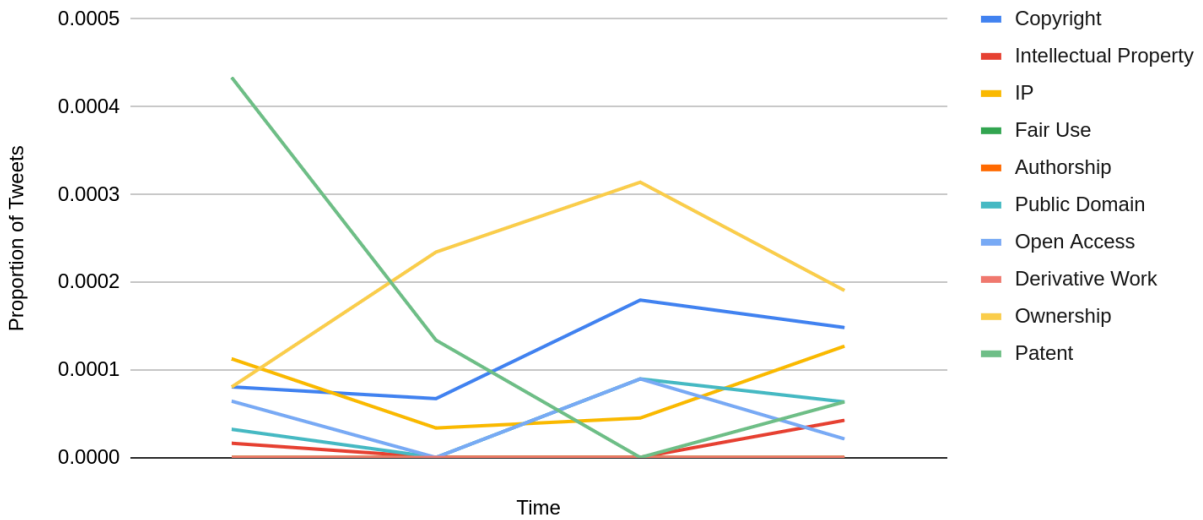
Figure 11. Frequencies of Themes Discussed by Programmers Over Time (Outliers Removed)

The third key social group, programmers, initially used the phrase "artificial intelligence" quite frequently, though this usage dropped off dramatically and appears to be replaced with the phrase "AI," representing a trend to use "AI" as a shorthand for the topic. Programmers initially

discussed patents with some frequency, but their interest sharply declined. Looking at the other topics more closely, we again see a slight increase in frequency of tweets regarding copyright and ownership. The release of Github Copilot, occurring in the part of the graph between the first two points, seems to only have increased the frequency of discussions around "ownership."

After examining each social group independently, it is clear that the groups have different priorities – not a surprise according to SCOT, considering that they experience different problems. However, all three social groups tended to discuss copyright and ownership more over time. The increase in frequency of the two topics was greatest for artists and smallest for authors. The greater change in artist priorities may be due to the fact that more artists use Twitter, which could mean that more artists engage in discussion with each other, hence amplifying issues and priorities within the community.

Analysis of U.S. Court Cases Regarding Copyright and AI

Using a search in two legal databases, Justia and CourtListener, I extracted 156 results, which I then filtered in multiple ways. First, I only considered court cases, not legal codes or regulations, because a court case inherently represents conflict between two parties, while the other types of document do not. Considering conflict among social groups is an essential part of the fourth and final step of SCOT, called wider context, during which a researcher must analyze how different social groups interact with each other about a technological artifact. Thus, only court cases are relevant to my research question. After limiting my results to court cases, I further filtered the results to only include cases that featured both the keyword "copyright" and the key phrase "artificial intelligence." At this point, I had 33 cases, each of which I manually reviewed in order to verify its relevance and to determine the social groups involved.

"Relevant" in this instance refers to any court case which involves a dispute about copyright or intellectual property that centers around an AI system (whether concerned with the inputs to, outputs of, or concepts/source code of the system). Fourteen of the 33 cases were not relevant and were thus discarded. In several cases, the documentation of the court case appears to have matched the term "artificial intelligence" because the judge or legal teams noted that artificial intelligence might have an influence in the courtroom. Although those cases were discarded because they did not concern AI systems, the phenomenon indicates that a social group not discussed in this paper – people who work in and uphold the judicial system in various capacities – have also had an attitude change due to AI. This concept is worth investigating in future work.

Returning to the analysis of relevant cases, three were repeat cases that represented a version of the same suit heard at a lower level court before the case got appealed. One might expect that hearing evidence for both sides of a case and seeing the legal precedence resulting from the case would be a fair way to satisfy conflicting social groups. The fact that such a phenomenon has *not* happened in several recent cases – in other words, that social groups were not satisfied with the ruling of their cases and thus appealed them – indicates that a consensus has yet to be reached among social groups about how to handle copyright in the age of generative AI. Thus, generative AI has not undergone "stabilization" (as defined by SCOT) on a societal level.

Of the sixteen relevant court cases, four occurred before the relevant time period of 2021 to the present. Nonetheless, all four of these cases occurred in the two preceding years, 2019 and 2020, indicating that copyright concerns with AI are relatively new. This novelty is interesting because AI that resembles the AI we produce today – which includes concepts such as the neural

19

network – has been around since at least the early 2000s, and AI as a concept has existed in various forms since the mid-twentieth century (Copeland, n.d.). Of the eleven relevant cases from 2021 onward, seven involved generative AI, while two involved facial recognition software, one involved object detection software, and one involved general AI principles. As the majority of relevant cases specifically involve generative AI, generative AI appears to be a driving factor in making people contest copyright and their intellectual property. Figure 12 shows the distribution of court cases through time, with cases specifically related to generative AI in orange. As seen in the figure, there is a greater concentration of court cases in the time period after the release dates of the three generative AI tools, and all generative AI-related cases appear after those dates.

Court Case Timeline
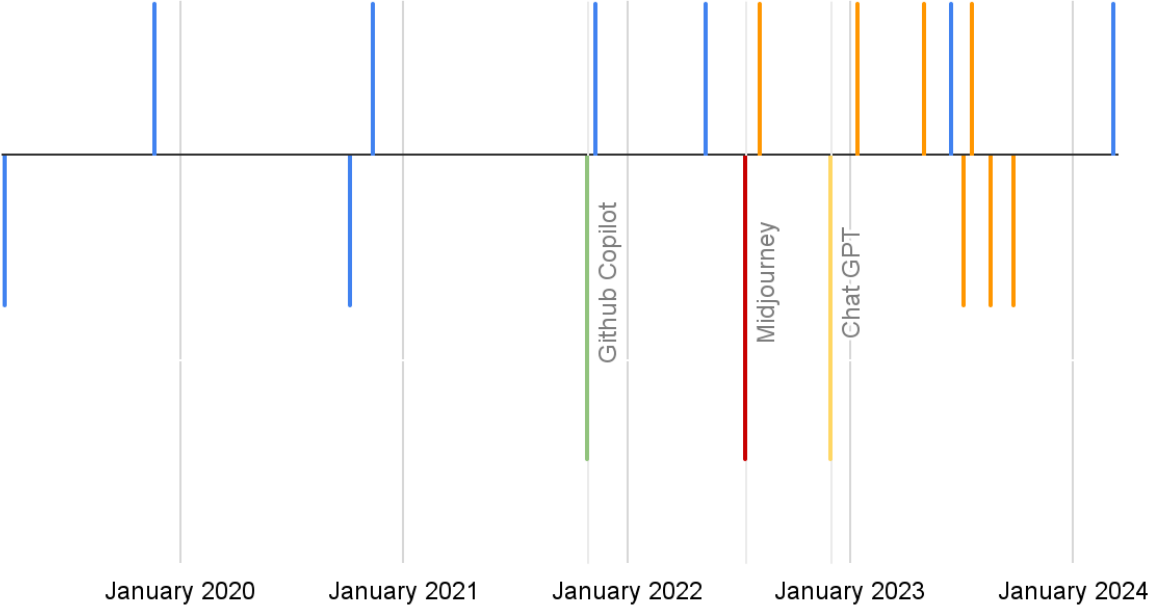
Figure 12. Timeline of Court Cases Regarding AI and Copyright

20

Based on the timeline of court cases, it is clear that generative AI has sparked conflict, but which social groups are involved in said conflicts? To answer this question, I examined the social groups involved in each of the 11 relevant cases. Figure 13 shows how many times a specific social group was involved in a court case, stratified by when the group appeared as a plaintiff and as a defendant. See Appendix D for a table of each case and the social groups of the plaintiff and defendant involved. Note that this paper distinguishes individuals from corporations – although corporations in an industry may share some values with the individuals in that field, they operate differently and have an additional set of problems to contend with, so I treated them as separate social groups in this analysis. In general, government entities and AI companies are the most likely to be involved in AI-related copyright cases, particularly as defendants – usually, plaintiffs sue AI companies for their *usage* of copyrighted data and governments for their *enforcement* of copyright policies.

Social Groups of Parties Involved in AI Copyright Cases
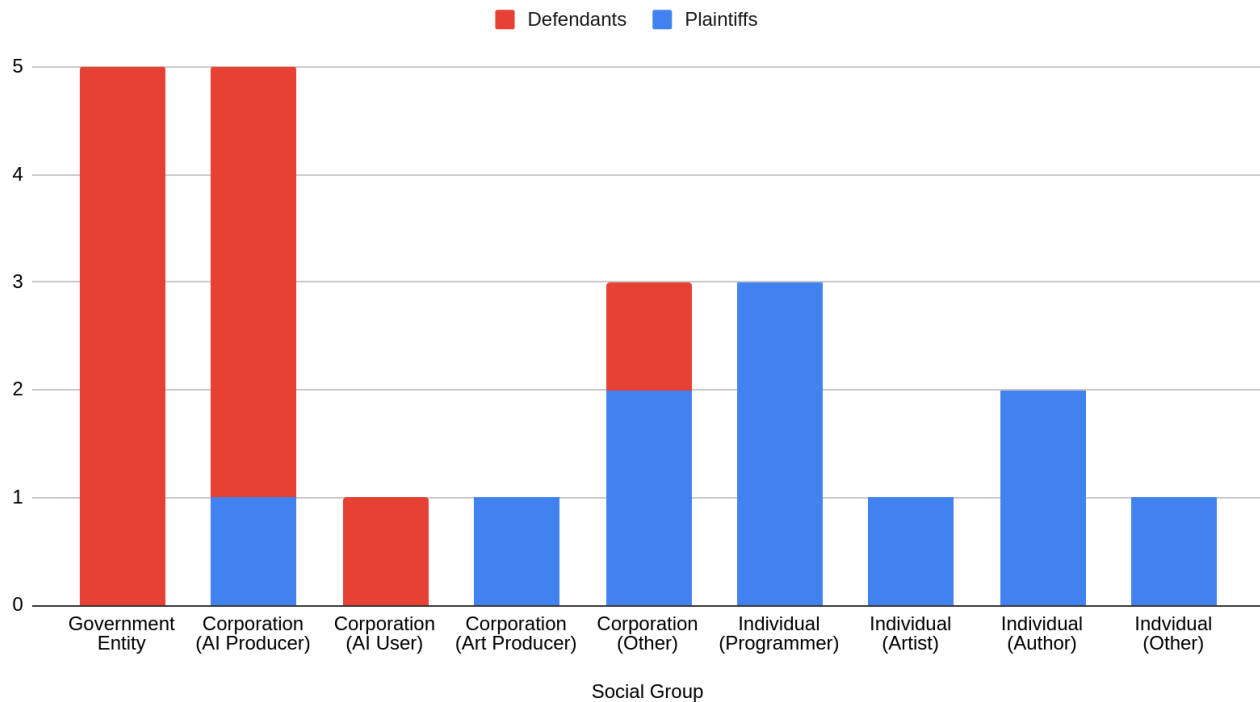
Figure 13. Social Groups of Parties Involved in AI Copyright Cases

Figure 13 indicates that plaintiffs are mostly individuals, though some corporations have instigated cases. Plaintiffs represent all three social groups of interest (artists, authors, and programmers). Meanwhile, defendants are exclusively corporations and government entities, which range from county governments to the U.S. copyright office to the entire United States.

Looking at the involved social groups on a case by case basis reveals even more trends. Individuals show a slight preference for suing governments over corporations, with three individuals suing corporations and four suing governments. On the other hand, plaintiffs that are corporations exclusively sue other corporations. All individuals and two of the three companies that sued corporations did so against companies that either created AI or employed it. Individuals in the fields of art and writing sued because their content was used to train AI systems, while

22

individuals in the field of computer science/programming sued either to get credit for their development of AI or to get ownership of works made by AI systems that they had created. Interestingly, no cases have yet emerged that involve a programmer suing a generative model for producing code influenced by their work, which is the case in most artist and author-instigated suits. Programmers seem to have multiple reasons to sue for copyright – not only would they want to prevent a generative AI from plagiarizing their work, but they might want to prevent other software creators from plagiarizing the actual code of the AI system. These reasons may account for why individual programmers are the most common type of plaintiff, though the small sample size of this data is a limitation. For example, a single individual, Stephen Thaler, accounts for two of the three occurrences of individual programmers, as he was involved in two cases.

Limitations and Future Work

Both methods of answering my research question pose some limitations. First, due to the large number of Twitter posts that I analyzed, I was unable to analyze each tweet individually, which resulted in some ambiguity about which definition of a key word appeared in each tweet. Additionally, as I was unable to access the Twitter API due to its high cost, I could not narrow down the tweets that I processed by topic or keyword before downloading them. Due to memory and computing restraints, I could not download more than four months' worth of Twitter data, which limited my ability to observe fine-grained trends over time. In the future, trends could be observed over more months and even on a daily or hourly basis within months.

In my analysis of court cases, I depended on manual filtering of the court cases, which could have produced some human error. Additionally, the two databases of court information that I used, Justia and CourtListener, likely do not provide complete records of all courts in the

United States. Future work could employ more robust data sources, such as the PACER system, that were unavailable to me in this project due to costs.

Conclusion: Importance of Acknowledging Social Group Conflicts Around Generative AI

The results of this paper indicate that people from all three social groups have more interest in – or perhaps greater awareness of – certain copyright topics now than before the release of several popular AI tools. Additionally, individuals sue for copyright infringement related to AI more than corporations do, indicating that they are more discontented with current copyright laws than corporations are. All but one of the recent court cases related to copyright and AI involved a plaintiff and defendant of two different social groups, showing that there is conflict among social groups and that a wider consensus around the proper use of generative AI has not yet been reached. Insights into the topics that worry social groups the most – as well as the outcomes that social groups would like to achieve in court – can help policymakers modify copyright law and regulate the use of AI in a fair, unbiased way that takes into account perspectives from all groups. Eventually, policymakers and social groups must resolve all conflicts about how generative AI tools may be trained, used, and profited from, but this process may be long and fraught. Identifying the *priorities* of social groups can help policymakers address the most pressing concerns first, leading to short term compromise before it is possible to create long-term solutions.

References

AI Industry Analysis: 50 Most Visited AI Tools and Their 24B+ Traffic Behavior - WriterBuddy.

(n.d.). Retrieved February 13, 2024, from https://writerbuddy.ai/blog/ai-industry-analysis

Artificial intelligence—Nouvelle AI, Machine Learning, Robotics | Britannica. (n.d.). Retrieved

March 23, 2024, from

https://www.britannica.com/technology/artificial-intelligence/Nouvelle-AI

Artificial Intelligence & the Future of Work Survey. (n.d.). Retrieved February 13, 2024, from

https://info.jff.org/artificial-intelligence-the-future-of-work-survey

Bogart, S. (n.d.). *SankeyMATIC: Build a Sankey Diagram*. SankeyMATIC.Com. Retrieved

March 24, 2024, from https://sankeymatic.com/build/

Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). *A large annotated corpus for*

learning natural language inference (arXiv:1508.05326). arXiv.

https://doi.org/10.48550/arXiv.1508.05326

ChatGPT. (n.d.). Retrieved February 13, 2024, from https://openai.com/chatgpt

Copeland, B. J. (n.d.). *Artificial intelligence—Nouvelle AI, Machine Learning, Robotics |*

Britannica. Retrieved March 23, 2024, from

https://www.britannica.com/technology/artificial-intelligence/Nouvelle-AI

Free Law Project. (2024). *Home | Free Law Project | Making the legal ecosystem more equitable*

and competitive. Free Law Project. https://free.law/

Generative Artificial Intelligence: A Social Construction of Technology Perspective. (2022).

[University of Virginia]. https://doi.org/10.18130/BHRK-8H70

GitHub Copilot · Your AI pair programmer. (n.d.). GitHub. Retrieved February 13, 2024, from

https://github.com/features/copilot/

Hayes, C. M. (2023). *Generative Artificial Intelligence and Copyright: Both Sides of the Black Box* (ssrn:4517799). SSRN. https://doi.org/10.2139/ssrn.4517799

Ho, K. K. (2024, January 2). Database of 16,000 Artists Used to Train Midjourney AI, Including 6-Year-Old Child, Garners Criticism. *ARTnews.Com*. https://www.artnews.com/art-news/news/midjourney-ai-artists-database-1234691955/

Justia Law. (2024, March 27). https://law.justia.com/

Kocetkov, D., Li, R., Allal, L. B., Li, J., Mou, C., Ferrandis, C. M., Jernite, Y., Mitchell, M., Hughes, S., Wolf, T., Bahdanau, D., von Werra, L., & de Vries, H. (2022). *The Stack: 3 TB of permissively licensed source code* (arXiv:2211.15533). arXiv. https://doi.org/10.48550/arXiv.2211.15533

Lai, G., Xie, Q., Liu, H., Yang, Y., & Hovy, E. (2017). *RACE: Large-scale ReAding Comprehension Dataset From Examinations* (arXiv:1704.04683). arXiv. https://doi.org/10.48550/arXiv.1704.04683

Leaked: The names of more than 16,000 non-consenting artists allegedly used to train Midjourney's AI. (2024, January 4). The Art Newspaper - International Art News and Events. https://www.theartnewspaper.com/2024/01/04/leaked-names-of-16000-artists-used-to-train-midjourney-ai

Malinsky, G. (2023, August 28). *AI could 'turn good jobs into bad jobs'—3 labor historians on what the future of work might hold*. CNBC. https://www.cnbc.com/2023/08/28/what-labor-historians-think-ai-could-do-to-some-jobs.html

Marr, B. (2023, May 19). *A Short History Of ChatGPT: How We Got To Where We Are Today*.

Forbes.

https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we

-got-to-where-we-are-today/

Midjourney Feed. (n.d.). Midjourney. Retrieved February 13, 2024, from

https://www.midjourney.com/showcase

Midjourney [@midjourney]. (2022, July 13). *We're officially moving to open-beta! Join now at*

https://discord.gg/midjourney. **Please read our directions carefully** or check out our

detailed how-to guides here: Https://midjourney.gitbook.io/docs. Most importantly, have

fun! [Tweet]. Twitter. https://twitter.com/midjourney/status/1547108864788553729

Miyazaki, K., Murayama, T., Uchiba, T., An, J., & Kwak, H. (2024). Public perception of

generative AI on Twitter: An empirical study based on occupation and usage. *EPJ Data*

Science, *13*(1), Article 1. https://doi.org/10.1140/epjds/s13688-023-00445-y

Pilato, V. (2021, November 3). *Research & Subject Guides: Copyright Guide: Glossary*.

https://guides.library.stonybrook.edu/copyright/glossary

Pinch, T., & Bijker, W. (1987). The Social Construction of Facts and Artifacts. In B. Wiebe, T.

Hughes, & T. Pinch (Eds.), *The Social Construction of Technological Systems* (pp.

17–50). MIT Press..

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T.,

Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K.,

Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). *LAION-5B: An open large-scale*

dataset for training next generation image-text models (arXiv:2210.08402). arXiv.

https://doi.org/10.48550/arXiv.2210.08402

Scott, J. (2012, December 6). *Archive Team: The Twitter Stream Grab: Free Web: Free*

Download, Borrow and Streaming: Internet Archive.

https://archive.org/details/twitterstream

Solomon, T. (2023, October 31). Judge Dismisses Several Copyright Allegations Against AI

Companies in Artist Class Action. *ARTnews.Com*.

https://www.artnews.com/art-news/news/judge-dismisses-several-copyright-allegations-a

gainst-ai-companies-in-artist-class-action-1234685299/

UNESCO. (2023, July 20). *Navigating intellectual property rights in the era of generative AI:*

The crucial role of educating judicial actors | UNESCO.

https://www.unesco.org/en/articles/navigating-intellectual-property-rights-era-generative-

ai-crucial-role-educating-judicial-actors

Welsh, M. (2022). The End of Programming. *Communications of the ACM*, *66*(1), 34–35.

https://doi.org/10.1145/3570220

Appendix A

AI Tool Name	Release Date
Github Copilot	October 29th, 2021
Midjourney	July 12th, 2022
ChatGPT	November 30th, 2022

Figure A1. Release Dates of Relevant Generative AI Tools

Analysis Era	Month Selected
Pre-release of any tools	May 2021
Post-release of Github Copilot	November 2021
Post-release of Midjourney	August 2022
Post-release of ChatGPT	December 2022

Figure A2. Eras Selected for Tweet Analysis

Appendix B

Social Group	Keywords Used for Identifying Group Members
Artists	"artist," "illustrator," "graphic designer," "art," "designer," "paint," "draw"
Authors	"author," "write," "novelist," "playwright," "poet," "dramatist"
Programmers	"programmer," "developer," "dev," "software," "swe," "code," "computer scien," "techie," "hacker"

Figure B1. Keywords Used for Identifying Social Groups

Note that some keywords are root words in order to capture any profile that uses a derivative form of the word. For example, the keyword "paint" captures profiles that read "I paint" or "Oil painter." All keyword searches were case-insensitive.

Social Group	May 21	Nov 21	Aug 22	Dec 22
Artists	238032	114376	87959	168358
Authors	64661	29317	22640	41970
Programmers	62322	29915	22320	47330
Others	2765325	1255824	854396	2453321

Figure C1. Tweet Counts by Social Group

Keyword	Count May 21	Count Nov 21	Count Aug 22	Count Dec 22
copyright	23	24	5	36
intellectual property	3	1	0	2
IP	20	5	4	16
fair use	0	0	1	3
authorship	0	0	1	0
public domain	4	2	2	10
open access	2	1	3	4
derivative work	0	0	0	0
artificial intelligence	64	28	4	19
ownership	10	12	8	26
patent	38	6	7	12
AI	56	27	59	949

Figure C2. Artist Keyword Counts

Keyword	Count May 21	Count Nov 21	Count Aug 22	Count Dec 22
copyright	5	3	4	8
intellectual property	3	0	0	3
IP	8	3	0	1

	0	0	0	0
fair use	0	0	0	0
authorship	0	0	0	3
public domain	0	1	1	0
open access	2	0	0	0
derivative work	0	0	0	0
artificial intelligence	4	1	4	3
ownership	3	6	4	14
patent	20	3	1	6
AI	7	11	7	71

Figure C3. Author Keyword Counts

Keyword	Count May 21	Count Nov 21	Count Aug 22	Count Dec 22
copyright	5	2	4	7
intellectual property	1	0	0	2
IP	7	1	1	6
fair use	0	0	0	0
authorship	0	0	0	0
public domain	2	0	2	3
open access	4	0	2	1
derivative work	0	0	0	0
artificial intelligence	226	14	4	5
ownership	5	7	7	9
patent	27	4	0	3
AI	85	21	19	168

Figure C4. Programmer Keyword Counts

Keyword	Count May 21	Count Nov 21	Count Aug 22	Count Dec 22
copyright	197	74	59	179
intellectual property	63	11	6	19

IP	181	44	49	179
fair use	4	5	2	9
authorship	2	5	0	8
public domain	14	6	5	46
open access	18	10	5	21
derivative work	1	0	0	1
artificial intelligence	428	51	24	95
ownership	283	170	115	426
patent	335	68	57	174
AI	386	178	250	2471

Figure C5. Other Groups Keyword Counts

Key word	May 21	Nov 21	Aug 22	Dec 22
copyright	230	103	72	230
intellectual property	70	12	6	26
IP	216	53	54	202
fair use	4	5	3	12
authorship	2	5	1	11
public domain	20	9	10	59
open access	26	11	10	26
derivative work	1	0	0	1
artificial intelligence	722	94	36	122
AI	534	237	335	3659
ownership	301	195	134	475
patent	420	81	65	195

Figure C6. Keyword Totals

Appendix D

Court Case	Plaintiff Social Group	Defendant Social Group
Jg Technologies, LLC v. United States	Corporation (software)	Government entity
Citizens Insurance Company of America v. Wynndalco Enterprises, LLC	Corporation (other)	Corporation (other)
Thaler v. Perlmutter	Individual (programmer/computer scientist)	Government entity
Moomaw et al. v. Geosnapshot Pty LTD et al.	Individual (Other)	Corporation (using AI)
Thomson Reuters Enterprise Centre GmbH et al. v. ROSS	Corporation (other)	Corporation (producer of AI)
Sarah Andersen, et al. v. Stability AI LTD., et al.	Individual (artist)	Corporation (producer of AI)
Silverman et al. v. OpenAI, Inc. et al.	Individual (author)	Government entity
Paul Tremblay, et al. v. OpenAI, INC., et al.	Individual (programmer/computer scientist)	Government entity
Aljindi v. US	Individual (author)	Corporation (producer of AI)
Thaler v. Vidal	Individual (programmer/computer scientist)	Government entity
Getty Images (US), INC., v. Stability AI, LTD. and Stability AI, INC.	Corporation (producer of art)	Corporation (producer of AI)