Campaign for AI Safety

Comments: US Copyright Office Notice of Inquiry on Copyright and Artificial Intelligence

Comments are made by the Campaign for AI Safety in response to the US Copyright Office's Notice of Inquiry on copyright and artificial intelligence. We trust our comments are of assistance. Please do not hesitate to contact us to discuss this submission.

The Campaign for AI Safety is a not-for-profit association established in Australia with members in Australia and other countries including the United States. We are concerned about the dangers AI poses to people and advocate for a stop on the advancement of certain AI capabilities. We also advocate for regulation that promotes and mandates ethical AI. We are not affiliated with any political group. Please visit campaignforaisafety.org for more information.

---

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail?

Implications for livelihoods of creative professionals

Generative AI arrived largely unanticipated by the public, creative professionals, publishers and relevant stakeholders. The vast majority of creators never imagined that their works would be used for machine learning. Copyright law never considered this novel use and the ease and speed with which content could be generated with AI. This has created a climate of uncertainty for creative professionals in the US and overseas as well as businesses using AI-generated content.

Strong copyright regime can help facilitate safer AI development

There are three reasons why enforcement of copyright matters for AI safety:

1. There is ongoing debate on the issue of "emergent abilities"[1] of large language models, including abstract thinking and dangerous capabilities (such as ability to

[1] Ability is defined as "emergent if it is not present in smaller models but is present in larger models", Jason Wei, et al., "Emergent Abilities of Large Language Models", https://arxiv.org/abs/2206.07682

suggest dangerous chemical compounds). Because they are seen as unpredictable, they can be hard to prevent, detect, limit or moderate.

Alternative explanations of these abilities have been proposed, including training data size[2]. Because the lists of references to training data of the most advanced models have not been released to the public, researchers have a difficulty in resolving the important issue of to what extent, which specific abilities are emergent or are regurgitation of their training data.

Our recommendation #5 below aims to rectify this situation.

2. According to the Campaign's recent poll[3], 66% of US adults agree that "artists and content creators should be paid if their work is used in training artificial intelligence (AI) models", which AI companies are not currently doing at scale.

   At the same time, AI labs often make reference to advancing AI "benefits all of humanity" and "avoid enabling uses of AI or AGI that harm humanity or unduly concentrate power"[4]

   There is an apparent disconnect between the behaviour of AI labs and their stated intentions, which should make one suspicious of labs' real motivations for building very powerful AI systems.

   Enforcing copyright will help reduce concentration of power and resources in the hands of makers of AI systems.

3. Enforcement of copyright in regards to training data would require some time and regrouping, which can help cool down competitive race dynamics in the AI industry, which are detrimental to ensuring robustness and safety of AI[5].

Preservation of competition and fair trading for suppliers in an oligopsonistic market

Only a few businesses have the financial and computational resources to train powerful AI models. The same businesses or their affiliates happen to own dominant social media and search engines, which puts them in positions of extreme market power. They may attempt to include waiver of copyright into their general terms and conditions of service[6]. Intervention is warranted because individuals will not have the resources or capability to fight against such overreach.

---

[2] "Are Emergent Abilities of Large Language Models a Mirage?", Rylan Schaeffer, et al., https://arxiv.org/abs/2304.15004
[3] "USA AI x-risk perception tracker", Campaign for AI Safety, https://www.campaignforaisafety.org/usa-ai-x-risk-perception-tracker/
[4] OpenAI Charter, https://openai.com/charter
[5] "AI Is Not an Arms Race", Katja Grace, Time, https://time.com/6283609/artificial-intelligence-race-existential-threat/
[6] "Google Says It'll Scrape Everything You Post Online for AI", Thomas Germain, Gizmodo, 3 July 2023

Policy recommendations to clarify and reinforce copyright in relation to AI training

1. Third parties who wish to use copyrighted materials in training AI models must obtain **specific consent** from the owners of materials.

2. Specific consent implies that third parties **must not be allowed to coerce copyright holders** to provide such consent. For example:

    a. Consent must not be part of terms and conditions of unrelated services (such as video or gaming distribution platforms). This way social media companies will not be able to deny services or features of their platforms to those who do not consent to such use of copyrighted materials.

    b. Specificity of consent means that it needs to be a separate agreement, ideally a contract with consideration, which copyright holders have an opportunity to take time to consider, review, and negotiate. It must not be a checkbox underneath a registration form on a website.

3. **Consent must be granular**. For example, copyright owners should be able to specify if they allow third parties to have technical means to generate works "in the style of" the author of the materials. This would be similar to granular consent for processing data under UK GDPR[7].

4. Copyright holders must have effective means of **negotiating and receiving compensation** for the use of their copyrighted materials (such as opt-in collective bargaining mechanisms).

5. **Data used in training AI models must be fully referenced**. References to the works used in training must be made publicly available.

6. Training models using "synthetic data" generated by other models trained on copyrighted materials must be considered equivalent to training on those copyrighted materials.

7. Legal liability in case of infringement of copyright must apply both to parties that train models and parties that use those models to generate content.

8. "Fair use" provisions in law must not apply to AI model training.

9. Materials generated completely or substantially by AI models must not be copyrightable.

These clarifications should be legislated, without waiting for courts to provide them. Appropriate legislation, if promptly enacted, will give certainty in relation to these important questions to the AI industry, the Australian business community, and copyright holders, including artists, writers, publishers, and the wider public.

We clarify the terms used in the recommendations above:

---

[7] What is valid consent?, UK Information Commissioner's Office

- **AI models** refer to generative AI models and systems (including constellations of models), including large language models (e.g. OpenA's GPT-3), multimodal models (e.g. Google Deepmind Gemini), large diffusion models (e.g. Stable Diffusion).
- **Training** includes fine-tuning[8].
- **Copyright** includes other substantially similar intellectual property rights.
- **Copyrighted materials** include confidential information and similarly protected content.

Case study 1: Chat-GPT powered by GPT

Chat-GPT is a famous AI-powered chatbot that uses a variety of models in the GPT series (including GPT-3.5 and GPT-4) to generate answers (or "completions") to user inputs ("prompts"). GPT models were trained on various undisclosed sources of information.

It has been alleged that many of those sources of data are copyrighted or privacy-protected[9]. Thanks to the training data, Chat-GPT is able to reproduce (albeit with some error) portions of copyrighted works and summarise them[10]. Authors face diminished demand for their works when imperfect reproductions and regurgitations are accessible from OpenAI and Microsoft.

Because of the secrecy around the training data used by OpenAI[11], copyright owners might not realise that their rights could be violated.

AI companies like OpenAI stand to unfairly gain from the use of works, whose creators never considered the possibility that they would be used for machine learning, do not consent to such use, are not compensated, and do not enjoy attribution of moral rights.

Case study 2: Google Deepmind's Gemini

Google Deepmind is working on a system called Gemini, which they claim will have a number of advanced capabilities[12]. It has been reported[13] that Gemini may be trained using materials uploaded onto Google-owned YouTube platform.

While it is not clear what data is being used for training, if Google indeed is using videos hosted on the YouTube platform, that would be a violation of users' trust and needs to be clarified in legislation as illegal because no specific consent was sought from the users who

---

[8] Inserting materials into the context window (prompt) of AI models may serve as an alternative to training and therefore must be treated similarly. Example long context window research: "[2307.02486] LongNet: Scaling Transformers to 1,000,000,000 Tokens"
[9] "ChatGPT Language Model Litigation", Joseph Saviery Law Firm, https://www.saverilawfirm.com/chatgpt-language-model-litigation.
[10] "OpenAI's ChatGPT may face a copyright quagmire after 'memorizing' these books", The Register, https://www.theregister.com/2023/05/03/openai_chatgpt_copyright/.
[11] Only cryptic names for datasets are given in OpenAI's GPT-3 paper, such as "Books1", "Books2", etc. "Language Models are Few-Shot Learners", https://arxiv.org/abs/2005.14165.
[12] "Google DeepMind's CEO Says Its Next Algorithm Will Eclipse ChatGPT", Wired, https://www.wired.com/story/google-deepmind-demis-hassabis-chatgpt/
[13] "Why YouTube Could Give Google an Edge in AI", The Information, https://www.theinformation.com/articles/why-youtube-could-give-google-an-edge-in-ai

created and uploaded their videos, and no mechanism of compensation was agreed to or established for such use.