

**US Copyright Office inquiry on copyright and artificial intelligence (AI)  
AI Initiative**

**Respondent  
Tavarus Blackmon Art LLP**

Tavarus (Blackmonster❤️) Blackmon

Tavarus Blackmon Art LLP

<https://tavarusblackmonart.com>

<https://www.linkedin.com/in/tavarus-blackmon/>

(Government ID Verified-CLEAR)

Part Time Faculty, CSU Sacramento

Chair, DEI Committee, Root Division, San Francisco, CA

Charter Member and Advisory Board Member,

UC Davis Art and Art History Department

Blackmonster Music

Creative Foundations

Art Music Lit Space

Speaking on behalf of Tavarus Blackmon Art LLP representing Partner's Blackmon and Elizabeth Cord, I (we) wish to answer the questions posed on AI, US Copyright and the copyright-ability of works created utilizing AI and Generative AI tools, models and algorithms.

Blackmon received their MFA in Studio Art at UC Davis, an MA in Studio Art at CSU Sacramento and further degrees in Digital Film and Video Production and Social Sciences. They're Thesis, *The Politics of the Cartoon and Contemporary Art*, is property of the Office of Graduate Studies, CSU Sacramento, published online at *ScholarWorks* <https://scholars.csus.edu/esploro/outputs/graduate/The-politics-of-the-cartoon-and/99257831093101671>, and is further developed into a Generative AI Text-to-Speech Sound Art Project included on the website *AwedioCodex* (2015).

Blackmon instructs in the CSU System and introduces AI Generative tools within the Foundational Drawing and Painting instruction. Blackmon is US Copyright holder of *Literary Hip-Hop: Capitalist Americans*, a registered US Copyright, 2013, establishing the use and rights of Text-to-Speech, AI Generative, works of Hip-Hop Music created with text-based Language Models. Blackmon is also holder of the US Copyright *Me & Karen*, an Album produced with the use and collaboration with AI Generative Text-to-Speech technology. *Blackmonster Family Reliquary*, US Copyright Registration, 2022, is a collection of published photographs created utilizing AI Generative Text-to-Image technology and hand illustrated and hand animated using human collaborative contributions. The completed work which is designed including the Art Direction of the project, by the Artist, utilizes Generative AI Text-to-Image based, Prompt-Engineering and Organizational Trademark and Copyright specific "text," and standard characters included within the Prompt-Engineering process.

Tavarus Blackmon Art LLP owns and operates Subsidiary Blackmonster Music publishing and the Sole Proprietorship's Creative Foundations, an Art Services Company and Art Music Lit Space, both a virtual gallery founded during the Global Pandemic Covid-19 and representing Artist's in the UC Davis Health Art Collection, Artist Pool.

Tavarus Blackmon, representing Tavarus Blackmon Art LLP, has served as a Creative Consultant on the City of Sacramento Music Census, conducted by Sound Music Cities, and the National Endowment for the Arts (NEA) and the City of Sacramento's, Guaranteed Basic Income for Artists (GBI) grant; Panelist for the Sacramento Office of Arts and Culture and the California Arts Counsel, Creative Youth Grant; nominator for the SECA Award (2022/2024), administered by the San Francisco Museum of Modern Art, and Organizational Vender with the City of Sacramento.

Blackmon has been utilizing AI Generative Text-to-Speech technology since 2012, specifically with the application *Natural Reader*. Notably, Blackmon combines Text-to-Speech technology and philosophical documents published in the Stanford Encyclopedia of Philosophy, other works and AI collaborations include adapting *The Encyclopedia of Universal Knowledge* and the *Ludas Patronymicus*. Blackmon has been utilizing First Generation AI Generative Text-to-Image technology and algorithms since 2021 with *Night Cafe Studio* including Image and Video. Blackmon has been utilizing AI-generated applications and databases such as [<https://www.behindthename.com/random/>] Random Name Generation, which experiment with the concept, *identity*, the profile picture *PfP*, and the US Copyright (*denied*) collection of photographs, *Messed Up Kids*, 2022. Blackmon has beta tested Open AI: DALL-E II, Chat-GPT, CAI, Content Credentials beta, Google, Generative AI Search beta and the Neural Net AI technology facilitating the transcription process, for example, the application *Veed*. Blackmon is also an individual member of the *Leonardo*, a not-for-profit 501(c)3 enterprising think tank.

## General Questions

1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

As a Business Owner, Fine Artist, Virtual Gallery Founder, Curator, Technologist and Educator sharing AI Generative tools with their students within the CSU system Art Department, I believe AI Generated works *should* be copyrightable.

[It is my professional opinion that regardless whether a “human,” contributes in parts representing adequate “human,” intervention of the completion of “work,” generated with AI tools, the “work,” generated by AI tools, models and algorithms is *beautiful*. Further, I propose that I have altered every AI output generated by my Organization to achieve what can be considered “authorship.” However, as a Fine Artist with two Master’s Degrees in Studio Art and the Studio Practice, AI Generative “Art,” or outputs posses qualities “expressive,” of what is otherwise considered *beautiful*, thereby Artistic in formal qualities. It is my opinion that if a “non-human,” or generative or AI learning model develops, generates or creates “work,” thereby representing an “image,” or “picture,” or “scene,” or “expression,” than considerations be made to develop, foster and facilitate working “relationships,” with AI models.]

Further, I acknowledge the potential holistic risks of the technology to culture and considering withstanding moral grounds and philosophical dilemma, but as an operator, owner and facilitator of AI Generative applications, datasets and Artwork, my organization works in good faith with the technology made available that also provides Diversity Equity and Inclusion DEI access to diverse systems made available through the utility of Artificial Intelligence.

As just stated, AI does pose risks. What risks and to what extent I am not inclined to yet fathom. However, I believe more in the application of AI for its DEI implications and not just Artistic or “profitable,” endeavors. For example, My introduction to AI models was the *Natural Reader* application. The App uses datasets and Language Models to develop natural sounding computer “reader’s.” Such applications including Kurzweil Technology, future iterations of Text-to-Speech Artificial Intelligence “Reader,” *models*, Text-to-Image Generative Applications

including image and video generations, Generative Language Models and Generative AI Search, create access for individuals including those with disabilities and impairments.

Further, Developmentally Disabled students and clients of School Districts, Academic Institutions and Clinics, including clinic-based Applied Behavioral Analysis, benefit from the underlying AI technology that makes natural reader's possible, further, the underlying technology and Language Models also provide access including for those with speech delays and others whom are non-verbal.

As a disabled Father and Artist and Father of children with Developmental Disabilities and verbal impairments, I utilize AI technology in the normal operation of my Art Services business, which also includes the creation of AI generated Artwork and Digital Assets tokenized on the Ethereum Layer 1 Blockchain Network, and I also provide my family and students accessible technology, adaptive technology and alternative forms of instructional delivery including AI Text-to-Speech Language Models that facilitate the learning process and ones ability to navigate and communicate in the community.

Finally, it is my belief that the underlying AI technology is affecting creators, copyright owners, technology developers, researchers, and the public.

In the case of Creators:

I have been instructing my students in the CSU System, Art Department, to familiarize themselves with AI Generative Text-to-Image applications. The Assignment is an Extra-Credit Assignment whereby I introduce the student's to the Application that I currently use: *Night Cafe Studio*. The student is instructed to visit the site, claim the free platform "credit," generate an image using a text-prompt, and then draw the output image with a traditional medium, such as graphite or charcoal, ink or watercolor, from "observation."

The course(s), Beginning Drawing and Intermediate Drawing, introduce students to a variety of Requirements and Learning Outcomes in the Humanities. My Extra-Credit assignment is designed to not only introduce AI Generative Models to my drawing students, but also methods to engage the technology and production of Traditional Art. It also allows students to see the potential of "scaleability," whereby my students do not have the business or financial means to "scale," a production of work. AI Generative models provide access to this necessary business model.

Further, the Extra Credit Assignment is designed to provide accessibility for students in my Beginning Drawing course, which is also an Elective in the C-1 Area of the Humanities. As many Freshmen and Transfer students take my course, including Art and Non Art Majors, this AI Generative Text-to-Image Extra Credit Assignment provides for an accessible, also technological, means to produce, manipulate, reproduce and scale through "observation," a work of creative output. In this case, a drawing made by the hand developed and generated with AI tools and a technologically-collaborative process, gives some students right and freedom to utilize their own device or CPU in the production of work. As an Artist and Business Owner utilizing technological devices as well as applications, this is a new and necessary component of the Fine Art Studio Lab that instructs Interdisciplinary Arts and the self-custodial management of Faculty and Student data through electronic and digital processes and devices.

In the case of Copyright Owners:

(I) we, Tavarus Blackmon Art LLP, has filed three *approved* US Copyright registrations that utilize AI Generative Technology: *Literary Hip-Hop: Capitalist Americans*, 2013, Sound Recording; *Works Published on the Album Me & Karen*, 2022, Sound Recordings from an Album; *Blackmonster Family Reliquary*, 2022, Published Photographs

In the case of these registrations, the Author, Artist or otherwise Creator of the work, Tavarus Blackmon representing Tavarus Blackmon Art LLP, utilizes AI Text-to-Image and Text-to-Speech Generative models and collaborates in the process by developing the dataset, Language Model, text-prompt, the addition of hand illustrated and hand animated elements within the composite and composition and further AI Text-to-Image Generative assets that are composited elements, included within the final Artwork.

However, it is not only that these US Copyright registrations include AI Generative outputs in the completion and fabrication of the work, but that these registrations include extensive assets external to the development of the AI “work,” and “composite,” that constitutes the larger Copyright.

For example, *Literary Hip-Hop: Capitalist Americans*, is created with AI Generative Text-to-Speech technology, has further been developed into an interactive and responsive element on a custom coded and stereoscopic website: *AwedioCodex*, made publicly available through the streaming platform *Soundcloud*, made publicly available on the desktop publishing application *Issuu*, developed into a Site-Specific Installation and Performance Artwork that premiered at Headlands Center for the Arts, Sausalito, CA, Building 961 (Building 961 has been deaccessioned and returned to the US Government). Finally, the work establishes the Registered US Copyright over AI Text-to-Speech Hip-Hop music created using Text-to-Speech technology, and was published in 2015 on an Interactive, stereoscopic, portfolio website and audio repository: *AwedioCodex*. The Ethereum Layer 1 Blockchain Network launched in the same year. At this time, *AwedioCodex* and the US Copyright Registration *Literary Hip-Hop: Capitalist Americans*, acted as a “poor-man’s,” Smart Contract on the interactive site: the interactive Computer Reader verbally “expresses,” the Copyright claim within the interactive website and recites the generative lyrics. The interactive vocal Text-to-Speech AI Generated “expression,” thereby “releases,” or proves provenance and establishes rightful ownership over Text-to-Speech AI Generative Hip-Hop using AI Language Models, for that every time the button is pressed, the reader recites the Copyright Claim. The claim and recitation is also the business owner’s US Copyright Registration work “deposit.” Therefore, the copyright “deposit,” is a Sound File that “expresses,” the copyright claim through the use of AI Text-to-Speech Generative Engineering and the AI Computer recitation encoded as an audio codec.

Lastly, as by “Generative Engineering,” it is meant that the Author, Artist or otherwise Creator, Blackmon, engineered the Generative Speech through the creation of language. That is to say that in utilizing AI Generative Text-to-Speech technology, including the utilization of various Language Models that synthesize *dialect*, Blackmon has created language by adjusting, adapting and altering Western Language to affect the Speech output.

Continues...

For Example:

Blackmon utilizes “Karen,” on a daily basis as a vocal AI Assistant. Further, Blackmon works with the Language Model “Karen,” on the works *Me & Karen*, published 2021. In the case of *Literary Hip-Hop: Capitalist Americans*, Blackmon utilizes and collaborates with “Fiona.” In the case of these two Language Models, “Karen,” and “Fiona,” are synthesized to “express,” themselves with a Western Australian and Western Scottish dialect. In these cases, pronunciation is different among the dialects despite the Western spelling being the same. It is the Authors “work,” and the Author’s Conceptual Intellectual Property to “craft,” or create language based on the Author’s intention. In the case of works created with Language Models “expressing,” disparate dialectical language output, Blackmon employs the use of grammatical alteration, adaptation, spacing, paragraph breaks and the employment of accents and diacritical marks, to “compose,” what can be considered the “expression.” Further, as Blackmon not only wrote the literary work but also “crafted,” Engineered, the Language Model to pronounce and thereby produce an achievable and conceptually “required,” output by the Reader, it is our opinion that the work is entirely the product of Tavarus Blackmon Art LLP, despite any AI Generative tools employed in the creation of the work.

In the case and Registration of *Me & Karen*,

It is not only that the Author or Artist or otherwise Creator utilizes AI Generative Language Models and Text-to-Speech technology in the production of this work, but that in addition to the use of the underlying AI technology to develop the lyrics and vocal recitation constituting an encodable audio file, the Artist produces by hand illustration and hand animation, discrete Motion Pictures associated with each “work,” of Music or *song* on the Album. These works are external to the process of AI Generative development but constituent to the larger Copyright. Further, that through the process of AI Language Model Engineering, the Artist has altered otherwise “natural,” written Western Language to produce the desired pronunciation or “composition,” through the use of a Computer Reader with a Western Australian dialect. It is thereby not only that the Artist has utilized AI Generative Text-to-Speech technology, but that through the Language Model Engineering process, made available in the User Interface UI of application’s *Natural Reader* and *Adobe Audition*, Blackmon has not “used,” the “Reader’s” voice in the creation of the work, rather, the AI Computer Reader recites in the “voice,” *designed* by the Author, Artist or otherwise Creator, in this case Tavarus Blackmon Art LLP.

In the case and Registration of *Blackmonster Family Reliquary*

It is not only that the Author, Artist or otherwise Creator utilizes AI Generative Text-to-Image assets in the creation of this work, but also contributes hand illustrations and hand animations, composites, to the final composition. Also, Blackmon has designed and thereby creates the Art Direction of the collection of published photographs. Further, in the process of Prompt-Engineering, Blackmon has included both their Given Name and Registered Trademark, Standard Character’s, in the Engineering process and text-prompt; thereby including their US Patent and Trademark Registration, Standard Characters, in the Engineering process. Lastly, in addition to the previously stated conditions of Production for this Project and US Copyright Registration, Blackmon has processed each AI Generative output to include Motion Pictures and Image files exported into appropriate codices. A small number of these works are also

composited and encoded with Improvised, live-recorded, Endurance Art vocal recitations by the Artist, further extending the Asset Class(es) of the US Copyright Registration and establishing Authorship of the Project which includes AI Generative Text-to-Image assets within the larger US Copyright.

#### In the case of Technology Developers

As my Organization utilizes AI Generative tools, models and algorithms, it is increasingly apparent that producing AI Generative works *is* a form of Development. That not only is the individual, user or organization utilizing AI Generative Models, tools and applications, but that in the process of engaging with the User Interface of an AI Application, they are given tools that facilitate the Development process.

*Night Cafe Studio* empowers these use cases to engage in the creation of AI Generative *creative* outputs including Image and Video. The User Interface allows for Developing the Generation and Engineering the dataset through Text-Prompting and customization criteria which uniquely Develop works that are specific to the Author, Artist or otherwise Creator. Further, that AI Seeds, within the UI, make it possible for the use case ensure reproducibility.

The processes I have described above outline the Engineering and Text-Prompting possibilities that can be employed in the creation of AI Generative works of creative output. In addition to these possibilities, applications like *Night Cafe Studio* provide User access to the dataset or Metadata used to and uniquely bound to the AI Generative output. This Metadata can be cataloged, recorded and further utilized by the Author, Artist or otherwise Creator within their Organization in both record-keeping and Creative ways. The Metadata associated with the Generative AI output is one dataset made available to this use case through the use of AI tools and applications that constitutes Developmental Data. Yet, in addition to this, the actual Metadata of the output file codices is different from and additional to the AI Generative Metadata, increasing the access to Developmental tools for Organizations.

#### In the Case of Researchers

My Fine Art Practice and Professional Practice is established and *verified* by *Research* and the two Master's Degrees in Studio Art that I earned in 2015 and 2018. In the process of earning the Master's of Art in Studio Art, I mounted and defended the Advancement to Candidacy Exhibition, *Dinner Theatre*, Else Gallery, CSU Sacramento 2014, I also mounted and defended the Master's of Art Thesis Exhibition, *Extra-Ordinary*, Library Annex Gallery, CSU Sacramento 2015 and wrote and Published the Master's Thesis, *The Politics of the Cartoon and Contemporary Art*, CSU Sacramento, Office of Graduate Studies, *ScholarWorks*. Further, In the process of earning the Terminal Master's of Fine Art Degree, I successfully mounted and passed Year 1 and Year 2 MFA Reviews under the direction of the UC Davis Art Studio Full-Time Faculty and successfully mounted and defended the Master's of Fine Art Thesis Exhibition, *Work Hurt*, as part of the UC Davis MFA Thesis Exhibition, *Present Tense*, under the direction of the MFA Thesis Committee and Graduate Advisor's: Darrin Martin, Shiva Ahmadi and Annabeth Rosen (Distinguished Professor of Art, Robert Arneson Endowed Chair,) Jan and Maria Manetti Shrem Museum of Art, 2018.

In addition to my *verified* Professional Practice including earning the Master's of Studio Art and Terminal Master's of Fine Art Degree(s), my practice has been involved in the study and

*Research* of Cryptology and Symbology including Epistemology, since 1999. Further Professional *Research* and investigation includes Western Color Theory, specifically the Munsell Color Theory and dimensional color-model, the phenomenological *physio-neurological* effects of Simultaneous Contrast, Dichromatic Contrast and their affect upon and relation to Dichromatosis and or, those with Dichromatic Vision, and their response to Visual Art expressed with Dichromatic Color pairs.

My *Research* includes the investigation and study of Digital Film and Video Production and Film Theory, including not only technological and Industry Standard training on equipment and the Production Pipeline, but training in a Production Studio, including a three-camera Studio with Control Panel or Video Switcher and full Lighting Grid. As an Undergraduate and Graduate student, I put myself through school in-part, as an employee of Academic Technology and Creative Services, as part of the E-Learning Department. In addition to earning a degree in *Film* as an Undergraduate student I was employed as a *Studio Operator* on the CSUS Campus, solely operating the Studio Classroom including two cameras, document cameras (ELMO), monitors, speakers, Professor's Wireless Microphone and backing up pre-recorded classroom content for dissemination on the University E-Learning portal and Cable Television. In this time I operated for courses in Computer Science, The College of Education and Nursing. Whereby through the process of recording these specific courses semester after semester, I learned about these fields while studying my Major and General Elective coursework.

Theoretical concerns of my Organization include the theories included in Papers by: Stuart Hall and their Cultural Studies on the *Encoding/Decoding Model* of Communications and the theory of *Hegemony* and *Mass Communication*; Laura Mulvey and *Film Theory* which presents the discursive topic of *Masculine Performativity* and other writings and papers discussing Film and dynamics of performance, including the paper, *Visual Pleasure and Narrative Cinema*. Inclusive theories to my Practice and *Research* involve the writing of Marshal McLuhan, specifically how Techno-Cultural advances are affecting the present state of American Culture, concerning the concepts of *Space and Place* and the Contemporary and Post-Modern concepts of *Placing* and or *Situating* in relationship to cultural and identity politics.

Further *Research* includes topics such as *Sound Art* and *Sound Installation* which are Art Practices and Fine Art Mediums including the theories and formal processes involved in *Tape Music*, *Music Concrete* and historically, includes the once controversial and illegal technique of *Back-Masking*, and dates back to the *Fluxus*, *Dada* and *Bauhaus* movements. Impetus for such *Research* includes the Doctoral Dissertation, *Sound Art and Spatial Practices: Situating Sound Installation since 1958*, Garcia Ouzounain, University of San Diego, 2008. Continued *Research* in the *Sound Space* includes my Graduate Studies at UC Davis, whereby I was mentored by the, then, Chair and Professor of The Music Department, Kurt Rhode and, the current Chair of The Music Department and Professor of Teaching, Sam Nichols. *Research* includes Composition, *Sound Art*, music technology and Computational *Audio Processing*. Additional tutelage provided by Eyvind Kang, Professor of Music, CalArts, Composition and Experimental Sound Practices, Performer-Composer Doctorate of Music Program.

My *Research* practice includes studies on and the production of works investigating AI Text-to-Speech Technological Models as combined with Philosophical texts, published on The Stanford Encyclopedia of Philosophy web database. Works include the transcribing of philosophical texts as to be "read," by Computer Reader's, as part of a broader Conceptual Sound Artwork, which includes Popular Music utilizing *Tape Music* and *Music Concrete*, techniques such as *Back-Masking*, tempo and pitch *shifting*, and the use of *Sound as Media*. Philosophical texts include the *Definition of Sprekken*, "expressed," by "Lekha," an AI Text-to-



Speech “reader,” digitally and electronically synthesized with a Hindi or (Indian) dialect; *The Definition of Morality*; *The Definition of Lying*; *The Prisoner’s Dilemma*; *The Encyclopedia of Universal Knowledge*, Chambers, 1863; and; *The Ludas Patronymicus: Or, the Etymology of Curious Surnames*, Charnock, 1868. Continued Research in this area includes Sound Artwork that combines both Creative Writing and Court Testimony, such as *When We Were Young We Killed the World*, paraphrased from *Romeo and Juliette*, and *Modality of Breath*, including Court Testimony from the Eric Garner wrongful death lawsuit.

Substantive Research includes the *Cartoon*, as expressed from the theoretical lens that spans the Comic-Strip, and such work like *Popeye*, Segar, 1929 and it’s adaptation to Moving Image by the Fleischer Brothers Studios, which includes racist and xenophobic portrayals of *Black*, and Middle Eastern Cultures to the the *cartoon*-like caricatures and portrayals of D.W. Griffith and modern Editing Theory, and such works as *The Birth of a Nation*, 1915, which depict and portray animalistic and predatory *Black* and African-American personas. Additionally, the larger Art Historical term ‘cartoon,’ which is the outline or “study,” for or of an Artwork: placed upon and reproduced upon Cotton Canvas, whereby the Artist or *Painter*, uses the ‘cartoon,’ as a preparatory step in the creation of and completion of the painting process. The term ‘cartoon,’ within the Fine Arts itself describes a process. However, this term can be applied broadly within Modern and Contemporary Fine Art to include the works of the Artist’s within the Chicago-based *Hairy-Who*, or the *Chicago Imagists* and the Northern California *Funk Art* Movement, who’s work and practitioners found scholarly work and research at the School of the Art Institute of Chicago and, University of California Davis respectively, and also included instructing in the Art Department at California State University, Sacramento. But this group can also be expanded to loosely include Artist’s Peter Saul who once exhibited at CSUS and Robert Colescott who taught at Encina High School. Both have connections to the Northern California *Funk Art* Movement through their connections to the practitioners in the Region and Teaching in the Region. Additional practitioners of the Fine Art Cartoon can be traced to the East Coast and Los Angeles, as well as the Bay Area, with the works of Carroll Dunham, Lari Pittman and Robert Williams (*Mr. Bitchin’*)-who also founded *Juxtapoze Magazine* and was also employed as an illustrator in The Bay Area, working with Robert Crumb and helping to establish *Car Culture*, working as an Artist with Ed *Big Daddy* Roth. Dunham also has roots in Abstract Expressionism and the Yale School of Art and Pittman’s career can be traced to Cal Arts. This larger loosely connected group of Artists and Educators work is involved with not only the *cartoon* from what can be considered a formal or cultural or aesthetic lens, but who’s work is informed by unique conditions paramount the identity of the Artist, including Feminism, Racism, Post-Modernism, Gay and Lesbian rights and Globalism in a Post-War America.

Continued Research includes access to and utilization of Databases such as the *MMI Facial Expression Database*, Beta access to AI Training and Text-to-Speech Language Models and Text-to-Image Generative algorithms, including the production of Image and Video output. Beta access and testing includes the applications: Dall-E II, Chat-GPT, Adobe Content Credentials beta (Content Authenticity Initiative/CAI), Generative AI Search beta, *Google*, and, utilization of Neural Network Technology facilitating the Transcription Process with applications *Otter* and *Veed*. Finally, my Research includes Blockchain technology broadly referred to as Web3. In the normal operation of my Fine Art Services Business, my organization has Deployed thirty (30) “Smart Contracts,” on the Ethereum Layer 1 Blockchain Network. This process includes complicated *verification* and *validation* on the Network through the process of the Ethereum Network Consensus Mechanism. This process includes both “human,” *validators*, a Proof-of-Work Era “human,” curatorial process or *consensus*, and, Artificially Intelligent Neural Networks that aid the process of securing the Ethereum Network, providing Provenance on the

Blockchain and further establishing the *verified* status of Creators on the Blockchain as to limit the 'mining,' of Copyright Protected works.

In the case of the Public

My daughter is a thirteen-year-old Artist and Creator with aspirations to compete Globally with Artist's around the World. In the normal operation of her Art practice, Drawing and Improvisation are work activities. As she does not have a formal Art training and insists on completing her own *Research* and study, she has been able to create countless Worlds, ecosystems and Intellectual Property. Her young perspective is that AI poses a threat to her work. That, as a traditional Drafter, her work is thereby diminished by the Output of AI Tools, applications and algorithms.

I am greatly concerned that my child feels that AI will and has the potential to "replace," her authenticity. It is my opinion that if the Public is *not* concerned about AI Models replacing day-to-day tasks, they must understand AI has already become a part of our daily lives. Which, if I were to seek Customer Service by any Service Provider, whether State, Federal or a Private Organization, that I would be met by as I have been met by for some years, an Artificial Intelligent Assistant that aides in the facilitation of my Service Call.

As AI Models and applications continue to develop at an exponentially increasing pace, with the assistance of AI Models that update the algorithm and Training Model and *code*, or, application, faster than was before possible, AI Models, tools, and applications will further become a part of our Culture and daily lives. Further, that as the Models are updated and the technology advances, that previously before unrealized or unimaginable access, including DEI accessibility, and utility may or can potentially be gained from the use of Artificial Intelligence.

It is my opinion that the Public has an opportunity with AI that was not before available. AI technology can create opportunity if employed equitably. But, that the use of AI Models, tools and applications does not constitute "work." An Author, Artist or Creator must be present in the creation of "work," as without the intervention, validation, verification or curation of a "human," AI Models, tools and applications are "processes," that do not constitute "work," because they do not employ both "mental," or "physical," labor. That in fact, an AI Model does not produce "mental," labor without a MIND and that an AI Model does not produce "physical," labor without a BODY.

But, as the Public is concerned so too am I. That despite these facts that cannot be stripped from the use of and utility of AI Models, tools and applications, they do produce something *beautiful*. In philosophical terms, and the Public should be concerned with this, what is *beautiful* is to say that something is worth "looking," at, "caretaking," providing *salvation* for or Conservatory efforts to maintain. And, simply put, what is *beautiful* is both not *ugly*, or in the writing of Immanuel Kant, what is beautiful is dichromatic-ally opposed to what is [not], in this case: *night, blackness*,

*darkness, animalistic: Sublime.* This thinking poses many problematic issues for Contemporary Global Discourse, Culture and Society. However, if the Public agrees that what is *beautiful* is inherently valuable or, that what is *beautiful* is intrinsically worth material *salvation*, than the Public should be concerned for AI Models, tools and algorithms. In fact, if “Culture,” in general, was able to caretake or further “adopt,” a technology, than AI is not merely a Model, tool or application to “use,” “do,” or manipulate. It may be a “relationship,” to foster. That, as the AI Technological Models, tools and applications advance through update the “expressions,” output by the AI Model and Algorithm are not simply “products,” to commodify but, *Expressive* in nature; that is to “show,” or “convey.” Within the Arts, historically, what is expressive is not merely looking upon what is shown but that to **be** expressive, is to employ Abstract [thought/critical] processes in the production of work as to, through causation, impart, engage or stimulate a “feeling,” or emotion with [a] the viewer. Which, the relationship between the Author, Artist or otherwise Creator, in this case an AI Generative Model, establishes the AI Model’s *intentionality* with the (a) viewer. That is, the AI Model uses Abstract “ideas,” composited into pictorial “expression(s),” that the AI Model must first intend to “touch,” through pre-conceived processes, the viewer. Therefore, the viewer will “feel,” through relation or connotation or denotation, what the AI Model intends. Thus, it is not simply that an AI Model or tool or application is “capable,” of producing something *beautiful*, but rather, I profess the AI Model was *able* to produce what can be considered *beautiful* because it intends to “show,” or “covey,” the existence of *Beauty*, thereby not merely “expressive,” but possessing qualities absent of body and mind but yet, inherently emotive.

[**Emotion** def: a natural instinctive state of mind deriving from one's circumstances, mood, or relationships with others.]

[**Instinct** def: a complex and species-specific pattern of activity that's carried out by the nervous system. It's not a reflex, but rather an adaptive response to a stimulus in one's environment.]

2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

In our opinion, the increasing use and distribution of AI-generated material raises unique issues for our sector of Fine Art and the wider field of the Arts. Recent articles about AI-generated works have interesting view-points concerning the use of such work including authorship and philosophical differences of opinion:

[<https://www.smithsonianmag.com/smart-news/artificial-intelligence-art-wins-colorado-state-fair-180980703/>]

However, recent Federal Court rulings have also drawn our attention:

[<https://www.nytimes.com/2023/08/21/arts/design/copyright-ai-artwork.html>]

Also, the *AI Registration Guidance* is an interesting issue concerning our business interests:

“the term ‘author,’ used in both the Constitution and the Copyright Act, excludes non-humans.”

“[i]f a work's traditional elements of authorship were produced by a machine, the work lacks human authorship and the Office will not register it.

As we have outlined thus far in this response, Tavarus Blackmon Art LLP and Creative Foundations have filed registered US Copyrights that utilize AI-generated “work,” within the composition or composite constituting an Artwork. But that in the process of using such “work,” it has been extensively Authored by Tavarus Blackmon Art LLP through processes external to the AI-generation and encoding, compositing and Publication across Platforms which constitute Authorship.

Further, that such technology as more recently concerned includes the output of generative AI images a video files. These files are representative of the codices that compress data into a visible format. Further, digital images and videos are *applications* with embedded metadata that facilitates the process of being visible on an Internet Browser. Also, that as Artificial Intelligence and AI-generated “work,” develop and advance, it will be difficult to distinguish what is real from misinformation. I hope that Initiatives such as the Content Authenticity Initiative are continued to develop more access for Developers and Creators to authentication software that prevents the misuse of Artist’s copyrightable material. That despite the continued use of AI-generated “work,” not only will processes be established to continue its fair and responsible utility, Industry Standard Applications are developed and made accessible for the Public to utilize AI-generated *utility services*, including those now available and future iterations that will have more direct benefit in the fields of Medicine, Psychology.

I believe that in the Arts AI-generated “work,” can be developed to be included within the Fine Art Studio Practice. As a practicing Artist and Educator, it has been exciting and fulfilling to work with this technology and apply the beneficial utility of the technology to my production, digital fabrication process and Studio Practice.

As this discussion continues and potential future Ruling(s) affect the use of such technology, (I) we wish to be a part of the broader discourse around AI-generated “work,” and the wider field of the Arts. As a Business Model, we are engaged in the use, *Research* and investigation of emerging technologies including AI. We have been involved in the use and utility of AI since 2012 and have invested thousands of hours crafting, Engineering and creating language with the use of AI Language Models. And as an Educator, I feel that students must be made available all and any resource that aids in the successful facilitation of moving from matriculation to

graduation. Finally, as a Father, I want to prepare my children for a future that I can not imagine. Which means that (I) we place faith in the work (I) we do to provide for our family. This work includes the use of AI-generated “expressions,” which, may not be present of Mind and Body yet no second to, none the less, a human Author without the necessity to create.

3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

Tavarus Blackmon Art LLP is referring to the following publication(s) regarding the economic effects of generative AI and the issue of AI in relationship to the wider field of the Arts:

*The Economics of Artificial Intelligence: An Agenda*

[<https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda>]

*The Impact of Artificial Intelligence On The Future Of Workforces in the European Union and The United State of America*

[<https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf>]

*The University of Oxford, AI and the Arts: How Machine Learning is Changing Creative Work*

[[https://www.oii.ox.ac.uk/wp-content/uploads/2022/03/040222-AI-and-the-Arts\\_FINAL.pdf](https://www.oii.ox.ac.uk/wp-content/uploads/2022/03/040222-AI-and-the-Arts_FINAL.pdf)]

*Arguments for the Rise of Artificial Intelligence Art: Does AI Art Have Creativity, Motivation, Self-awareness and Emotion?*

[[https://www.researchgate.net/publication/370109562\\_Arguments\\_for\\_the\\_Rise\\_of\\_Artificial\\_Intelligence\\_Art\\_Does\\_AI\\_Art\\_Have\\_Creativity\\_Motivation\\_Self-awareness\\_and\\_Emotion](https://www.researchgate.net/publication/370109562_Arguments_for_the_Rise_of_Artificial_Intelligence_Art_Does_AI_Art_Have_Creativity_Motivation_Self-awareness_and_Emotion)]

*The Creativity of Artificial Intelligence in Art*

[[https://www.researchgate.net/publication/359809445\\_The\\_Creativity\\_of\\_Artificial\\_Intelligence\\_in\\_Art](https://www.researchgate.net/publication/359809445_The_Creativity_of_Artificial_Intelligence_in_Art)]

*AI Art: Machine Visions and Warped Dreams*

[[http://openhumanitiespress.org/books/download/Zylinska\\_2020\\_AI-Art.pdf](http://openhumanitiespress.org/books/download/Zylinska_2020_AI-Art.pdf)]

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? <sup>[40]</sup> How important a factor is international consistency in this area across borders?

### *The EU AI Act*

[<https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>]

The above European Commission proposal and June 14, 2023, *Parliaments negotiating position on the AI Act* are in discussions with EU countries on the Council to form the Final Law by years end.

The proposal outlines Public protections, safeguards and requirements that AI systems be overseen by people, not automation processes, to prevent harm. This would be the World's first Law on AI, its systems, uses and measures enacted to promote the benefits of AI including the Industries of Healthcare, Transportation and Manufacturing whilst preventing harmful and unwanted outcomes through the use of the technology. This includes procedures for making complaints about the affects of the use of AI on the Public including authorized use by Law Enforcement that infringes on the Rights and Laws protecting the Public.

(I) we feel that as the access to this technology is made available through electronic Global web dissemination, that international consistency should be considered in the area of AI technology, and that proposed Laws to protect the Public from the harms of the use of AI in the EU can be applied in the US, including Legal recourse that infringes upon the Rights of a US Citizen by Law Enforcement, using AI-generative tools in the process of conducting an Arrest, Search or Investigation, should be considered. The EU proposal outlines a framework and policy for establishing transparency and a "technology-neutral," definition of AI technology, ensuring an interoperable Framework for the management of and development within AI systems is available now and in the future. (I) we feel that as AI, Blockchain Technology, Content Authenticity Initiative and Financial Technology develop and mature in the US including Globally, that the US should position itself to Responsibly Develop AI Technology. This includes AI technology that not merely, currently, has applicable Creative use cases, but [it's] the potential developments, achievable, with the use of AI tools that are as yet integrated into Infrastructure including Communications and Information Technology. That AI is considered, as to be not just a novel resource but a *user service*, facilitating not just creative processes but daily tasks and activities.

See also:

*From China to Brazil, here's how AI is regulated around the world*

[<https://www.washingtonpost.com/world/2023/09/03/ai-regulation-law-china-israel-eu/>]

The following article describes regulations around the world concerning AI Models, tools and applications. Brazil, for example, has a draft AI law, product of three years on this topic. And, has produced a 900-page Senate committee report on AI. China has published draft regulation on AI and produced the *Next Generation Artificial Intelligence Development Plan*. Israel's Ministry of Innovation, Science and Technology has published draft policy on AI regulation which aims to be "compatible," with global AI best practices. Italy, at once, banning CHAT-GPT, has allocated funding to support work-displacement due to "digital transformation." Japan has revised the country's Copyright Act which includes AI as a consideration to the new laws. And, the UAE's *National Strategy for Artificial Intelligence*, seeks to develop the technology and aims to "observe and integrate," global best practices.

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

As I have outlined in this response, there are several methods that my organization have undertaken to establish what can be considered Authorship regarding works that include AI-generative assets as part of the fabrication process. Text-prompting, Language Model Engineering, Generative Engineering, adapting Language Model output with linguistic devices, and the inclusion of USPTO (pending) registrations, Standard Characters within the text-prompt, etc. But what we would like to address and consider is the application of the US Copyright to works that have not been "authored," by a human. That is to say, (I) we wish to seek US Copyright for works considered *authored*, by our human interest, but also not created by a "human." That significant customization within the UI of AI-generative applications provides straightforward access to generate uniqueness, randomness and reproducibility. And through the process of working *with* and utilizing this technology as a Graduate Student, Professional Artist and Business Owner, we wish to seek protection from the unauthorized use of our Organizations Creative Production Intellectual Property and Copyright, including our Inventory and Archive which includes the Intellectual Property such as AI-generative assets used in the creation of such work.

## Training

*If your comment applies only to a specific subset of AI technologies, please make that clear.*

6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?

Tavarus Blackmon Art LLP will utilize Google, Generative-AI Search and CHAT-GPT to furnish responses to the following Section and questions:

*Google AI Generative Search beta*

Training AI models with copyrighted data is likely considered fair use. However, using copyrighted works to generate content may not be considered fair use

To train an AI with images, you need to create a training dataset that contains the images you want to feed into the algorithm. These images may have been scraped from the internet and will likely include works that are protected by copyright

AI models can scan copyright-protected content at scale to distill an image's style, a novel's plot, or a program's logic

AI does not have any rights under copyright law. Therefore, there is no legal obligation to indicate that AI was used to generate the work

Companies that have developed and are responsible for AI systems may be able to be sued for copyright infringement



6.1. How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

*CHAT-GPT*

<https://chat.openai.com/share/802abf58-181f-4c88-a911-67cf4dfabe2e>

How or where do developers of AI models acquire the materials or datasets that their models are trained on?

AI developers and researchers acquire datasets for training their models from various sources:

- **Publicly Available Datasets:** There are many publicly available datasets for a wide range of tasks, including:
  - ImageNet for image classification
  - COCO for object detection and segmentation
  - SQuAD for question answering
  - IMDb for sentiment analysis, etc.
- **Academic Institutions:** Universities and research institutions often release datasets related to their research.
- **Government and Public Institutions:** Datasets like census data, climate data, or transportation statistics are sometimes made available by governments for public use.
- **Data Brokers and Marketplaces:** There are organizations and platforms that aggregate and sell datasets. Some marketplaces like AWS Data Exchange facilitate the purchase and exchange of datasets.
- **Web Scraping:** With permissions and respecting robots.txt and terms of service, one can gather data from the internet, e.g., text from websites, images, etc.
- **Self-collection:** In some cases, researchers or organizations collect their own data. This is common in domain-specific applications where unique data is required.
- **Simulated and Synthetic Data:** Sometimes, real-world data is hard to obtain, so researchers use simulators or

synthetic data generators. This is common in robotics and some medical imaging tasks.

- **Partnerships and Collaborations:** Companies might form partnerships to share non-sensitive data.
- **Crowdsourcing Platforms:** Websites like Amazon Mechanical Turk allow for data collection and labeling by distributing tasks to a large number of human workers.
- **Transfer and Pre-trained Models:** Instead of training a model from scratch, researchers often use a pre-trained model on a large dataset and fine-tune it on a smaller, domain-specific dataset.

However, when obtaining datasets, it's crucial to consider:

- **Ethical Implications:** Respect user privacy, avoid bias, and ensure that the data is sourced ethically.
- **Licensing and Permissions:** It's essential to make sure you have the right to use the data, especially for commercial purposes.
- **Data Quality:** Raw data often requires preprocessing and cleaning to be useful.
- **Representativeness:** Ensure that the dataset accurately represents the problem space and doesn't introduce biases.

In all cases, AI developers need to be conscious of privacy, ethics, and the legal implications of how they source and use data.

*Continued...*

To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

Third-party entities, including academic researchers, private companies, and even individuals, play a significant role in collecting and curating datasets for AI model training. The extent of their involvement varies based on the type of data, its domain, and the intended application. Here's a breakdown:

- **Academic Research:** Many foundational datasets used in the AI community originate from academic research. For example:
  - **ImageNet**, a large-scale image classification dataset, was developed by researchers at Stanford and Princeton.
  - **SQuAD (Stanford Question Answering Dataset)** came from Stanford University.
  - **MNIST**, a basic dataset for handwritten digit recognition, was created in collaboration between NIST (a U.S. government agency) and researchers.
- **Private Companies:** Companies with access to vast amounts of data often release subsets for research purposes (after ensuring privacy and compliance with regulations):
  - **Google** has released datasets like OpenImages for computer vision tasks.
  - **Yelp** provides a dataset for sentiment analysis and classification.
- **Specialized Data Collection Companies:** Some companies specialize in curating and selling datasets tailored for specific AI applications. They often collect data explicitly for machine learning purposes, ensuring high-quality annotations and comprehensive coverage of the problem space.

- **Collaborations & Consortia:** Occasionally, multiple entities, including companies, academic institutions, and non-profits, come together to curate datasets. Such collaborations can lead to vast and high-quality datasets.
- **Crowdsourcing:** Both academic researchers and companies often resort to platforms like Amazon Mechanical Turk to collect and label data. The advantage here is that diverse data points can be gathered from a wide range of participants.
- **Competitions & Challenges:** Platforms like Kaggle host data science competitions where companies or researchers provide datasets for specific challenges, allowing for data collection and model development to happen in tandem.
- **Domain-specific Data:** In domains like healthcare, finance, or geospatial analysis, specialized entities collect and curate data due to the domain's unique requirements. For instance, medical imaging datasets might come from hospitals or medical research institutions.
- **User-generated Content:** Platforms with user-generated content (e.g., social media platforms, forums, review sites) can be a rich source of data. However, data collection from these sources raises significant ethical and privacy concerns and usually requires anonymization and other preprocessing.

To sum up, third-party entities are pivotal in the AI ecosystem when it comes to data collection. Whether through rigorous academic research, targeted business efforts, or collaborative ventures, they provide the foundational material upon which many AI models are built. However, as data becomes increasingly central to AI, questions of privacy, ethics, and representation become paramount, influencing how and from where data is sourced.

6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?

### *CHAT-GPT*

To what extent are copyrighted works licensed from copyright owners for use as training materials?

AI developers and researchers acquire datasets for training their models from various sources:

- **Publicly Available Datasets:** There are many publicly available datasets for a wide range of tasks, including:
  - ImageNet for image classification
  - COCO for object detection and segmentation
  - SQuAD for question answering
  - IMDb for sentiment analysis, etc.
- **Academic Institutions:** Universities and research institutions often release datasets related to their research.
- **Government and Public Institutions:** Datasets like census data, climate data, or transportation statistics are sometimes made available by governments for public use.
- **Data Brokers and Marketplaces:** There are organizations and platforms that aggregate and sell datasets. Some marketplaces like AWS Data Exchange facilitate the purchase and exchange of datasets.
- **Web Scraping:** With permissions and respecting robots.txt and terms of service, one can gather data from the internet, e.g., text from websites, images, etc.
- **Self-collection:** In some cases, researchers or organizations collect their own data. This is common in domain-specific applications where unique data is required.
- **Simulated and Synthetic Data:** Sometimes, real-world data is hard to obtain, so researchers use simulators or

synthetic data generators. This is common in robotics and some medical imaging tasks.

- **Partnerships and Collaborations:** Companies might form partnerships to share non-sensitive data.
- **Crowdsourcing Platforms:** Websites like Amazon Mechanical Turk allow for data collection and labeling by distributing tasks to a large number of human workers.
- **Transfer and Pre-trained Models:** Instead of training a model from scratch, researchers often use a pre-trained model on a large dataset and fine-tune it on a smaller, domain-specific dataset.

However, when obtaining datasets, it's crucial to consider:

- **Ethical Implications:** Respect user privacy, avoid bias, and ensure that the data is sourced ethically.
- **Licensing and Permissions:** It's essential to make sure you have the right to use the data, especially for commercial purposes.
- **Data Quality:** Raw data often requires preprocessing and cleaning to be useful.
- **Representativeness:** Ensure that the dataset accurately represents the problem space and doesn't introduce biases.

In all cases, AI developers need to be conscious of privacy, ethics, and the legal implications of how they source and use data.

To your knowledge, what licensing models are currently being offered and used?

Third-party entities, including academic researchers, private companies, and even individuals, play a significant role in collecting and curating datasets for AI model training. The extent of their involvement varies based on the type of data, its domain, and the intended application. Here's a breakdown:

- **Academic Research:** Many foundational datasets used in the AI community originate from academic research. For example:
  - **ImageNet**, a large-scale image classification dataset, was developed by researchers at Stanford and Princeton.
  - **SQuAD (Stanford Question Answering Dataset)** came from Stanford University.
  - **MNIST**, a basic dataset for handwritten digit recognition, was created in collaboration between NIST (a U.S. government agency) and researchers.
- **Private Companies:** Companies with access to vast amounts of data often release subsets for research purposes (after ensuring privacy and compliance with regulations):
  - **Google** has released datasets like OpenImages for computer vision tasks.
  - **Yelp** provides a dataset for sentiment analysis and classification.
- **Specialized Data Collection Companies:** Some companies specialize in curating and selling datasets tailored for specific AI applications. They often collect data explicitly for machine learning purposes, ensuring high-quality annotations and comprehensive coverage of the problem space.

- **Collaborations & Consortia:** Occasionally, multiple entities, including companies, academic institutions, and non-profits, come together to curate datasets. Such collaborations can lead to vast and high-quality datasets.
- **Crowdsourcing:** Both academic researchers and companies often resort to platforms like Amazon Mechanical Turk to collect and label data. The advantage here is that diverse data points can be gathered from a wide range of participants.
- **Competitions & Challenges:** Platforms like Kaggle host data science competitions where companies or researchers provide datasets for specific challenges, allowing for data collection and model development to happen in tandem.
- **Domain-specific Data:** In domains like healthcare, finance, or geospatial analysis, specialized entities collect and curate data due to the domain's unique requirements. For instance, medical imaging datasets might come from hospitals or medical research institutions.
- **User-generated Content:** Platforms with user-generated content (e.g., social media platforms, forums, review sites) can be a rich source of data. However, data collection from these sources raises significant ethical and privacy concerns and usually requires anonymization and other preprocessing.

To sum up, third-party entities are pivotal in the AI ecosystem when it comes to data collection. Whether through rigorous academic research, targeted business efforts, or collaborative ventures, they provide the foundational material upon which many AI models are built. However, as data becomes increasingly central to AI, questions of privacy, ethics, and representation become paramount, influencing how and from where data is sourced.



6.3. To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?

#### *CHAT-GPT*

To what extent is non-copyrighted material (such as public domain works) used for AI training?

Non-copyrighted material, including public domain works, is often used for AI training to a significant extent. These materials are valuable resources because they can be freely used, modified, and distributed without copyright restrictions. Here's how non-copyrighted material is commonly used in AI training:

- **Text Corpora:** Public domain texts, such as classic literature, historical documents, and government publications, are frequently used to train natural language processing (NLP) models. Examples include the use of works by Shakespeare, Jane Austen, or U.S. government reports.
- **Image Datasets:** Public domain images, artwork, and photographs are used for training computer vision models. For instance, collections of artwork from museums or historical photographs are often used to create image datasets.
- **Audio and Music:** Public domain audio recordings, classical music compositions, and historical speeches serve as training data for speech recognition and music-related AI models.
- **Databases:** Various public domain databases, such as census data, scientific data, and historical records, are used for training machine learning models for tasks like data analysis and prediction.
- **Open Source Software:** Codebases and software that are released under open source licenses provide valuable

training data for AI models designed for code generation, code completion, or programming-related tasks.

- **Publicly Available Research:** Academic papers, preprints, and research articles are often used as text corpora for training NLP models. While not all research papers are public domain, many are freely accessible.
- **Publicly Accessible Websites:** Data can be scraped from websites that are publicly accessible and do not have restrictions against web scraping. This data is often used for various NLP and data analysis tasks.
- **Educational Materials:** Publicly available educational resources, such as textbooks, lecture transcripts, and course materials, can be used for training educational AI models and chatbots.
- **Historical Data:** Historical datasets, including genealogical records, historical weather data, and ancient texts, are valuable for historical research and AI applications.

While non-copyrighted material is a valuable resource for AI training, it's essential to ensure that the materials used are indeed in the public domain or have a permissive open license.

Researchers and developers must also be mindful of ethical considerations, especially when dealing with historical or culturally sensitive data, to avoid unintended consequences and respect privacy and cultural norms.

Alternatively, to what extent is training material created or commissioned by developers of AI models?

Developers of AI models often create or commission training material to a significant extent, especially when dealing with specific or domain-specific tasks where readily available datasets are limited or when they require data that closely aligns with their application. Here are some scenarios in which training material is created or commissioned:

- **Domain-specific Tasks:** In industries like healthcare, finance, or law, developers often need specialized datasets that reflect the nuances of their domain. They may collect and annotate data themselves to ensure relevance and accuracy.
- **Customized Corpora:** In natural language processing (NLP) tasks, developers may create customized text corpora by curating and aggregating content from various sources. This can involve collecting and cleaning text data specific to their application, such as medical records, legal documents, or industry-specific reports.
- **Data Augmentation:** Developers may generate additional training data by applying data augmentation techniques. This involves creating variations of existing data points, such as rotating, cropping, or adding noise to images, or paraphrasing text to increase dataset size.
- **Simulated Data:** In robotics, autonomous vehicles, and certain scientific applications, simulators are used to generate synthetic data for training. Simulated data allows developers to create diverse scenarios and control environmental conditions.
- **Annotated Data:** When building models for tasks like object detection or sentiment analysis, developers often

need labeled data. They may manually annotate data themselves or hire annotators to label data points accurately.

- **User-generated Content:** In some cases, developers may encourage user-generated content to generate training data. For example, social media platforms may use user interactions and content as training material for recommendation systems.
- **Collaborative Efforts:** Developers might collaborate with domain experts, researchers, or other organizations to create and share datasets for specific research or application areas.
- **Crowdsourcing:** Developers may use crowdsourcing platforms to collect or verify data. For example, they might use platforms like Amazon Mechanical Turk to gather labeled data or verify information.
- **Fine-tuning on Proprietary Data:** In industries like finance or e-commerce, companies often have access to proprietary data that is not publicly available. Developers can fine-tune models on this data to make them more effective for their specific business needs.
- **Artificial Intelligence Generated Data:** In some cases, AI-generated data can be used to supplement training material. For instance, Generative Adversarial Networks (GANs) can create realistic synthetic images.

While creating or commissioning training material offers greater control over data quality and relevance, it can be resource-intensive in terms of time, effort, and cost. Developers must also ensure that the data collection and annotation processes adhere to ethical guidelines, especially concerning privacy and bias. Additionally, they should consider how representative the created or commissioned data is of the real-world scenarios their models will encounter.

6.4. Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.

*CHAT-GPT*

Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)?

Developers of AI models typically retain at least a portion of their training materials after the training process is complete, and they do so for various purposes. The decision of which training materials to retain and how long to keep them depends on several factors, including data privacy, storage costs, legal requirements, and the specific use case. Here are some common purposes for retaining training materials:

- **Model Evaluation and Improvement:** Retaining a representative subset of the training data allows developers to evaluate their AI model's performance over time. This is crucial for identifying and addressing issues, improving model accuracy, and monitoring for concept drift (when the model's performance degrades due to changing data patterns).
- **Re-training and Fine-tuning:** Training materials are valuable for re-training and fine-tuning models. As new data becomes available or the model's objectives change, developers can use retained data to update and improve the model without starting from scratch.
- **Benchmarking and Research:** Developers may keep datasets used in research or model development to facilitate benchmarking and comparisons with future models or external research.
- **Debugging and Analysis:** When issues arise, having access to the original training materials can aid in debugging and root cause analysis. Developers can use this data to understand why a model made specific predictions or errors.

- **Compliance and Auditing:** In regulated industries or applications with legal requirements, retaining training materials can be necessary for compliance, auditing, and providing evidence of model behavior and data quality.
- **Ethical Considerations:** Retaining data can be important for ethical reasons, such as addressing bias and fairness concerns. Developers can review past training data to identify and mitigate biases in their models.
- **Reproducibility:** For transparency and reproducibility in research, it's essential to retain the training materials used to train a model. This allows other researchers to replicate experiments and verify results.
- **Legal and Liability:** In some cases, legal and liability considerations may require retaining data. For instance, if a model makes decisions that impact individuals' rights, it may be necessary to have a record of the data and processes involved in those decisions.
- **Security and Incident Response:** Retaining training data can be part of a broader data security and incident response strategy. In the event of a data breach or security incident, having access to the original training materials can help assess the impact and respond appropriately.
- **Knowledge Preservation:** Training materials can be retained for knowledge preservation within an organization. They serve as a historical record of the data used for AI model development.

It's important to note that data retention should be done in compliance with data protection regulations and privacy best practices. Sensitive or personally identifiable information should be appropriately anonymized or protected, and data retention policies should be established and communicated to ensure data is retained only for legitimate purposes and for an appropriate duration.

Please describe any relevant storage and retention practices.

Relevant storage and retention practices for training materials in AI development involve careful management of data to ensure compliance with privacy regulations, maintain data security, and facilitate the ongoing improvement and evaluation of AI models. Here are some key considerations and practices:

- **Data Classification:** Categorize training data based on its sensitivity and potential impact. Some data may be highly sensitive (e.g., personal health information), while others are less sensitive (e.g., publicly available text).
- **Anonymization and De-identification:** Prior to retention, ensure that any personally identifiable information (PII) or sensitive data is properly anonymized or de-identified. This helps protect individuals' privacy and reduces the risk of data breaches.
- **Data Retention Policies:** Develop clear data retention policies that specify how long training data should be retained based on its type and purpose. These policies should comply with legal requirements and best practices.
- **Secure Storage:** Store training materials in secure and controlled environments. This may include encrypted storage, access controls, and regular security audits to protect against unauthorized access or data breaches.
- **Version Control:** Implement version control for training data and models. This enables tracking changes and helps ensure that older versions of data are retained for reproducibility and auditing purposes.
- **Data Cataloging:** Maintain a catalog or metadata repository that describes the training data, its source, usage, and any associated documentation. This facilitates data discovery and understanding.



- **Regular Data Audits:** Conduct periodic audits of retained training data to verify its quality, relevance, and compliance with data retention policies. Remove data that is no longer needed.
- **Data Minimization:** Minimize the amount of training data retained to only what is necessary for model evaluation, improvement, and compliance purposes. Avoid unnecessary data storage.
- **Redundancy and Backups:** Ensure redundancy and regular backups of retained training data to prevent data loss due to hardware failures or accidents.
- **Legal and Ethical Compliance:** Stay informed about relevant data protection laws (e.g., GDPR, CCPA) and comply with legal requirements for data retention, transparency, and user rights.
- **Data Destruction:** Implement processes for secure data destruction when training data is no longer needed or when it reaches the end of its retention period. This may involve secure data erasure methods.
- **Documentation:** Maintain comprehensive documentation of data retention practices, including data inventory, retention schedules, and procedures for data handling.
- **Data Access Controls:** Restrict access to retained training data to authorized personnel only. Implement role-based access controls and monitor access logs for any suspicious activities.
- **Data Ethics Review:** Establish a data ethics review process to assess the ethical implications of retaining and using training data, especially when dealing with sensitive or potentially biased data.
- **Transparency and Communication:** Clearly communicate data retention practices to stakeholders,



including users whose data may be involved, to build trust and transparency.

These practices are essential for responsible AI development, ensuring data privacy, and maintaining data security throughout the AI model's lifecycle. They help strike a balance between the need for data retention for model improvement and compliance and the protection of individuals' privacy and data rights.

7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:

#### *CHAT-GPT*

To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained.

I don't possess personal experiences or knowledge, but I can provide information based on the training data and models used to develop me. AI models like mine, based on the GPT-4 architecture, are trained through a process known as unsupervised learning on a massive dataset. This training typically involves the following steps:

- **Data Collection:** A diverse and extensive dataset is collected from various sources, such as books, articles, websites, and more. The data can include text, images, or other types of information depending on the model's purpose.
- **Preprocessing:** Data preprocessing is performed to clean and format the raw data. This may involve removing irrelevant information, standardizing text, and converting data into a suitable format for training.
- **Tokenization:** Text data is typically divided into smaller units called tokens, which can be words, subwords, or

characters. Tokenization allows the model to process and understand the text.

- **Model Architecture:** AI models like mine have a specific architecture, often based on neural networks. GPT-4, for example, uses a transformer architecture, which is well-suited for natural language understanding and generation tasks.
- **Training Objective:** The model is trained to predict the next token in a sequence based on the context of previous tokens. This objective encourages the model to learn patterns, semantics, and grammar in the data.
- **Training Process:** The model iteratively learns from the dataset, adjusting its internal parameters to minimize prediction errors. This involves forward and backward passes through the neural network, updating weights and biases.
- **Loss Function:** A loss function measures the difference between the model's predictions and the actual data. The goal is to minimize this loss during training.
- **Optimization:** Optimization algorithms like stochastic gradient descent (SGD) or its variants are used to adjust the model's parameters in the direction that reduces the loss function.
- **Validation:** A portion of the dataset is set aside for validation. The model's performance is assessed on this validation data to prevent overfitting and gauge how well it generalizes to new data.
- **Hyperparameter Tuning:** Various hyperparameters, such as learning rates, batch sizes, and model architecture details, are fine-tuned to optimize the model's performance.
- **Iterative Training:** The training process is often carried out in multiple epochs or iterations to allow the model to learn progressively from the data.

- **Model Evaluation:** Once training is complete, the model is evaluated on a separate test dataset to assess its performance and generalization to unseen data.
- **Deployment:** The trained model can then be deployed for various applications, such as natural language understanding, text generation, or image recognition, depending on its intended use.

It's important to note that the training process can be computationally intensive and require access to substantial computing resources. Additionally, the quality and diversity of the training data significantly impact the performance and capabilities of the resulting AI model.

7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.

*CHAT-GPT*

Training an AI model involves the use and reproduction of training materials, which can raise questions about copyright implications. Here's a breakdown of how training materials are used and reproduced during the training process and the associated copyright considerations:

- **Nature of Reproduction:**
  - **Data Ingestion:** The training process begins with the ingestion of training data, which may include text, images, or other forms of information. During this step, copies of the data are made to be processed by the AI model.
  - **Data Augmentation:** To improve model generalization and robustness, data augmentation techniques may be applied. This involves creating variations of the original data, such as rotating images

or paraphrasing text, leading to additional copies or versions of the data.

- **Model Weights and Parameters:** The AI model's weights and parameters are adjusted during training to minimize prediction errors. These adjustments involve mathematical operations on the model's parameters and can be seen as transformations of the model itself.
- **Model Output:** As the model learns, it generates outputs, such as text or images, as part of the training process. These outputs are not direct reproductions of training data but are generated based on what the model has learned from the data.
- **Duration of Reproduction:**
  - **During Training:** Reproduction of training data, augmentation, and model parameter updates occur iteratively during the training process, which can span hours, days, or even weeks, depending on the complexity of the model and the size of the dataset.
  - **Inference:** After training, when the AI model is put into production or used for specific tasks, it may generate new content based on its training. This can be seen as a form of ongoing reproduction when the model generates predictions, text, or other outputs.
- **Copyright Implications:**
  - **Fair Use and Transformative Use:** AI model training, which involves copying and transforming data for the purpose of creating a new model with different capabilities, may be considered transformative and potentially fall under fair use exceptions in copyright law. Fair use allows for certain uses of copyrighted materials without permission from the copyright owner, provided it meets certain criteria (e.g., it doesn't harm the market value of the original work).

- **Data Licensing:** The copyright status of training data matters. If the data used for training is in the public domain, openly licensed, or provided with explicit permissions, copyright concerns are mitigated. However, if copyrighted data is used without proper authorization, it can lead to legal issues.
- **Model Output:** When AI models generate content as part of their operation, the output may not directly implicate copyright, as it's the result of the model's learned patterns and not a direct reproduction of training data. However, if the model generates copyrighted content verbatim, it could raise concerns.

It's important to note that copyright law can be complex, and the implications can vary based on jurisdiction and specific circumstances. AI developers should be aware of copyright considerations, seek legal advice when necessary, and ensure they have the appropriate rights or permissions for the data used in training. Additionally, they should consider ethical aspects, such as data privacy, bias, and fairness, when working with training materials.

7.2. How are inferences gained from the training process stored or represented within an AI model?

*CHAT-GPT*

Inferences gained from the training process are stored or represented within an AI model through the adjustment of the model's internal parameters and structures. The model learns patterns, relationships, and information from the training data, which is then encoded into its parameters. Here's how this representation and storage typically work:

- **Model Parameters:**
  - **Neural Network Weights:** In deep learning models, such as neural networks, inferences are stored in the form of weights and biases associated with each neuron or node in the network. These weights are adjusted during training to minimize prediction errors.
  - **Embeddings:** For natural language processing tasks, words or tokens are often represented as embeddings—dense vectors learned during training. These embeddings capture semantic relationships between words and are a form of inference about word meanings.
- **Internal States:**
  - **Hidden States:** In recurrent neural networks (RNNs) and transformers, hidden states capture information about the input sequence. These hidden states are updated at each step of processing and store context and knowledge gained from the data.
  - **Memory Cells:** In models like Long Short-Term Memory (LSTM) networks, memory cells store information over longer sequences and help the model capture dependencies in the data.
- **Architectural Components:**

- **Attention Mechanisms:** In transformer-based models, attention mechanisms capture relationships between different parts of the input data. These relationships are encoded into the model's architecture and used to make predictions.
- **Convolutional Filters:** In convolutional neural networks (CNNs), convolutional filters learn features and patterns in images. The patterns detected by these filters represent inferences about the visual data.
- **Output Layers:**
  - **Logits or Probabilities:** For classification tasks, the model's output layer represents inferences as logits or probability distributions over possible classes. The values of these logits reflect the model's confidence in each class.
- **Activation Functions:**
  - **Non-linear Activation Functions:** Activation functions like ReLU or sigmoid introduce non-linearity into the model, allowing it to capture complex relationships and inferences in the data.
- **Loss Function Minimization:**
  - **Minimization of Loss:** During training, the model adjusts its parameters to minimize a loss function. This process involves comparing model predictions to the ground truth labels in the training data. The adjustments reflect the model's efforts to make more accurate inferences.
- **Interconnected Layers:**
  - **Hierarchical Structure:** In deep models, such as deep neural networks, inferences are represented hierarchically across multiple layers. Lower layers capture low-level features, while higher layers build more abstract representations.
- **Embedding Spaces:**

- **Latent Spaces:** Variational autoencoders (VAEs) and generative adversarial networks (GANs) learn latent spaces where inferences about data variations are stored. These latent spaces can be used for data generation and manipulation.

In summary, inferences are represented within an AI model through the adjustment of its internal parameters, structures, and architectural components during the training process. These representations are the model's learned understanding of patterns and relationships in the training data, and they enable the model to make predictions and generate responses when presented with new input data.

7.3. Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to “unlearn” inferences from training?

#### *CHAT-GPT*

Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material?

AI models, including deep learning models, typically do not have a built-in mechanism to “unlearn” inferences gained from training on particular pieces of training material once the training process is complete. The learning process in these models is generally unidirectional, where they learn and adapt their parameters to minimize prediction errors during training. Once learned, these parameters are relatively fixed unless the model undergoes additional training or fine-tuning.

However, there are some ways to mitigate or update learned inferences:



- **Fine-tuning:** After initial training, models can be further trained on new data to adapt to changing patterns or to correct biases. This is a form of learning and adaptation, not unlearning.
- **Re-training:** If it's determined that the model's learned inferences are problematic or biased, developers may choose to retrain the model with different training data or training objectives. This process aims to replace the problematic inferences with improved ones but doesn't directly "unlearn" previous knowledge.
- **Bias Mitigation Techniques:** Post-processing techniques can be applied to the model's outputs to reduce bias or undesirable inferences. This doesn't erase the original inferences but rather adjusts model outputs to be more aligned with desired outcomes.
- **Knowledge Distillation:** Knowledge distillation is a technique where a large, complex model (teacher) transfers its knowledge to a smaller model (student). This process can be used to refine inferences or make them more interpretable.
- **Model Pruning:** In some cases, certain model parameters can be pruned or removed to reduce the impact of specific inferences. Pruning doesn't "unlearn" but reduces the model's reliance on certain knowledge.
- **Retraining with Diverse Data:** Training a model with more diverse and representative data can help it generalize better and adapt to new circumstances, potentially reducing the impact of specific inferences.

It's important to note that once a model has learned from training data, the original inferences remain encoded in its parameters. The goal is often not to "unlearn" but to adapt, mitigate, or improve these inferences as needed for specific applications. Furthermore, ethical considerations and responsible AI practices should guide any attempts to adjust model behavior to ensure

fairness, transparency, and compliance with legal and ethical standards.

If so, is it economically feasible?

The feasibility of "unlearning" inferences gained by an AI model depends on several factors, including the nature of the model, the specific inferences to be addressed, the availability of new training data, and the computational resources required. Here are some considerations related to the economic feasibility of unlearning:

- **Model Complexity:** The feasibility of unlearning may vary based on the complexity of the AI model. Simpler models may be more amenable to retraining or fine-tuning, while extremely complex models may require significant computational resources.
- **Data Availability:** To unlearn or retrain a model, you typically need access to new, diverse, and relevant training data. The cost and feasibility of obtaining such data can vary widely based on the domain and application.
- **Computational Resources:** Retraining or fine-tuning an AI model can be computationally intensive, especially for deep learning models. The availability of powerful hardware (e.g., GPUs or TPUs) and cloud computing services can affect the economic feasibility.
- **Development Effort:** Modifying or adapting an AI model to unlearn specific inferences often involves significant development and engineering effort. The cost of hiring skilled machine learning engineers and data scientists should be considered.
- **Ethical and Legal Considerations:** Adjusting or unlearning inferences must be done carefully to avoid introducing new biases or ethical issues. Compliance with

legal and ethical standards can add complexity and cost to the process.

- **Benefits vs. Costs:** The economic feasibility depends on the perceived benefits of unlearning. If retaining specific inferences harms the model's effectiveness or reputation, it may justify the cost of mitigation.
- **Iterative Approach:** In some cases, unlearning and model improvement can be an iterative process, requiring multiple rounds of training and evaluation. The cost can accumulate over time.
- **Resource Constraints:** Smaller organizations or research projects with limited budgets may find it less economically feasible to unlearn inferences compared to larger enterprises with more resources.

In summary, while it is possible to address or mitigate undesirable inferences learned by AI models, the economic feasibility depends on various factors. In some cases, it may be cost-effective and essential for maintaining model fairness, transparency, and effectiveness. In other situations, the cost and resource requirements may be prohibitive, and alternative approaches, such as bias mitigation techniques or ethical guidelines, may be more practical. Decisions regarding unlearning should be based on a careful cost-benefit analysis and consideration of ethical and regulatory factors.

In addition to retraining a model, are there other ways to “unlearn” inferences from training?

While retraining is one of the primary methods to adapt or adjust an AI model's learned inferences, there are several other techniques and approaches that can be used to "unlearn" or mitigate specific inferences without undergoing a full retraining process. These approaches focus on post-processing or modifying model outputs rather than changing the underlying model parameters. Here are some common methods:

- **Post-processing and Bias Mitigation Techniques:**
  - **Calibration:** This technique involves adjusting the confidence scores or probabilities assigned to model predictions to align them with the true accuracy of the model. It helps reduce overconfidence or underconfidence in predictions.
  - **Re-ranking:** After the model makes predictions, re-ranking can be used to prioritize or deprioritize certain outputs based on predefined criteria or fairness considerations.
  - **De-biasing:** Post-processing methods can be applied to model outputs to reduce biases in predictions, particularly for sensitive attributes like gender, race, or age. This can involve adjusting probabilities or re-weighting predictions.
- **Adversarial Attacks and Defense:**
  - **Adversarial Attacks:** Techniques like adversarial training introduce adversarial examples during model training to make the model more robust to certain types of inferences. Adversarial attacks can also be used to identify and expose model biases.
  - **Adversarial Defense:** Models can be fine-tuned with adversarial training or robustness techniques to

defend against unwanted inferences or adversarial inputs.

- **Rule-based Filtering:**
  - **Post-processing Filters:** Deploying rule-based or heuristic filters can help reject or modify model outputs that violate predefined rules or ethical guidelines. This approach is common in content moderation and chatbot development.
- **Interpretable Models:**
  - **Interpretability Techniques:** By using interpretable models or techniques like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (SHapley Additive exPlanations), developers can gain insights into how a model makes specific inferences and identify areas for adjustment.
- **Human-in-the-Loop:**
  - **Human Review:** Incorporating a human-in-the-loop approach involves having human reviewers assess model outputs and make adjustments or corrections when inferences are undesirable or biased.
- **Dynamic Thresholds and Policies:**
  - **Threshold Adjustments:** Models can be configured to have dynamic decision thresholds, allowing developers to control the trade-off between precision and recall, which can impact the types of inferences made.
  - **Policy-based Control:** Organizations can implement policies that govern the use of AI systems and establish guidelines for acceptable inferences.
- **Data Augmentation and Balancing:**
  - **Data Augmentation:** By augmenting training data with counterexamples or specific types of data, models can be trained to recognize and balance different inferences more effectively.

- **Regularization Techniques:**
  - **Regularization Losses:** Additional regularization terms can be added during model training to discourage specific types of inferences. For example, fairness-aware regularization can be used to reduce discriminatory inferences.

It's important to note that these techniques are often used in combination, and their effectiveness can vary depending on the specific use case and the nature of the inferences that need to be addressed. The choice of approach should be guided by the nature of the problem and the ethical considerations surrounding the AI system's use.

7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

*CHAT-GPT*

Identifying whether an AI model was trained on a particular piece of training material without access to the underlying dataset can be challenging, especially if the dataset is large and diverse. However, there are some indirect methods and techniques that may provide clues or hints about the model's training data. Here are a few approaches:

- **Bias or Artifacts:** If the training data contains specific biases, artifacts, or unique characteristics, these may manifest in the model's behavior or outputs. For example, if the training data contains biased language or specific types of errors, the model might exhibit similar biases or errors in its responses.
- **Fine-tuning and Transfer Learning:** If the model was fine-tuned on a smaller, domain-specific dataset, it may

exhibit knowledge or biases related to that domain. Examining the model's fine-tuning history and any additional data used in fine-tuning can provide insights.

- **Prompt Engineering:** For language models, like GPT-4, prompt engineering can influence the model's responses. The choice of prompts or input queries can lead the model to produce answers related to specific topics or datasets.
- **Specific Knowledge or Omissions:** The model's responses or knowledge may reveal hints about its training data. It may exhibit knowledge gaps or provide specific details that align with the content of the training material.
- **Model Artifacts:** Certain artifacts or patterns in the model's behavior, such as overuse or underuse of specific terms or concepts, may suggest the influence of particular training data.
- **Data Attribution:** Some models may include metadata or information about their training data sources, although this is not always the case. Checking model documentation or associated metadata can provide information about the training data.
- **Stylometric Analysis:** Stylometric analysis involves examining the writing style of text generated by the model. While it may not pinpoint the exact source, it can reveal similarities or differences in writing style that could be associated with specific training material.
- **External Data Sources:** Cross-referencing the model's responses with publicly available data or information from the web can sometimes lead to clues about its training data sources.

It's important to note that these methods are not foolproof, and they may not provide definitive proof of a specific training dataset. AI developers and organizations often consider model transparency and disclosure to address concerns related to



training data and model behavior. Ethical considerations, privacy, and intellectual property rights also play a role in determining whether and how such information can be accessed and disclosed.

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

### *CHAT-GPT*

Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use?

Determining whether the unauthorized use of copyrighted works to train AI models constitutes fair use is a complex legal question that depends on several factors. Fair use is a doctrine in copyright law that allows limited use of copyrighted material without permission from or payment to the copyright owner under certain conditions. These conditions are typically evaluated on a case-by-case basis. Here are some circumstances in which the unauthorized use of copyrighted works for AI model training might potentially be considered fair use:

- **Transformative Purpose:** If the use of copyrighted material is transformative, meaning it adds new value or meaning to the original work by repurposing it for a different and innovative use, it is more likely to be considered fair use. For example, using copyrighted text to train a machine learning model for sentiment analysis or language translation could be transformative.
- **Nature of the Copyrighted Work:** Fair use is more likely to apply when the copyrighted work used for training is factual or non-fictional rather than highly creative or fictional. For instance, using news articles for training may have a stronger fair use argument than using entire novels.



- **Amount and Substantiality:** The extent to which the copyrighted material is used matters. Using a small, non-substantial portion of a work for training is more likely to be considered fair use than using the entirety of a work. However, the evaluation also considers whether the portion used is the "heart" or most important part of the copyrighted work.
- **Effect on the Market:** If the use of copyrighted material for AI model training does not significantly affect the potential market for the original work, it is more likely to be considered fair use. If the use competes with or replaces the market for the original work, it is less likely to be fair use.
- **Research and Education:** Fair use considerations may be more favorable when the use of copyrighted material is for non-profit educational or research purposes, as long as the other fair use factors are met.
- **Public Interest and Innovation:** Courts may weigh the public interest and the potential for technological advancement when assessing fair use claims. Uses of copyrighted material that promote innovation and public benefit may be more likely to be considered fair use.

It's important to emphasize that fair use is a complex and context-specific legal doctrine, and determinations are made on a case-by-case basis. There are no absolute rules, and the outcome can vary depending on the specific circumstances, the jurisdiction, and the legal arguments presented. If you intend to use copyrighted material for AI model training without authorization, it is advisable to consult with legal counsel to assess the potential for fair use and to understand the associated legal risks.

Please discuss any case law you believe relevant to this question.

I don't have access to real-time data or specific legal cases beyond my last knowledge update in September 2021. However, I can mention some relevant copyright-related legal principles and landmark cases that were relevant to the fair use doctrine in the context of AI and data use. Please note that the legal landscape may have evolved since then, and consulting with legal experts is essential for up-to-date guidance.

- **Google LLC v. Oracle America, Inc. (2021):** This U.S. Supreme Court case revolved around the use of Java APIs in Google's Android operating system. The Court found in favor of Google, stating that Google's use of Java APIs constituted fair use, primarily due to transformative use and the functional nature of APIs.
- **Authors Guild, Inc. v. Google, Inc. (2015):** In this case, Google was sued for digitizing and indexing millions of books without obtaining permission from copyright owners. The court ruled in favor of Google, stating that its digitization efforts constituted fair use because it provided significant public benefits, including facilitating search and academic research.
- **American Geophysical Union v. Texaco Inc. (1994):** This case established that unauthorized copying of scientific journal articles for internal research purposes can be fair use. However, the fair use determination depends on the specifics of each case.
- **Campbell v. Acuff-Rose Music, Inc. (1994):** In this Supreme Court case, the Court held that the parody of a copyrighted song by the rap group 2 Live Crew constituted fair use because it was transformative and did not harm the market for the original work.

- **HathiTrust Digital Library Case (2012):** The HathiTrust Digital Library, a digital repository of books, was sued for digitizing copyrighted works for the purpose of creating accessible copies for people with print disabilities. The court ruled in favor of HathiTrust, citing fair use and the transformative nature of the use.

It's essential to note that while these cases provide guidance on the fair use doctrine, they do not specifically address the use of copyrighted works in AI model training. The application of fair use to AI model training will depend on the unique circumstances of each case, including factors such as purpose, amount used, market effect, and transformative nature. Courts evaluate each case individually, considering these factors to determine whether the unauthorized use constitutes fair use or not. Legal advice from experts well-versed in copyright law is crucial when dealing with such matters.

8.1. In light of the Supreme Court's recent decisions in *Google v. Oracle America* [\[41\]](#) and *Andy Warhol Foundation v. Goldsmith*,[\[42\]](#) how should the “purpose and character” of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning,[\[43\]](#) raise different considerations under the first fair use factor?

#### CHAT-GPT

In light of the Supreme Court's recent decisions in *Google v. Oracle America* [\[41\]](#) and *Andy Warhol Foundation v. Goldsmith*,[\[42\]](#) how should the “purpose and character” of the use of copyrighted works to train an AI model be evaluated?

The Supreme Court's decisions in "Google v. Oracle America" and "Andy Warhol Foundation v. Goldsmith" provide important guidance on how the "purpose and character" of the use of copyrighted works to train an AI model should be evaluated in the context of the fair use doctrine. These decisions emphasize the transformative nature of the use as a key factor in the fair use

analysis. Here's how the "purpose and character" should be considered:

- **Transformative Use:**
  - **Google v. Oracle America:** In this case, the Supreme Court emphasized the transformative nature of Google's use of Java APIs in its Android operating system. The Court recognized that Google's use transformed the APIs for a new and different purpose, enabling the creation of a successful mobile platform. This transformative aspect favored a finding of fair use.
  - **Andy Warhol Foundation v. Goldsmith:** The Supreme Court held that Andy Warhol's use of a copyrighted photograph to create a series of artworks, while not transformative in a traditional sense, still had transformative elements because Warhol added new expression and meaning to the original work. The transformative nature of the use favored fair use.
- **Commercial vs. Non-Commercial Use:**
  - The commercial or non-commercial nature of the use is a relevant factor. In both cases, the fact that the uses had commercial aspects did not preclude a fair use finding. The Court emphasized that a transformative use can still be fair use even if there is a commercial motive.
- **Nature of the Original Work:**
  - The nature of the copyrighted work is another factor to consider. Both cases acknowledged that the nature of the original works (APIs and a photograph) had varying degrees of creativity and functional aspects. However, the fair use analysis was not solely determined by the nature of the works.
- **Market Effect:**

- Consideration of the market effect is important. In "Google v. Oracle America," the Court noted that Google's use did not harm the market for Oracle's Java platform, which was a factor supporting fair use. In "Andy Warhol Foundation v. Goldsmith," the Court emphasized that the transformative nature of the use did not harm the market for the original photograph.

Applying these principles to the use of copyrighted works to train an AI model, the "purpose and character" of the use should be evaluated with a focus on whether the use is transformative, whether it adds new value or meaning to the original work, and whether it facilitates innovation and new creative expression. Additionally, the commercial or non-commercial nature of the use should not be determinative if the use is transformative.

It's important to note that fair use determinations are highly fact-specific and depend on the specific circumstances of each case. Legal advice from experts is crucial when assessing whether the use of copyrighted works in AI model training qualifies as fair use, considering these recent Supreme Court decisions and other relevant factors.

What is the relevant use to be analyzed?

When analyzing the relevant use of copyrighted works in the context of fair use, it's important to focus on the specific use or purpose for which the copyrighted material is being used. The relevant use refers to how the copyrighted work is employed, and it is a critical factor in the fair use analysis. Here are the key aspects of the relevant use that should be analyzed:

- **Purpose and Character of the Use:**

- Is the use transformative? Does it add new value, meaning, or creative expression to the original work? Transformative uses, such as commentary, criticism, parody, or research, are more likely to favor fair use.
- Is the use for commercial or non-commercial purposes? While a commercial purpose does not necessarily preclude fair use, it may be considered in the analysis.
- **Nature of the Use:**
  - What is the specific context in which the copyrighted material is used? Is it used for educational, research, news reporting, commentary, or other purposes?
  - Does the use involve reproduction, adaptation, or modification of the original work, or is it a more straightforward and unaltered use?
- **Amount and Substantiality of the Use:**
  - How much of the copyrighted work is used? Is it a small portion or the entirety of the work? Using only what is necessary for the intended purpose may weigh in favor of fair use.
  - Does the use include the "heart" or most significant and creative part of the copyrighted work?
- **Effect on the Market for the Original Work:**
  - Does the use potentially harm the market for the original work? If the use competes with or substitutes for the original work, it may weigh against fair use.
  - Does the use potentially create a new market or audience that does not substitute for the original work?
- **Context and Intent:**
  - What is the intent behind the use? Is it intended to educate, inform, criticize, or comment on the original work?

- What is the broader context of the use? Is it part of a larger transformative work or commentary?
- **Legal and Ethical Considerations:**
  - Are there legal or ethical considerations surrounding the use, such as privacy or data protection issues?
  - Does the use comply with relevant laws and regulations, including copyright law and any licensing agreements?

The fair use analysis involves a balancing of these factors, and no single factor is determinative. Courts consider the overall context and purpose of the use when making fair use determinations. Additionally, it's essential to recognize that fair use is a nuanced and fact-specific legal doctrine, and the outcome can vary depending on the specific circumstances of each case. Legal advice from experts is often necessary to assess whether a particular use qualifies as fair use.

Do different stages of training, such as pre-training and fine-tuning,<sup>[43]</sup> raise different considerations under the first fair use factor?

1 / 1

Yes, different stages of training in the context of AI model development, such as pre-training and fine-tuning, can raise different considerations under the first fair use factor, which evaluates the "purpose and character of the use." Here's how each stage may be analyzed:

- **Pre-training:**
  - **Transformative Nature:** The pre-training stage often involves training a model on a large corpus of data, which may include copyrighted materials. The



transformative aspect of this stage can be complex. On one hand, the pre-training process may not be immediately transformative, as it aims to build a foundational understanding of language or data patterns. However, it is essential for enabling transformative use in subsequent stages.

- **Commercial vs. Non-Commercial:** Pre-training, typically performed by organizations for research and development purposes, can have both commercial and non-commercial aspects. The commercial nature of the organization's goals may be less relevant at this stage.
- **Fine-tuning:**
  - **Transformative Nature:** Fine-tuning is where the AI model is adapted to perform specific tasks, such as natural language understanding or image recognition. The fine-tuning process can be highly transformative, as it tailors the model's knowledge and capabilities for a particular purpose. For example, fine-tuning a pre-trained language model for sentiment analysis or translation involves a clear transformative intent.
  - **Commercial vs. Non-Commercial:** Fine-tuning may include commercial applications, such as building AI products or services. The commercial aspect at this stage does not necessarily preclude fair use, particularly if the use remains transformative.

In the fair use analysis, the determination of whether pre-training and fine-tuning qualify as fair use should consider the overall context and purpose of each stage:

- Pre-training serves as a foundational step that enables subsequent transformative uses. While it may not be immediately transformative on its own, it can be seen as a necessary preparatory step.



- Fine-tuning is often where the primary transformation occurs, as the model is adapted for specific tasks. If fine-tuning involves creative or expressive elements, it may strengthen the fair use argument, especially if it leads to innovative and transformative applications.

As with all fair use considerations, the specific facts and circumstances of each case are essential. Courts will assess the purpose and character of each stage in the broader context of AI model development. Legal advice and a thorough understanding of copyright law are crucial when navigating fair use considerations in AI model development.

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

*CHAT-GPT*

Entities that collect and distribute copyrighted material for training purposes but do not engage in the actual training of AI models may still raise copyright considerations. The fair use analysis in such cases may focus on the entity's role in collecting and distributing copyrighted material. Here's how the fair use factors may apply:

- **Purpose and Character of the Use:**
  - The purpose of collecting and distributing copyrighted material for training purposes may be relevant. If the entity's goal is primarily educational, research, or facilitating technological advancement, it may be more likely to favor fair use.
  - Whether the entity is involved in transformative uses, such as creating educational resources or datasets for research, can also influence the analysis.

- **Nature of the Copyrighted Work:**
  - The nature of the copyrighted material collected and distributed matters. If the material consists of factual, non-fictional, or publicly available works, it may weigh in favor of fair use. However, if the material is highly creative and unpublished, it may raise more copyright concerns.
- **Amount and Substantiality of the Use:**
  - The quantity of copyrighted material collected and distributed can impact the analysis. If only a small portion of copyrighted works is used for specific purposes, it may be more likely to favor fair use.
  - However, if the entity collects and distributes extensive portions or entire works without authorization, it may raise concerns.
- **Effect on the Market for the Original Work:**
  - Assessing the impact on the market for the original work is important. If the entity's actions do not harm the potential market for the copyrighted material, it may weigh in favor of fair use.
  - If the entity's actions compete with or substitute for the market for the original work, it may raise questions about fair use.
- **Context and Intent:**
  - The context in which the entity collects and distributes copyrighted material matters. If it is for educational, research, or non-profit purposes, this context may be relevant.
  - The entity's intent and whether it seeks to enable transformative uses or innovative applications through the distribution of copyrighted material can influence the analysis.
- **Ethical and Legal Considerations:**

- Compliance with legal and ethical standards, such as copyright law and licensing agreements, is essential. Adhering to fair dealing provisions or educational exceptions in copyright law may strengthen the fair use argument.

Entities that collect and distribute copyrighted material for training without engaging in the actual training should be mindful of copyright law, ethical considerations, and licensing agreements. They should also consider the broader purposes and contexts of their activities, which can impact whether their actions are viewed as fair use. Legal advice from experts in copyright law is advisable to navigate these complex issues effectively.

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes.<sup>[44]</sup> How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? <sup>[45]</sup> Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

#### *CHAT-GPT*

The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes.<sup>[44]</sup> How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature?

The fair use analysis for the use of copyrighted materials in a training dataset or to train generative AI models can evolve if those models or datasets are later adapted for commercial purposes. When copyrighted materials initially used for noncommercial or research purposes transition to commercial applications, several factors come into play:

- **Purpose and Character of the Use:**
- The original noncommercial or research use may have been considered more favorably under the fair use doctrine. However, the transition to commercial use may change the character of the use. Courts may

assess whether the subsequent commercial use is still transformative or whether it is primarily for commercial gain.

- **Commercial Nature:**

- The commercial aspect of the adaptation is a critical factor. Commercial use, by itself, does not necessarily preclude fair use. However, it can affect how the "purpose and character" factor is weighed. Courts may consider whether the commercialization is incidental or essential to the use.

- **Amount and Substantiality of the Use:**

- The extent to which copyrighted materials were used and whether this usage expands in the commercial adaptation is significant. Using more copyrighted material than necessary for a commercial purpose can weigh against fair use.

- **Effect on the Market for the Original Work:**

- The potential impact on the market for the original copyrighted material is a key consideration. If the commercial adaptation competes with or substitutes for the market for the original work, it may weigh against fair use.

- **Context and Intent:**

- Courts may consider whether the entity's original intent for noncommercial or research purposes aligns with the commercial adaptation. An entity's intent to facilitate transformative uses or innovation may still be relevant.

- **Ethical and Legal Considerations:**

- Compliance with copyright law, ethical standards, and any licensing agreements is important, especially in the transition to commercial use. Adherence to fair dealing provisions or licensing agreements may be relevant.

It's important to recognize that the evolution from noncommercial or research purposes to commercial use can impact the fair use analysis, particularly under the first fair use factor. The analysis will be influenced by the specifics of the transition, the degree of transformation, the extent of commercialization, and the potential impact on the market for the original work.

Entities making this transition should be aware of potential changes in the fair use analysis and consider legal counsel to ensure compliance with copyright law and other relevant regulations. It's also important to continually assess and document the purpose, character, and nature of the use as it evolves over time.

Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

Yes, the source of funding for noncommercial or research uses of copyrighted materials, especially when provided by for-profit developers of AI systems, can be a factor that affects the fair use analysis. The fair use analysis, particularly under the first factor ("purpose and character of the use"), may take into account the source of funding in the following ways:

- **Impact on Commercial Character:**
  - If the funding for noncommercial or research uses is provided by for-profit developers of AI systems, it may raise questions about the underlying commercial nature of the project. Courts may consider whether the source of funding reflects an intent to eventually commercialize the project.
- **Intent and Motive:**

- The source of funding can be indicative of the entity's intent and motive. Courts may assess whether the entity initially received funding with the expectation of transitioning the project to commercial use. The intent behind the research or noncommercial use can influence the fair use determination.
- **Transformation and Innovation:**
  - The fair use analysis often focuses on whether the use is transformative and whether it adds new value or innovation. If for-profit developers fund research that leads to transformative uses, it may strengthen the fair use argument. However, if the funding primarily supports uses that benefit the for-profit entity without transformative elements, it may weaken the fair use argument.
- **Commercial Viability:**
  - The financial support provided by for-profit developers can affect the assessment of commercial viability. If the research or noncommercial use demonstrates potential for commercial applications that benefit the for-profit entity, this may be considered in the analysis.

It's important to note that while the source of funding can be a relevant factor, it is just one element in the overall fair use analysis. Courts consider multiple factors, including the purpose and character of the use, the nature of the copyrighted work, the amount used, the effect on the market, and more. The fair use determination is fact-specific and requires a holistic assessment of all relevant factors.

Entities engaged in research or noncommercial uses funded by for-profit developers should carefully consider the impact of funding on their projects, document their intent and motives, and ensure that their uses align with the transformative and non-

commercial goals that may support a fair use argument. Legal advice may also be beneficial to navigate complex fair use considerations in such scenarios.

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

*CHAT-GPT*

What quantity of training materials do developers of generative AI models use for training?

The quantity of training materials used by developers of generative AI models can vary significantly based on several factors, including the specific model architecture, the complexity of the task, the desired performance level, and the available computational resources. Here are some general considerations:

- **Model Size and Capacity:**
  - Large AI models, such as those with billions of parameters, typically require vast amounts of training data to perform well. These models have a high capacity for learning, and developers often use extensive datasets to exploit their full potential.
- **Task Complexity:**
  - The complexity of the task that the AI model is being trained for can influence the quantity of training materials required. More complex tasks may demand larger and more diverse datasets to achieve high performance.
- **Data Diversity:**
  - Diverse datasets can improve the generalization and robustness of AI models. Developers often aim to include a wide variety of data to ensure that the model can handle different inputs and scenarios.
- **Data Preprocessing:**

- Data preprocessing techniques can impact the quantity of training materials needed. Cleaning, filtering, and augmenting datasets can help improve the quality of training data and potentially reduce the required quantity.
- **Resource Constraints:**
  - Available computational resources, including processing power and memory, can influence the quantity of training materials that can be practically used. More resources may enable the training of larger models with more data.
- **Overfitting Considerations:**
  - Developers must be cautious about overfitting, where a model becomes too specialized to the training data and performs poorly on new, unseen data. Balancing the quantity of data to avoid overfitting is a critical consideration.
- **Domain and Application:**
  - The specific domain and application of the AI model can impact the dataset requirements. Some domains may have limited data available, while others have access to extensive datasets.
- **Transfer Learning:**
  - In some cases, developers may use pre-trained models as a starting point and then fine-tune them on a smaller dataset specific to their task. This approach can reduce the quantity of training materials required.

It's important to emphasize that there is no fixed quantity of training materials that universally applies to all generative AI models. Developers typically perform data analysis and experimentation to determine the optimal dataset size for their specific use case. The goal is to strike a balance between having enough data to train a robust model and avoiding data overfitting. This process often involves iterative testing and evaluation.



In many cases, developers seek to collect or curate the largest and most diverse dataset possible to improve the model's performance, but practical constraints, including data availability and resource limitations, play a role in determining the actual quantity used for training.

Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

The volume of material used to train an AI model can affect the fair use analysis, particularly under the second fair use factor, which considers the "nature of the copyrighted work." Here's how the volume of material may influence the fair use analysis:

- **Amount and Substantiality of the Use:**
  - The fair use analysis evaluates the quantity of copyrighted material used in relation to the whole work. Using a large volume of copyrighted material, especially if it constitutes a substantial portion of the original work, can weigh against fair use.
  - Courts often assess whether the amount used is necessary for the intended transformative purpose. Using excessive material beyond what is required may weaken the fair use argument.
- **Nature of the Copyrighted Work:**
  - The nature of the copyrighted work factor can also be influenced by the volume of material. If a substantial volume of copyrighted material used for training is highly creative or unpublished, it may raise more copyright concerns than using smaller portions of factual or publicly available works.
- **Market Effect:**
  - The volume of material can indirectly impact the market effect analysis. If a large volume of copyrighted material is used in AI model training, and this results in

competition with or substitution for the market for the original work, it may weigh against fair use.

- **Transformative Use:**
  - The volume of material used in a transformative manner can affect the fair use analysis positively. If a large volume of material is necessary to achieve the transformative purpose, this may strengthen the fair use argument.

It's important to note that while the volume of material is a relevant factor, fair use determinations are multifaceted and consider multiple factors, including the purpose and character of the use, the nature of the copyrighted work, the effect on the market, and the overall context. The quantity of training material alone does not determine fair use. Developers and courts must consider the interplay of all relevant factors when evaluating whether a particular use qualifies as fair use.

Ultimately, the fair use analysis remains highly case-specific, and the outcome depends on the unique circumstances of each case. Legal advice from experts in copyright law is crucial when navigating fair use considerations in AI model development, especially when dealing with large volumes of copyrighted material.

8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? <sup>[46]</sup> Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

#### *CHAT-GPT*

Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured?

Under the fourth factor of the fair use analysis, which assesses the effect on the potential market for or value of a copyrighted work used to train an AI model, the measurement of this effect involves a complex evaluation. Courts typically consider various aspects and evidence to determine whether the use in question adversely impacts the market or value of the copyrighted work. Here's how this measurement can be approached:

- **Direct Market Harm:**
  - Courts assess whether the use competes with or directly substitutes for the market of the original work. If the AI model's use significantly diminishes the market for the copyrighted work, it may weigh against fair use.
  - Evidence of direct market harm could include a decline in sales, licensing opportunities, or demand for the original work due to the use of AI-generated content.
- **Derivative Works:**
  - Courts consider whether the AI model's use results in the creation of derivative works or secondary markets that adversely affect the original work. If the AI-generated content serves as a substitute for derivative works based on the copyrighted material, it may be viewed negatively.
  - Evidence of potential harm may include a decrease in demand for adaptations, spin-offs, or derivative products related to the original work.
- **Potential Future Markets:**
  - Courts may analyze whether the AI model's use impedes potential future markets or value that the copyright owner could reasonably exploit. This involves considering how the use affects the copyright owner's ability to adapt to evolving markets and technology.

- Evidence may include market research, industry trends, and expert testimony about potential future opportunities.
- **Licensing and Permissions:**
  - The availability of licensing or permission mechanisms can be relevant. If a market exists for licensing the copyrighted material for similar purposes, and the AI model's use bypasses or undermines licensing opportunities, it may be viewed negatively.
  - Evidence may include discussions or negotiations for licensing the copyrighted material.
- **Market Impact on Original Work's Value:**
  - Courts assess whether the AI model's use has an adverse impact on the overall economic value of the original work. This involves considering whether the use detracts from the original work's unique qualities or market appeal.
  - Evidence may include market studies, consumer preferences, and expert testimony regarding the work's intrinsic value.
- **Fair Market Impact:**
  - Some transformative uses may result in a fair market impact that doesn't harm the original work. Courts may consider whether the AI model's use enhances the overall market for creative and transformative works.

It's important to emphasize that the effect on the potential market for or value of a copyrighted work is one factor among several in the fair use analysis. The assessment is fact-specific and takes into account the unique circumstances of each case. The ultimate determination depends on a balancing of all fair use factors, including the purpose and character of the use, the nature of the copyrighted work, the amount used, and the overall context.

Legal experts often play a critical role in assessing the impact on the market and whether a specific use qualifies as fair use under copyright law.

Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

In the fair use analysis, when assessing the effect on the potential market for or value of a copyrighted work, the inquiry typically considers the broader market for that general class of works rather than focusing narrowly on competition with a particular copyrighted work or the works of the same author. Here's why:

- **Broad Market Consideration:**
  - Copyright law aims to protect the overall market for creative works and to incentivize the creation of new works. Therefore, the fair use analysis often takes into account how the use of copyrighted material impacts the market for similar types of works, not just a single work or author.
  - Courts are concerned with whether the use could potentially harm the economic interests of copyright owners across the category or genre of works to which the copyrighted material belongs.
- **Preventing Evasion of Copyright:**
  - Focusing solely on competition with a particular copyrighted work could allow potential infringers to evade copyright protection by arguing that their use does not directly compete with a specific work. This would undermine the intent of copyright law to protect the broader market.
- **Consideration of Substitutability:**

- Courts often assess whether the AI system's outputs, which incorporate the model, serve as substitutes for the type of works covered by the copyright. If the AI-generated content competes with or serves as a viable substitute for a category of works, it may impact the fair use analysis.
- **Market Transformation:**
  - The digital age and advancements in technology have transformed markets for creative works. Courts recognize that new technologies, such as AI, can create new markets and distribution channels. Therefore, they may consider how AI-generated content affects these evolving markets.

In summary, the fair use analysis typically evaluates whether the AI system's outputs, which incorporate a model trained on copyrighted material, have the potential to adversely impact the broader market for works of a similar class, genre, or category. It aims to strike a balance between protecting the rights of copyright owners and fostering innovation and transformative uses.

The assessment remains fact-specific and considers various factors, including the nature of the copyrighted work, the nature of the AI-generated content, the intent of the use, and the overall market dynamics. Legal experts play a key role in guiding this analysis within the context of specific cases.

9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

*CHAT-GPT*

The question of whether copyright owners should have to affirmatively consent (opt in) to the use of their works for training materials or whether they should be provided with the means to object (opt out) is a matter of legal and policy debate, and it has implications for the balance between copyright protection and the development of AI technologies. Here are considerations for both approaches:

#### **Opt-In (Affirmative Consent):**

- **Strong Copyright Protection:** The opt-in approach places a high emphasis on copyright owners' rights and control over their works. It ensures that copyrighted material cannot be used without explicit permission.
- **Clear Rights Clearance:** This approach simplifies the process of obtaining permissions and may provide more legal certainty for developers, as they would only use materials for which they have received explicit consent.
- **Protection for Unique or Sensitive Works:** It can be particularly beneficial for copyrighted works that are unique, sensitive, or require special consideration, such as unpublished manuscripts, personal data, or culturally significant materials.

#### **Opt-Out (Providing Means to Object):**

- **Fostering Innovation:** The opt-out approach is often seen as more permissive and conducive to fostering innovation, especially in AI development. It allows developers to use a wide range of materials without seeking individual permissions.

- **Practicality and Efficiency:** For large-scale AI model training, obtaining opt-in consent from every copyright owner may be impractical or even impossible, given the sheer volume of data used. The opt-out approach can be more efficient.
- **Fair Use Consideration:** In some cases, AI model training might qualify as fair use, making the opt-out approach more aligned with existing copyright law principles. Fair use allows for certain uses of copyrighted material without permission.
- **Balancing Interests:** This approach attempts to strike a balance between copyright protection and technological advancement, recognizing that strict opt-in requirements could stifle innovation.

The choice between opt-in and opt-out approaches depends on the legal and policy framework of each jurisdiction and the specific goals and values of copyright law in that context. Some jurisdictions, like the European Union, have implemented provisions like the Copyright Directive that require platforms to obtain licenses for copyrighted materials, shifting the responsibility more toward opt-in. Others, like the United States, have more permissive fair use principles, which align with an opt-out approach in certain situations.

Ultimately, the approach should consider the need to protect copyright holders' rights while also facilitating innovation and technological progress. Finding a balance that achieves both goals is a complex and evolving challenge in the digital age. Legal experts, policymakers, and stakeholders continue to debate and shape the rules and norms in this area.



9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses? [\[47\]](#)

CHAT-GPT

The requirement for consent from the copyright owner to use copyrighted works to train AI models can vary depending on the jurisdiction, legal framework, and the nature of the use. There is no one-size-fits-all answer, and different approaches exist:

- **All Uses (Including Non-Commercial):**
  - Some legal frameworks or policies may require consent for all uses of copyrighted works, regardless of whether the use is for commercial or non-commercial purposes. This approach places a strong emphasis on copyright protection and control.
- **Commercial Uses Only:**
  - Other jurisdictions or frameworks may differentiate between commercial and non-commercial uses. They may require consent primarily for commercial applications of AI models while allowing more flexibility for non-commercial, research, or educational uses.

The choice between these approaches involves a balance between copyright protection and promoting innovation and research. Here are considerations for both:

#### Requiring Consent for All Uses:

- **Strong Copyright Protection:** This approach ensures that copyright owners have control over all uses of their works, regardless of the purpose. It aligns with the traditional copyright principle of exclusive rights.
- **Protection of All Copyrighted Works:** It extends protection to all types of copyrighted works, including those that might

not have immediate commercial value but are still important to creators.

- **Clarity and Certainty:** It provides clarity for users and copyright owners regarding their rights and obligations.

#### **Requiring Consent for Commercial Uses Only:**

- **Fostering Innovation and Research:** Allowing non-commercial uses without consent can encourage research, development, and innovation, particularly in academia and the nonprofit sector.
- **Efficiency:** Requiring consent for every non-commercial use, such as research projects, may be impractical, given the volume of AI research.
- **Fair Use Consideration:** It aligns with fair use principles in some jurisdictions, which allow certain non-commercial uses of copyrighted materials without consent.

The choice between these approaches may also depend on how copyright law is evolving in response to technological advancements. Some countries have introduced specific provisions or exemptions to address AI model training, while others are adapting traditional copyright principles.

It's important to note that this is a complex and evolving legal and policy area, and the approach can vary significantly from one jurisdiction to another. Legal advice and compliance with applicable copyright laws and regulations are essential, especially when dealing with copyrighted materials in AI model development. Additionally, legal frameworks may continue to evolve to strike a balance between copyright protection and fostering innovation.

9.2. If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?<sup>[48]</sup>

CHAT\_GPT

If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training?

An "opt-out" approach in the context of using copyrighted works for training AI models would typically involve a mechanism by which copyright owners can express their objection to such use. The specific process can vary depending on legal frameworks and platforms, but here's a general outline of how it might work:

- **Notice and Request for Removal:**
  - Under an opt-out system, AI developers or platforms using copyrighted works would provide a clear and accessible process for copyright owners to object to the use of their materials.
  - Copyright owners who wish to opt out would typically submit a formal notice or request for removal. This notice should include specific information, such as the copyrighted work in question and details about the objection.
- **Designated Contact or Portal:**
  - AI developers or platforms may establish a designated contact point or an online portal for receiving opt-out requests. This contact point should be easily accessible to copyright owners.
- **Verification and Review:**
  - Upon receiving an opt-out request, AI developers or platforms would typically verify the identity and rights of the copyright owner and assess the validity of the objection.

- This process may involve evaluating whether the use falls under copyright exceptions, such as fair use or fair dealing, or other applicable legal defenses.
- **Prompt Response and Removal:**
  - If the objection is deemed valid, AI developers or platforms would promptly remove or cease using the copyrighted material in question.
  - The timeframe for removal would depend on the specific platform's policies or legal requirements.
- **Communication and Records:**
  - AI developers or platforms should maintain clear communication with the copyright owner throughout the opt-out process and keep records of the objection and resolution.
  - This helps ensure transparency and compliance with legal obligations.
- **Appeals Process (Optional):**
  - Some systems may include an appeals process for situations where AI developers or platforms dispute the validity of the opt-out request. This process would involve a review by a neutral party or authority.
- **Public Documentation:**
  - AI developers or platforms might maintain a public record of copyrighted materials for which opt-out requests have been honored. This record can help prevent repeated use of the same material.

It's important to note that the specifics of an opt-out system, including the processes and mechanisms, can vary widely depending on jurisdiction-specific laws, industry standards, and platform policies. In some cases, legislative or regulatory measures may be introduced to formalize the opt-out process and establish clear legal obligations.

AI developers and platforms should be proactive in complying with any applicable opt-out requests and should have mechanisms in place to address objections from copyright owners while ensuring transparency, fairness, and legal compliance. Legal advice and consultation with experts in copyright law can be crucial in implementing effective opt-out procedures.

Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?

1 / 1

Yes, there are technical tools and mechanisms that can facilitate the opt-out process and help copyright owners indicate their objection to the use of their works for AI training purposes. These tools often involve metadata or identifiers that convey permissions or restrictions related to copyrighted materials. Here are a few examples:

- **Rights Management Metadata:** Metadata standards like RightsML or ODRL (Open Digital Rights Language) can be used to embed machine-readable information about the usage rights and permissions associated with a specific work. Copyright owners can use these standards to specify their objections to certain uses, including AI training.
- **Digital Watermarks:** Digital watermarks can be embedded within multimedia files (images, audio, video) to carry information about copyright and usage restrictions. Watermarks can include instructions that specify the conditions under which a work can or cannot be used for AI training.
- **Machine-Readable Licensing:** Copyright owners can provide machine-readable licenses using standardized

formats like Creative Commons licenses. These licenses can indicate the terms under which the work can be used and whether AI training is allowed or prohibited.

- **Robots.txt and Meta Robots Tags:** For web content, website owners can use the robots.txt file and meta robots tags in web pages to communicate with web crawlers and automated services. They can specify which content should not be indexed or used for AI training purposes.
- **APIs and Permissions Systems:** Some platforms and content management systems offer APIs (Application Programming Interfaces) that allow copyright owners to set permissions and restrictions on the use of their content. Developers can integrate these APIs to respect copyright owner preferences.
- **Content ID and Fingerprinting:** In some cases, content identification and fingerprinting systems are used to identify and manage copyrighted content. Copyright owners can use these systems to mark their content and specify usage policies.
- **Blockchain and Smart Contracts:** Emerging technologies like blockchain can be used to create and enforce smart contracts that specify usage rights and restrictions for digital content, including AI training.

It's important to note that the effectiveness of these technical tools may depend on widespread adoption and interoperability with AI training systems and platforms. Additionally, legal frameworks may play a role in recognizing and enforcing the preferences indicated by copyright owners through these tools.

Copyright owners interested in using these technical tools to indicate their objections should also be aware of the legal requirements and standards that apply in their jurisdiction, as well as industry best practices for communicating usage preferences to AI developers and platforms.

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

*CHAT-GPT*

Establishing and using a process that allows copyright owners to indicate their objections to the use of their works for AI training purposes, while technically feasible, may encounter various legal, technical, and practical obstacles. Here are some of the key challenges:

#### **Legal Obstacles:**

- **Copyright Complexity:** Copyright law varies by jurisdiction and can be complex. Determining the appropriate legal framework for enabling such objections and ensuring compliance with regional copyright laws can be challenging.
- **Standardization:** Achieving global standardization for metadata, watermarks, or licensing terms that indicate objections is difficult. Harmonizing these standards across different platforms and jurisdictions is a significant challenge.
- **Enforceability:** Even if technical tools are used to indicate objections, enforcing these objections across the internet and within AI training datasets can be difficult, particularly if data has been widely distributed.
- **Privacy and Data Protection:** The use of technical tools to indicate objections must also consider privacy and data protection laws. Some metadata or watermarking methods may inadvertently reveal sensitive information about the copyright owner.

#### **Technical Obstacles:**

- **Metadata Adoption:** Widespread adoption of metadata standards or watermarks by copyright owners and AI



developers may take time, and not all content may be adequately tagged.

- **Data Collection:** AI training often involves collecting data from a wide range of sources, including the internet. Identifying and respecting objections within such diverse datasets can be challenging.
- **Interoperability:** Ensuring that technical tools used to indicate objections are compatible with various AI training platforms, libraries, and tools requires standardization and collaboration among tech providers.
- **False Positives and Negatives:** Automated systems that identify objections may produce false positives (flagging content that doesn't object) or false negatives (missing content that objects), requiring manual review.

#### **Practical Obstacles:**

- **Awareness and Education:** Copyright owners, AI developers, and platform operators need to be aware of the existence and use of objection mechanisms. Education and awareness campaigns may be necessary.
- **Content Identification:** Copyright owners may find it challenging to identify all instances of their content, particularly if it has been widely disseminated on the internet.
- **Resource Intensity:** Managing objections and compliance can be resource-intensive for both copyright owners and AI developers, especially if there are numerous objections to process.
- **Legal Disputes:** Objections can lead to legal disputes about the validity of copyright ownership, licensing terms, and fair use, which can be time-consuming and costly.
- **International Considerations:** Addressing objections globally involves navigating differing copyright laws, languages, and cultural sensitivities.



- **Platform Policies:** Platforms and AI development environments may have their own policies and mechanisms for addressing objections, adding complexity to the process. Overall, while establishing a process to address copyright objections in AI training is technically possible, it requires careful consideration of legal, technical, and practical challenges. Effective solutions may involve a combination of legal reforms, standardized technical tools, education and awareness efforts, and cooperation among copyright owners, AI developers, and platform operators. Legal advice and compliance with regional copyright laws are crucial when implementing such processes.

Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

Obtaining advance consent from copyright owners for the use of their works in AI model training, especially when dealing with a large volume of works, can be extremely challenging and, in many cases, impractical. Several factors contribute to this challenge:

- **Scale of Data:** AI models, particularly large and complex ones, often require vast amounts of training data. Collecting advance consent from each copyright owner for every individual work within a dataset of this scale can be logistically complex and time-consuming.
- **Data Collection from the Internet:** AI training datasets often include content collected from the internet, where it can be challenging to identify and contact copyright owners for consent, especially if content has been widely shared.
- **Orphan Works:** Many works used in AI training may fall into the category of orphan works, where the copyright owner is unknown or untraceable. Obtaining consent for orphan works is virtually impossible.

- **Variety of Works:** AI models benefit from diverse training data, including various types of content. Seeking consent for a wide range of works, from text and images to audio and video, can be exceptionally complex.
- **Timeliness:** AI research and development often require quick access to data. Waiting for individual consents can slow down research and innovation significantly.
- **Data Updates:** Training datasets may require frequent updates and additions, making it impractical to obtain continuous consent.
- **Fair Use and Public Domain:** Some training data may be considered fair use or in the public domain, where consent is not required. Determining the copyright status of every work in advance is resource-intensive.

Due to these challenges, AI developers often rely on a combination of strategies, including:

- **Using Publicly Available Data:** Utilizing datasets that consist of publicly available and openly licensed content for which advance consent is not needed.
- **Data Curation and Filtering:** Employing data curation techniques to remove copyrighted material for which consent cannot be obtained.
- **Fair Use and Fair Dealing:** Complying with fair use or fair dealing provisions in copyright law where applicable.
- **Licensing Agreements:** Entering into licensing agreements with content providers or aggregators who have obtained the necessary permissions.
- **Alternative Data Sources:** Exploring alternative sources of data, such as user-generated content, creative commons material, and data that falls under exceptions and limitations in copyright law.

It's important to acknowledge that the balance between copyright compliance and technological advancement is a complex issue.

Legal frameworks and industry practices are still evolving to address the challenges associated with using copyrighted material in AI training. Developers should seek legal guidance and strive to comply with copyright laws and standards while considering the practical constraints of AI model development.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

*CHAT-GPT*

If an objection is not honored, what remedies should be available?

The remedies available when an objection to the use of copyrighted works for AI training is not honored can vary depending on legal frameworks and jurisdiction-specific laws. Here are some common remedies that may be available:

- **Cease and Desist Letters:** Copyright owners can send cease and desist letters to the AI developer, platform, or entity using their copyrighted material without consent. These letters demand the immediate cessation of the infringing use and may outline potential legal consequences if the use continues.
- **Injunctions:** Copyright owners may seek injunctive relief through the legal system, requesting a court order to stop the infringing use of their works in AI training.
- **Takedown Notices:** In some jurisdictions, copyright owners can issue takedown notices under the Digital Millennium Copyright Act (DMCA) in the United States or similar laws in other countries. Platform operators are required to remove infringing content in response to valid notices.

- **Damages:** If copyright owners can demonstrate that their copyrighted material was used without consent and suffered financial harm as a result, they may seek damages in the form of monetary compensation through litigation.
- **Account Suspension or Content Removal:** Platform operators and AI developers may take action to suspend user accounts, remove content, or disable access to infringing materials in response to valid objections and legal requests.
- **Legal Action:** Copyright owners have the option to file lawsuits against infringing parties, seeking remedies such as damages, injunctions, and court-ordered removal of infringing content.
- **Alternative Dispute Resolution:** In some cases, parties may choose to engage in alternative dispute resolution methods, such as mediation or arbitration, to resolve copyright disputes without going to court.
- **Statutory Damages:** In some jurisdictions, copyright laws provide for statutory damages, which are predetermined amounts of damages that may be awarded to copyright owners without the need to prove actual harm.
- **Public Relations and Reputational Impact:** Copyright owners may choose to engage in public relations efforts to raise awareness of copyright infringement issues, potentially causing reputational damage to infringing parties.

The specific remedies available and the process for pursuing them can vary significantly depending on the applicable copyright laws and the legal systems of individual countries. It's essential for both copyright owners and AI developers to seek legal advice and guidance when copyright disputes arise. Engaging in good-faith negotiations and dispute resolution efforts can sometimes lead to mutually acceptable resolutions without resorting to formal legal action.

Are existing remedies for infringement appropriate or should there be a separate cause of action?

The appropriateness of existing remedies for copyright infringement in the context of AI training can be a subject of debate and may vary depending on the specific circumstances and legal frameworks of different jurisdictions. Whether a separate cause of action is necessary depends on various factors, including the extent of copyright infringement concerns related to AI training. Here are some considerations:

### Existing Remedies for Copyright Infringement:

- **Adaptability:** Existing copyright laws and remedies have been developed over time to address a wide range of copyright infringement scenarios, including traditional forms of content reproduction and distribution. These remedies can be applied to address copyright concerns in the context of AI training.
- **Case-by-Case Assessment:** Copyright infringement cases related to AI training are often assessed on a case-by-case basis, taking into account factors such as the nature of the copyrighted material, the extent of use, and the legal principles of fair use or fair dealing where applicable.
- **Legal Precedent:** Existing copyright laws and remedies are supported by legal precedent and established case law, providing guidance for both copyright owners and users of copyrighted material.

### Considerations for a Separate Cause of Action:

- **Unique Challenges:** AI training presents unique challenges and complexities related to the volume of data used, automated processes, and the transformation of copyrighted content. These challenges may warrant specialized legal provisions.

- **Clarification:** A separate cause of action or specific provisions related to AI training could provide greater clarity and legal certainty for all parties involved, helping to address concerns related to AI's transformative capabilities.
- **Balancing Interests:** Specialized provisions could help strike a balance between protecting copyright holders' rights and fostering technological innovation by addressing AI-specific considerations.
- **International Coordination:** Given the global nature of AI development, harmonized international provisions could facilitate consistency in addressing copyright concerns related to AI training.
- **Transparency and Compliance:** A separate cause of action could emphasize the importance of transparency and compliance with copyright laws in AI model development and data acquisition processes.

Whether a separate cause of action is deemed necessary or appropriate would depend on legislative and policy decisions made by individual jurisdictions. Policymakers, legal experts, and stakeholders in the AI and copyright domains play crucial roles in shaping the legal framework and determining whether specialized provisions are needed to address copyright concerns in AI training effectively.

Ultimately, the legal landscape is evolving in response to advancements in technology, including AI, and ongoing discussions will continue to shape the approach to addressing copyright infringement in this context. Legal frameworks must strike a balance between protecting intellectual property rights and enabling innovation.

9.5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?

*CHAT-GPT*

Whether a human creator who does not own the copyright but was involved in the creation of a work should have the right to object to an AI model being trained on their work depends on the specific terms of their contractual agreement, the applicable copyright laws, and the legal relationship between the creator, the copyright holder, and any involved parties. Here are some considerations:

- **Contractual Agreements:** The rights and obligations of creators who do not own the copyright are often determined by contractual agreements. In many cases, creators may have signed contracts that stipulate the rights and compensation associated with their creative contributions. These contracts can include clauses that address the use of their work in various contexts, including AI training.
- **Work Made for Hire:** If the work was created under a "work made for hire" agreement, the copyright typically belongs to the employer or the hiring party, not the creator. In such cases, the creator may have limited rights to object to the use of the work, as they are not the copyright holder.
- **Moral Rights:** Some jurisdictions recognize moral rights, which are distinct from copyright and provide creators with certain non-economic rights related to the integrity and attribution of their work. Depending on local laws, creators may have the right to object to uses that could damage their reputation or involve modifications to their work that they find objectionable.
- **Assignment and Licensing:** Creators who have assigned their copyright or granted licenses to others may



have relinquished their rights to object to certain uses of their work. The terms of these agreements are crucial in determining the extent of their rights.

- **Collaborative Works:** In cases where multiple contributors have worked on a project, the rights and objections of individual creators may be influenced by the nature of their contributions, their contractual agreements, and the specific circumstances of the collaboration.
- **Legal Framework:** Copyright laws and regulations vary by jurisdiction and can impact the rights of creators who do not own the copyright. Some countries provide more extensive moral rights protections, while others prioritize the rights of the copyright holder.

The ability of a creator to object to the use of their work in AI training depends on the interplay of these factors. In many cases, creators who have transferred their copyright or worked under "work made for hire" agreements may have limited rights to object. However, they may still have moral rights or contractual rights that allow them to assert objections under certain conditions.

To navigate these complex issues, it's essential for creators, copyright holders, and AI developers to review and understand the terms of their agreements, seek legal advice when necessary, and comply with applicable copyright laws and regulations in their respective jurisdictions.



If so, how would such a system work?

A system that allows creators who do not own the copyright to object to the use of their work in AI training would need to consider a range of legal, technical, and practical aspects. The specifics of such a system can vary, but here's a general outline of how it might work:

### **1. Clear Contractual Agreements:**

- Creators involved in collaborative projects or commissioned work should have clear contractual agreements outlining their rights, compensation, and any restrictions on the use of their contributions. These contracts should include provisions related to AI training.

### **2. Rights Acknowledgment:**

- AI developers and copyright holders should acknowledge the rights and contributions of creators who do not own the copyright, as specified in their contractual agreements.

### **3. Objection Mechanism:**

- Creators should have a mechanism to formally raise objections if they believe that the use of their work in AI training violates their contractual rights or moral rights.

### **4. Dispute Resolution:**

- In cases where objections are raised, a dispute resolution process may be necessary. This process could involve negotiation, mediation, or, if necessary, legal action to determine the validity of the objection.

### **5. Record-Keeping:**

- All parties involved, including AI developers and copyright holders, should maintain records of contractual agreements and objections raised. This ensures transparency and accountability.

## **6. Compliance and Remedies:**

- If an objection is deemed valid, the AI developer and copyright holder should take appropriate actions, such as ceasing the use of the work in AI training or providing compensation as specified in the contract.

## **7. Legal Framework:**

- The system should operate within the framework of existing copyright laws and regulations. Some jurisdictions may have specific provisions related to moral rights or contractual rights of creators.

## **8. Education and Awareness:**

- Creators, copyright holders, and AI developers should be educated about their rights and responsibilities within this system. Clear guidelines and best practices can help prevent disputes.

## **9. International Considerations:**

- For cross-border collaborations and AI development projects, the system should consider the complexities of international copyright law and agreements.

## **10. Technical Tools:**

- Technical tools, such as metadata or watermarking, may be used to indicate the rights and objections of creators. These tools can help identify and respect objections during AI training.

The effectiveness of such a system relies on clear, enforceable contractual agreements, cooperation among all parties involved, and adherence to legal requirements. It also requires an understanding of the legal and ethical considerations related to copyright, moral rights, and the use of AI in content creation.

In practice, the specifics of this system would likely vary by industry, jurisdiction, and individual projects. Legal guidance is crucial for creators, copyright holders, and AI developers to navigate these complexities and ensure compliance with applicable laws and agreements.

10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?

*CHAT-GPT*

Obtaining licenses to use copyrighted works in training generative AI models involves a legal and contractual process. Here are steps that can be taken to secure licenses:

### **1. Identify Copyright Owners:**

- Determine the copyright owners of the works you intend to use. This may involve tracking down individual creators, publishers, or rights-holding entities.

### **2. Contact Copyright Owners:**

- Reach out to copyright owners to express your interest in obtaining a license for the specific use of their works in AI training. Provide clear and detailed information about the intended use.

### **3. Negotiate License Terms:**

- Negotiate the terms of the license, including the scope of use, duration, compensation, and any restrictions. Be prepared for discussions on pricing, usage rights, and other relevant terms.

#### **4. Draft a License Agreement:**

- Once terms are agreed upon, draft a formal license agreement that outlines all the negotiated terms, responsibilities, and obligations of both parties. Seek legal advice to ensure the agreement is legally sound.

#### **5. Obtain Written Consent:**

- Secure written consent or an executed copy of the license agreement from the copyright owner. This provides a legally binding record of the agreement.

#### **6. Compliance with Legal Requirements:**

- Ensure that the license agreement complies with all relevant copyright laws and regulations in your jurisdiction. Legal advice is essential to address any legal intricacies.

#### **7. Payment and Royalties:**

- If compensation is part of the agreement, arrange for payment according to the agreed-upon terms, which may involve one-time payments, ongoing royalties, or other arrangements.

#### **8. Record Keeping:**

- Maintain accurate records of all license agreements, communications with copyright owners, and payments made. Good record-keeping is essential for compliance and dispute resolution.

#### **9. Adhere to License Terms:**

- Strictly adhere to the terms and conditions specified in the license agreement. Any deviation may result in a breach of contract and legal consequences.

#### **10. Renew or Extend Licenses (if necessary):**

- If the AI training project extends beyond the original license period, negotiate renewals or extensions with copyright owners, if required.

#### **11. Monitor Usage and Compliance:**

- Continuously monitor and ensure compliance with the terms of the license agreement throughout the AI training process.

#### **12. Seek Legal Advice:**

- At every stage of the licensing process, consult with legal experts who specialize in copyright law to ensure full compliance and protection of rights.

It's important to note that the licensing process can be complex, time-consuming, and costly, especially when dealing with a large volume of copyrighted works. Additionally, the willingness of copyright owners to grant licenses can vary, and not all copyright owners may be open to licensing their works for AI training.

In cases where obtaining licenses is impractical or impossible, AI developers may consider alternative data sources, such as public domain works, openly licensed materials, or creating their own datasets. Additionally, understanding and complying with fair use or fair dealing provisions, where applicable, can be important for AI model development.

### 10.1. Is direct voluntary licensing feasible in some or all creative sectors?

CHAT-GPT

Direct voluntary licensing is feasible in many creative sectors and is a common practice for obtaining permission to use copyrighted works. The feasibility of such licensing arrangements can vary depending on several factors, including the nature of the creative work, the copyright owner's willingness to license, and the intended use. Here are some creative sectors where direct voluntary licensing is often feasible:

- **Music:** The music industry frequently involves direct voluntary licensing, where musicians, composers, and music publishers grant licenses for the use of their songs in various contexts, such as film, television, advertising, and digital platforms.
- **Publishing:** Authors, publishers, and literary agents often negotiate direct licenses for the use of written works, including books, articles, and written content for educational materials.
- **Visual Arts:** Visual artists, photographers, and illustrators may grant licenses for the reproduction and use of their artworks in print, digital media, and advertising campaigns.
- **Film and Television:** Producers and filmmakers negotiate licenses for the use of copyrighted material, such as music, scripts, and visuals, in the production of films and television shows.
- **Software:** Software developers frequently license their software products to end-users under specific terms and conditions, often through end-user license agreements (EULAs).

- **Photography:** Photographers and stock photo agencies offer licenses for the use of their photographs in various applications, from advertising to editorial use.
- **Video Games:** Developers and publishers obtain licenses for music, characters, and other copyrighted content used in video games.
- **Advertising and Marketing:** Advertisers often seek licenses for music, images, and other creative content to use in advertising campaigns.

In these sectors, the licensing process typically involves negotiation, agreement on terms, and the payment of licensing fees or royalties. Digital technology has made it easier to facilitate and track licensing agreements, contributing to the feasibility of direct voluntary licensing.

However, it's essential to recognize that not all copyright owners may be willing to grant licenses, especially in cases where the use could potentially harm their interests or when the work has exclusive distribution agreements. Additionally, the terms of licensing agreements can vary widely, so negotiation skills and legal expertise are often necessary to navigate these arrangements effectively.

In some cases, collective licensing organizations, such as performing rights organizations (PROs) for music, simplify the licensing process by collectively representing copyright owners and licensing their works to users. These organizations play a critical role in facilitating licensing in certain creative sectors.

Overall, while direct voluntary licensing is feasible in many creative sectors, the specific terms and conditions of licensing agreements can vary significantly based on the industry, the parties involved, and the intended use of the copyrighted works.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? <sup>[49]</sup> Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

CHAT-GPT

Is a voluntary collective licensing scheme a feasible or desirable approach?

A voluntary collective licensing scheme can be both feasible and desirable in certain contexts, particularly for managing copyright permissions and licensing in cases involving large volumes of copyrighted works or when dealing with complex licensing scenarios. Here are considerations for its feasibility and desirability:

### **Feasibility:**

- **Complex Licensing Scenarios:** In situations where obtaining individual licenses for each copyrighted work is impractical due to the sheer volume of works or the diversity of copyright owners, a collective licensing scheme can simplify the process.
- **Administrative Efficiency:** Collective licensing organizations can streamline the administrative burden associated with negotiating and managing licenses for multiple works, making it more feasible for both copyright owners and users.
- **Standardized Terms:** Collective licensing often involves standardized licensing terms and fees, which can simplify negotiations and reduce the need for extensive individual agreements.
- **Economies of Scale:** By pooling together the rights of multiple copyright owners, collective licensing organizations



can negotiate more favorable terms and conditions, potentially benefiting both copyright owners and users.

### **Desirability:**

- **Convenience:** A collective licensing scheme can provide convenience for users who need access to a broad range of copyrighted works, such as digital platforms, broadcasters, and educational institutions.
- **Consistency:** Standardized licensing terms can promote consistency and transparency in licensing agreements, reducing potential disputes and litigation.
- **Revenue Distribution:** Collective licensing organizations often have established mechanisms for distributing royalties to copyright owners, ensuring that creators receive compensation for the use of their works.
- **Promoting Legal Use:** By providing a straightforward path to obtain licenses, collective licensing can encourage legal and compliant use of copyrighted materials, reducing the risk of copyright infringement.
- **Promoting Access:** In some cases, collective licensing can expand access to copyrighted works for educational, research, or cultural purposes, fostering creativity and innovation.

However, it's important to acknowledge that collective licensing schemes may not be suitable for all situations, and there are potential challenges and considerations:

- **Consent:** Participation in a collective licensing scheme is typically voluntary for copyright owners. Not all copyright owners may choose to join, which can limit the effectiveness of such schemes.
- **Market Dynamics:** Collective licensing organizations may hold significant market power, and negotiations with users may not always result in terms favorable to all parties.

- **Transparency:** Ensuring transparency in revenue distribution and the allocation of royalties to individual creators can be a complex process that requires robust tracking and reporting mechanisms.
- **Legal and Regulatory Considerations:** The legal and regulatory framework in a given jurisdiction may influence the feasibility and desirability of collective licensing schemes.
- **Industry-Specific Variations:** The appropriateness of collective licensing can vary by industry, with some sectors, such as music and broadcasting, being more established in this regard.

In summary, a voluntary collective licensing scheme can offer benefits in terms of simplifying copyright licensing, promoting access to creative works, and ensuring fair compensation for creators. However, the feasibility and desirability of such a scheme depend on the specific context, industry, and the willingness of copyright owners and users to participate. Effective governance, transparency, and adherence to legal and ethical principles are crucial for the success of collective licensing efforts.

Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role?

Yes, there are existing collective management organizations (CMOs) that are well-suited to provide licenses for the use of copyrighted works in various contexts. These CMOs are established in different creative sectors and play a key role in managing rights, collecting royalties, and facilitating licensing agreements. Some well-known examples include:

- **Music:**
  - Performing Rights Organizations (PROs) like ASCAP, BMI, and SESAC in the United States.

- Phonographic Performance Limited (PPL) and PRS for Music in the United Kingdom.
- **Visual Arts:**
  - Visual Arts Rights Societies (VARs) that manage the rights of visual artists and photographers.
- **Literature and Text:**
  - Authors' societies and literary agencies that represent authors and publishers in licensing written works.
- **Audiovisual:**
  - CMOs representing filmmakers, screenwriters, and composers in the audiovisual sector.
- **Software:**
  - Organizations that manage software licensing and royalties for software developers.

These CMOs are equipped to manage licenses, collect royalties, and distribute payments to copyright owners based on standardized or negotiated terms. They have experience in navigating licensing agreements and can provide valuable services to both copyright owners and users of copyrighted works.

However, there are legal and operational considerations that may affect the ability of CMOs to perform this role effectively:

### **Legal Impediments:**

- **Jurisdiction-Specific Laws:** Copyright laws and regulations vary by jurisdiction, and the ability of CMOs to manage licenses may be subject to compliance with local legal requirements.
- **Competition Laws:** In some cases, competition laws and antitrust regulations may impose limitations on the market power and activities of CMOs.

- **Conflict of Interest:** CMOs must navigate potential conflicts of interest between copyright owners and licensees to ensure fair and transparent licensing.

#### Operational Challenges:

- **Technological Infrastructure:** Effective licensing and royalty distribution require robust technological systems to track usage and manage payments accurately.
- **Transparency:** Ensuring transparency in royalty distribution and reporting is crucial to maintaining trust among copyright owners and licensees.
- **Broad Representation:** CMOs need to have a broad and diverse membership to effectively represent the interests of copyright owners.
- **Education and Outreach:** Effective communication and education efforts are necessary to inform copyright owners and licensees about licensing options and benefits.

In summary, while existing CMOs are well-suited to provide licenses for copyrighted works, the effectiveness of their role may be influenced by legal, operational, and industry-specific factors. Collaboration between CMOs, copyright owners, and users, as well as compliance with legal requirements, is essential to overcome potential impediments and ensure fair and efficient licensing processes.

Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

Whether Congress should consider statutory or other changes, including potential antitrust exceptions, to facilitate the negotiation of collective licenses is a complex question that involves a careful balancing of various interests, including those of copyright owners, users of copyrighted material, and competition in the market. Here are some considerations:

## Potential Benefits:

- **Streamlined Licensing:** Statutory changes could simplify and streamline the licensing process, making it easier for users to obtain licenses for a wide range of copyrighted works, potentially fostering innovation and creativity.
- **Efficient Royalty Distribution:** Collective licensing can help efficiently collect and distribute royalties to copyright owners, ensuring that creators are fairly compensated for the use of their works.
- **Consistency:** Standardized licensing terms provided by collective management organizations (CMOs) can promote consistency and transparency in licensing agreements, reducing disputes.
- **Market Negotiation Power:** Collective licensing can provide a mechanism for smaller copyright owners to collectively negotiate terms with large users, potentially balancing negotiation power.

## Potential Concerns and Challenges:

- **Market Dominance:** Some CMOs may hold significant market power and could potentially abuse that power in negotiations, leading to concerns about anti-competitive behavior.
- **Lack of Competition:** An antitrust exception or other changes that overly favor CMOs could reduce competition in the licensing market, potentially limiting choices for users and copyright owners.
- **Transparency and Accountability:** Effective governance, transparency, and accountability mechanisms would be necessary to address concerns related to CMO operations and royalty distribution.

- **International Considerations:** Changes to collective licensing in the United States may have implications for international copyright agreements and treaties.
- **Industry-Specific Variation:** The appropriateness of collective licensing and any potential changes may vary by creative sector, so a one-size-fits-all approach may not be suitable.
- **Balance of Interests:** Policymakers must carefully consider the balance between promoting innovation and protecting the rights of creators and copyright owners.

Given these considerations, any legislative or regulatory changes should be approached with caution and a thorough understanding of the potential implications. It may be beneficial to conduct comprehensive studies, engage stakeholders, and seek expert input to develop balanced and effective solutions that address the unique challenges and opportunities in different creative sectors.

In cases where statutory changes are contemplated, it's essential to strike a balance between promoting efficient licensing and ensuring competition and fairness in the marketplace. Careful drafting of legislation, oversight mechanisms, and ongoing evaluation of the impact of any changes are critical elements of such an approach.

10.3. Should Congress consider establishing a compulsory licensing regime? [\[50\]](#) If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?

*CHAT-GPT*

Should Congress consider establishing a compulsory licensing regime

The consideration of establishing a compulsory licensing regime is a complex policy matter that involves weighing the interests of copyright owners, users, and the broader public. Compulsory licensing would require copyright owners to license their works under specific terms and conditions set by the government, regardless of their individual preferences. Here are some key considerations:

### **Potential Benefits of Compulsory Licensing:**

- **Access to Works:** Compulsory licensing can ensure broader access to copyrighted works, particularly in cases where copyright owners are unwilling to license their works or where obtaining licenses is impractical.
- **Promoting Innovation:** By facilitating access to copyrighted material, compulsory licensing could promote innovation by enabling the development of new technologies, services, and creative works.
- **Balancing Interests:** Compulsory licensing can strike a balance between the interests of copyright owners and users by providing a framework for access while ensuring fair compensation.

### **Potential Concerns and Challenges:**

- **Rights of Copyright Owners:** Compulsory licensing may infringe on the property rights of copyright owners by requiring them to license their works under terms they did not agree to.
- **Market Distortion:** It may distort the market by reducing the incentive for copyright owners to create and distribute works, potentially limiting the production of new creative content.



- **Fair Compensation:** Determining fair compensation for copyright owners can be challenging under a compulsory licensing regime, potentially leading to disputes and legal complexities.
- **Complex Implementation:** Implementing a compulsory licensing system requires careful design, oversight, and management to ensure its effectiveness and fairness.
- **International Implications:** Compulsory licensing may have implications for international copyright agreements and treaties, potentially affecting the ability of U.S. creators to protect their works abroad.
- **Sector-Specific Variation:** The appropriateness of compulsory licensing may vary by creative sector and should consider the unique characteristics of each industry.
- **Alternative Solutions:** Before resorting to compulsory licensing, alternative solutions such as collective licensing, standardized licensing terms, or fair use/fair dealing provisions should be explored.

In the United States, compulsory licensing has been established for certain uses of copyrighted works, such as for the mechanical reproduction of musical compositions. The Mechanical Licensing Collective (MLC) was created to administer these licenses. However, extending compulsory licensing to other creative sectors or broader applications would require careful consideration and thorough analysis of its potential impact.

Ultimately, whether Congress should consider establishing a compulsory licensing regime depends on the specific context, industry needs, and the balance of interests in promoting access to creative works while respecting the rights of copyright owners. Policymakers would need to engage stakeholders, conduct comprehensive studies, and weigh the pros and cons before making such a decision.



If so, what should such a regime look like?

Designing a compulsory licensing regime involves numerous legal, operational, and policy considerations. If Congress were to consider establishing such a regime, it would need to carefully craft the framework to balance the interests of copyright owners, users, and the public. Here are key elements to consider when designing a compulsory licensing regime:

### **1. Eligible Works and Users:**

- Define which types of copyrighted works and which categories of users are eligible for compulsory licensing. This could be tailored to specific industries or uses.

### **2. Clear Terms and Conditions:**

- Establish clear and standardized terms and conditions under which compulsory licenses would be granted, including pricing, royalties, and duration.

### **3. Compensation Mechanism:**

- Develop a mechanism for determining and distributing royalties to copyright owners. This should be fair and transparent, considering factors like usage, market value, and the economic interests of creators.

### **4. Opt-Out Provisions:**

- Determine whether copyright owners should have the option to opt out of compulsory licensing for their works if they wish to retain full control.

### **5. Oversight and Regulation:**

- Create a regulatory body or mechanism to oversee the implementation of compulsory licenses, administer royalty

payments, and resolve disputes between copyright owners and users.

#### **6. Compliance and Reporting:**

- Set requirements for compliance, reporting, and accountability to ensure that users of compulsory licenses adhere to the terms and conditions.

#### **7. Sunset Provisions:**

- Consider whether the regime should include sunset provisions, specifying a timeframe for the availability of compulsory licenses or periodic reviews to assess the regime's impact.

#### **8. International Considerations:**

- Address potential international implications and ensure that the regime complies with international copyright agreements and treaties.

#### **9. Public Interest:**

- Incorporate safeguards to protect the public interest and prevent abuses of compulsory licensing, including safeguards against market distortions.

#### **10. Industry-Specific Tailoring:**

- Recognize that the design of the regime may need to be tailored to the specific needs and dynamics of different creative sectors.

#### **11. Collaboration and Stakeholder Engagement:**

- Engage stakeholders, including copyright owners, users, and experts, in the development and refinement of the regime to

ensure that it addresses industry-specific challenges and balances competing interests.

## **12. Alternative Solutions:**

- Consider whether alternative solutions, such as voluntary collective licensing or standardized licensing terms, could address the goals of the regime without resorting to compulsory licenses.

## **13. Education and Outreach:**

- Implement educational efforts to inform copyright owners and users about the regime's provisions and processes.

The design of a compulsory licensing regime is a nuanced process that requires thorough legal and policy analysis, stakeholder input, and a careful consideration of the potential impact on creativity, innovation, and copyright protection. Any such regime should be implemented with a clear understanding of its goals, benefits, and consequences for all parties involved.

What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out?

The scope of a compulsory licensing regime would involve defining which activities the license covers, specifying which works are subject to the license, and considering whether copyright owners have the ability to opt out. These decisions are crucial in shaping the impact and effectiveness of the regime. Here's an overview of each aspect:

### **1. Activities Covered:**

The license could cover a range of activities, and the specific activities should be carefully defined. Some activities that a compulsory license might cover include:

- **Reproduction:** Allowing the reproduction of copyrighted works for specific purposes, such as educational use or certain types of digital distribution.
- **Distribution:** Permitting the distribution or dissemination of copyrighted works through various channels, including digital platforms or educational institutions.
- **Public Performance or Display:** Allowing the public performance or display of works, which could apply to sectors like music, film, or visual arts.
- **Derivative Works:** Enabling the creation of derivative works or adaptations based on copyrighted material.

The choice of activities covered would depend on the goals of the regime and the specific needs of the creative sectors it aims to address.

## **2. Works Subject to the License:**

Determining which works are subject to the compulsory license is another key consideration. The regime could encompass various types of copyrighted works, such as:

- Musical compositions and sound recordings.
- Literary works, including books, articles, and educational materials.
- Visual arts, including paintings, photographs, and illustrations.
- Audiovisual works, including films, videos, and television programs.
- Software, including code and applications.

The inclusion or exclusion of specific categories of works should be guided by the objectives of the regime and the industries it aims to support.

### 3. Copyright Owners' Ability to Opt Out:

Whether copyright owners should have the ability to opt out of the compulsory licensing regime is a policy decision with implications for both copyright owners and users. Here are two possible scenarios:

- **Opt-In Model:** Copyright owners would have the option to voluntarily participate in the compulsory licensing regime. Those who choose to participate would subject their works to the terms and conditions of the license, including compensation provisions.
- **Opt-Out Model:** Copyright owners would automatically be subject to the compulsory license unless they actively opt out. Those who choose to opt out would retain full control over their works and would not be subject to the license's terms and conditions.

The choice between these models would depend on policy goals, legal considerations, and the desire to balance the interests of copyright owners and users. An opt-in model might provide copyright owners with more control but could limit the effectiveness of the regime in achieving broader access to works. An opt-out model could expand access but might be viewed as more intrusive on copyright owners' rights.

Ultimately, the specific design of the compulsory licensing regime would require careful consideration, taking into account industry needs, legal principles, and the balance of interests between creators and users of copyrighted material. Stakeholder

engagement and expert input would be essential in making informed decisions about the scope and provisions of the license.

How should royalty rates and terms be set, allocated, reported and distributed?

Setting, allocating, reporting, and distributing royalty rates and terms in a compulsory licensing regime require a fair and transparent process that considers the interests of copyright owners, users, and the public. Here's an outline of how these elements could be managed:

## 1. Setting Royalty Rates and Terms:

- **Market-Based Approach:** Consider using a market-based approach to determine royalty rates. This might involve analyzing prevailing industry rates for similar uses of copyrighted works and adjusting them based on specific factors such as the nature of the use, the impact on the market, and the economic interests of creators.
- **Stakeholder Input:** Solicit input from stakeholders, including copyright owners, users, industry experts, and organizations representing creators. Public consultations and expert studies can help gather diverse perspectives and data.
- **Review and Adjustment:** Periodically review and adjust royalty rates and terms to account for changes in the market, technology, and economic conditions.

## 2. Allocation of Royalties:

- **Transparent Allocation Rules:** Develop transparent rules for allocating royalties among copyright owners. This may involve considering factors like the popularity of works, usage data, and equitable distribution principles.
- **Collective Management:** Consider using collective management organizations (CMOs) to administer and

allocate royalties. CMOs can have established mechanisms for distributing payments to copyright owners based on agreed-upon criteria.

### 3. Reporting Requirements:

- **User Reporting Obligations:** Place obligations on users of copyrighted works to report their usage, including the type, volume, and context of use. This reporting should be regular and detailed.
- **Transparency:** Ensure transparency in reporting requirements and procedures, making it clear how usage data will be collected, verified, and shared with copyright owners.

### 4. Distribution of Royalties:

- **Efficient Distribution:** Implement mechanisms for efficiently distributing royalties to copyright owners. CMOs can play a role in this process by aggregating payments and handling distribution.
- **Clear Payment Procedures:** Establish clear and accessible procedures for copyright owners to claim their royalties and resolve any disputes related to payment.
- **International Distribution:** If applicable, consider how royalties will be distributed internationally, especially when works have global reach.

### 5. Oversight and Regulation:

- **Regulatory Body:** Create an oversight body or regulatory authority responsible for monitoring compliance with reporting and payment obligations, as well as ensuring that royalties are distributed in accordance with the rules.
- **Dispute Resolution:** Develop mechanisms for resolving disputes related to royalty rates, allocation, reporting, and

distribution. Mediation and arbitration processes can be effective in this regard.

## **6. Transparency and Accountability:**

- **Public Reporting:** Promote transparency by requiring the public reporting of royalty rates, allocation criteria, and distribution processes. This can help build trust and accountability.
- **Auditing:** Allow for auditing of royalty allocation and distribution processes to verify that payments are accurate and fair.

## **7. Ongoing Evaluation:**

- **Periodic Reviews:** Conduct periodic reviews of the royalty-setting, allocation, reporting, and distribution processes to ensure that they are achieving their intended goals and are fair to all stakeholders.

The design and management of royalty rates and terms in a compulsory licensing regime require a delicate balance between providing access to creative works, ensuring fair compensation for copyright owners, and preventing market distortions. Public input, stakeholder collaboration, and expert analysis are essential in developing and maintaining a robust and equitable system. Additionally, compliance mechanisms and dispute resolution procedures should be clearly defined to address any issues that may arise in practice.



10.4. Is an extended collective licensing scheme [\[51\]](#) a feasible or desirable approach?

*CHAT-GPT*

An extended collective licensing (ECL) scheme can be both feasible and desirable in certain contexts, depending on the specific goals and needs of copyright management and licensing within a given jurisdiction or industry. ECL schemes are designed to facilitate collective licensing beyond the members of a specific collective management organization (CMO), allowing for broader access to copyrighted works. Here are some considerations for the feasibility and desirability of ECL schemes:

**Feasibility:**

- **Simplified Licensing:** ECL can simplify licensing processes, particularly in industries where it's challenging to identify and contact individual copyright owners, such as in the digital sphere.
- **Efficiency:** ECL can provide an efficient mechanism for licensing and distributing royalties when many rights holders are involved, streamlining the process for both users and copyright owners.
- **Access to Orphan Works:** ECL schemes can help address the issue of orphan works (works for which the copyright owner is unknown or cannot be located) by enabling their use under standardized terms and conditions.

**Desirability:**

- **Balancing Interests:** ECL schemes can strike a balance between the interests of copyright owners, users, and the public by providing access to works while ensuring fair compensation for creators.
- **Market Development:** ECL can encourage the development of markets for certain types of works, such as

out-of-print books or archival materials, which might not otherwise be readily available.

- **Cultural and Educational Uses:** ECL can be especially desirable for cultural and educational institutions seeking to digitize and make available large collections of works, fostering research and cultural preservation.
- **Access to Digital Resources:** In the digital age, ECL can help libraries, archives, and educational institutions provide access to digital resources while respecting copyright.

However, the desirability and effectiveness of ECL schemes can vary based on several factors:

- **Legal Framework:** The legal and regulatory framework of a given jurisdiction must support the establishment and operation of ECL schemes.
- **Industry Specificity:** The appropriateness of ECL schemes may vary by industry and should consider the unique characteristics of each sector.
- **Stakeholder Collaboration:** Effective ECL schemes require collaboration among copyright owners, CMOs, users, and regulatory authorities. Balancing the interests of these stakeholders is essential.
- **Transparency and Accountability:** Ensuring transparency and accountability in ECL operations is crucial to prevent misuse and promote fair distribution of royalties.
- **Market Impact:** The introduction of ECL schemes should be assessed for its potential impact on the market, including competition and pricing.
- **Compliance and Enforcement:** Mechanisms for compliance with ECL terms and conditions and enforcement against misuse or infringement should be established.

In summary, ECL schemes can offer practical solutions for copyright management and licensing, particularly in situations

involving large-scale digitization, cultural preservation, and educational access. However, their feasibility and desirability depend on the legal framework, industry-specific needs, stakeholder cooperation, and the balance of interests in a given jurisdiction or industry. Careful planning, transparency, and compliance mechanisms are essential to make ECL schemes effective and fair for all parties involved.

10.5. Should licensing regimes vary based on the type of work at issue?

*CHAT-GPT*

1 / 1

Yes, licensing regimes can vary based on the type of work at issue, and in many cases, it is both practical and desirable to tailor licensing approaches to the specific characteristics and needs of different types of works. Different creative works have unique attributes, market dynamics, and considerations, which can influence how they are licensed. Here are some reasons why licensing regimes may vary based on the type of work:

- **Diverse Market Dynamics:** Different types of works, such as music, literature, visual arts, software, and audiovisual content, have distinct market dynamics, pricing models, and consumer behavior. Licensing regimes should reflect these differences to optimize access and compensation.
- **Rights Management:** The rights associated with various types of works may vary. For example, music involves both composition and sound recording rights, while literary works primarily involve reproduction and distribution rights. Licensing regimes must address these differences.

- **Creative Intent:** Creators may have different intentions for their works. Some may prioritize widespread distribution, while others may seek to maintain tight control over access and use. Licensing regimes should respect and accommodate these intentions.
- **User Needs:** Users of creative works have varying requirements. Educational institutions may need licenses for textbooks, libraries may require licenses for digitizing archival materials, and digital platforms may seek licenses for streaming music or video content. Licensing should align with user needs.
- **Public Interest:** Works that have significant cultural or educational value, such as historical documents or public domain materials, may warrant special considerations to ensure broad public access.
- **Technology and Format:** The format and technological requirements of different works can impact licensing. For example, software licensing often involves complex terms, while streaming music may require performance licenses.
- **Market Practices:** Industry-specific licensing practices and norms can influence how licensing is structured. For instance, the music industry has established standards for performance and mechanical licenses.
- **International Copyright Treaties:** International agreements, such as the Berne Convention and the WIPO Copyright Treaty, may impose specific obligations and exceptions for different types of works. Licensing regimes should align with these treaty obligations.
- **Collective Management:** Some licensing is effectively managed through collective management organizations (CMOs), which may specialize in specific types of works, such as music or visual arts.
- **Emerging Technologies:** New technologies and distribution methods may require innovative licensing

models. For example, licensing for virtual reality or augmented reality content may differ from traditional audiovisual licensing.

Given these considerations, licensing regimes should be flexible and adaptable to accommodate the diversity of creative works and evolving industry practices. Legal frameworks, industry standards, stakeholder collaboration, and technology trends should inform the design of licensing approaches tailored to the characteristics of each type of work.

11. What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?

*CHAT-GPT*

What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training?

Obtaining appropriate licenses for training AI models, particularly when using copyrighted material, can be a complex process that involves various legal, technical, and practical considerations. Here are some key issues that may arise:

#### **Legal Issues:**

- **Copyright Clearance:** Determining which works are protected by copyright and obtaining the necessary permissions or licenses from copyright owners can be challenging, especially when dealing with a large and diverse dataset.
- **Licensing Terms:** Negotiating and complying with licensing terms, including usage restrictions, duration, and

fees, can be intricate, and terms may vary based on the type of work and intended use.

- **Fair Use or Fair Dealing:** Assessing whether the use of copyrighted material qualifies as fair use (in the U.S.) or fair dealing (in other jurisdictions) can be a complex legal analysis and may require consultation with legal experts.
- **Orphan Works:** Addressing orphan works (works whose copyright owners cannot be identified or located) presents legal challenges, as there may be no clear entity to grant licenses.

### Technical Issues:

- **Data Integration:** Integrating licensed data into AI training pipelines and ensuring that it is appropriately tagged and managed can be technically challenging, particularly when dealing with a wide variety of data sources and formats.
- **Data Privacy:** Ensuring compliance with data privacy regulations, such as GDPR in Europe, when using licensed datasets containing personal information can be complex and requires robust data anonymization and protection measures.
- **Data Security:** Maintaining data security and protecting licensed content from unauthorized access or breaches is a paramount concern.

### Practical Issues:

- **License Management:** Managing licenses for a vast amount of training data can be administratively demanding. Tracking and renewing licenses as needed can be a logistical challenge.
- **Costs:** Licensing fees for copyrighted material can be substantial, and budgeting for these costs is necessary.

Negotiating favorable terms and understanding the total cost of licensing is crucial.

- **Risk Assessment:** Evaluating the legal and reputational risks associated with using licensed data is essential, as legal disputes or violations of licensing agreements can have serious consequences.
- **Clearance Procedures:** Establishing clear internal procedures for obtaining, documenting, and retaining licenses is vital to ensure compliance and mitigate legal risks.
- **Contractual Compliance:** Ensuring that all parties involved in AI model training comply with licensing agreements is crucial. This may require education and training for personnel involved in the training process.

To address these issues effectively, organizations should engage legal counsel with expertise in intellectual property and data licensing, establish robust licensing and compliance procedures, and invest in technologies and systems that facilitate the management and tracking of licensed data. Additionally, staying informed about copyright laws, emerging licensing models, and industry best practices is essential in navigating the complexities of licensing for AI model training.

Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?

The responsibility for securing appropriate licenses for training AI models can vary depending on the roles and relationships among different entities involved in the AI development process. Here are considerations for different scenarios:

- **Curator of a Training Dataset:** If an entity is responsible for curating or assembling the training dataset,



that entity is typically responsible for securing licenses for the copyrighted material included in the dataset. This may involve obtaining licenses from copyright owners, ensuring compliance with licensing terms, and documenting the licensing agreements.

- **Developer Who Trains the AI Model:** The developer who trains the AI model using the curated dataset bears responsibility for complying with licensing agreements and ensuring that the use of copyrighted material aligns with the terms and conditions specified in the licenses. Developers should be aware of the licensing status of the dataset they are using.
- **Company Employing the AI Model:** The company that employs the AI model in its AI system is responsible for ensuring that the use of the AI model, which includes any underlying training data, aligns with legal and contractual obligations. This responsibility extends to both commercial and noncommercial uses of the AI system.

In practice, there may be collaboration and shared responsibility among these entities, particularly in cases where the curator of the dataset, the developer, and the company employing the AI system are separate entities. Clear contractual agreements and communication channels should be established to define roles and responsibilities related to licensing and compliance.

It's essential for all parties involved to understand and respect the licensing terms, monitor compliance, and take corrective actions if violations are identified. Additionally, organizations should consider legal counsel to navigate licensing complexities and ensure that they have appropriate policies and procedures in place to manage licensing and compliance effectively. This approach helps mitigate legal risks and ensures that the AI development process remains in compliance with copyright laws and licensing agreements.



12. Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.

*CHAT-GPT*

Identifying the degree to which a particular work contributes to a specific output from a generative AI system can be challenging and may not always be straightforward. Several factors contribute to this complexity:

- **Complex Neural Networks:** Generative AI models, such as deep learning neural networks, are complex and often opaque systems. They process vast amounts of training data, making it difficult to pinpoint the precise influence of a specific data point (e.g., a particular training work) on a given output.
- **Non-Linear Transformations:** AI models apply non-linear transformations to input data during training. As a result, the influence of individual training data points can become highly distributed and intertwined across the model's parameters, making it hard to attribute specific outcomes to specific works.
- **High-Dimensional Representations:** AI models often operate in high-dimensional spaces where input data is transformed into abstract representations. The contribution of a single data point to these representations may be challenging to measure directly.
- **Layered Processing:** AI models typically have multiple layers or stages of processing. A specific input may contribute differently at various stages, making it difficult to isolate its impact on the final output.
- **Latent Variables:** Many generative models involve latent variables that capture hidden patterns. These variables can further complicate the attribution of outputs to individual works.

- **Overfitting and Generalization:** AI models aim to generalize from their training data rather than memorize it. This means that specific data points may not have a one-to-one correspondence with model outputs.
- **Interplay of Data:** Outputs from generative AI systems often result from the interplay of numerous training data points, making it challenging to disentangle the influence of a single work.
- **Adaptability:** Some AI models are designed to adapt to input data during inference (e.g., fine-tuning). This adaptability means that their behavior may not be solely determined by their training data.

While it may be theoretically possible to design experiments to estimate the influence of specific data points on AI model outputs in controlled scenarios, this process can be resource-intensive and may not provide precise results. Moreover, it may not be feasible to apply such experiments comprehensively in real-world AI systems.

In practice, efforts to understand the role of specific works in AI-generated outputs often involve techniques like attention mechanisms or feature visualization, but these approaches provide only partial insights. Ultimately, AI model behavior is a complex emergent property influenced by its entire training dataset, architecture, parameters, and other factors, making it challenging to attribute specific contributions to individual works with a high degree of precision.

13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?

*CHAT-GPT*

The economic impacts of a licensing requirement on the development and adoption of generative AI systems would depend on several factors, including the scope of the requirement, licensing terms, industry practices, and the goals of the licensing regime. Here are some potential economic impacts to consider:

### **1. Development Costs:**

- **Increased Costs:** Licensing requirements could increase the costs of developing generative AI systems, as organizations may need to acquire licenses for training data, negotiate licensing agreements, and potentially pay royalties.
- **Administrative Overhead:** Managing licenses, compliance, and reporting could introduce administrative overhead, including legal and operational expenses.

### **2. Access to Training Data:**

- **Restricted Access:** Licensing requirements may restrict access to certain types of training data, particularly if copyright owners are unwilling to grant licenses or if licensing fees are prohibitively high. This could limit the availability of diverse and valuable datasets for AI development.
- **Impact on Research:** Licensing requirements could impact academic and research institutions by limiting their access to data for AI research, potentially hindering innovation.

### **3. Market Entry:**

- **Barriers to Entry:** Smaller or newer players in the AI industry, startups, and research organizations may face barriers to entry due to the costs and complexities associated with licensing. This could potentially reduce competition.

#### 4. Pricing Models:

- **Pricing Structure:** The pricing structure of licensing agreements could affect the affordability of AI systems. High licensing fees may lead to higher costs for AI services and products, potentially affecting consumers and businesses.

#### 5. Innovation:

- **Incentives for Innovation:** Licensing requirements could create incentives for the development of alternative AI technologies that rely on open or publicly available datasets to avoid licensing complexities.
- **Focus on Commercialization:** Companies may focus more on commercial applications of AI to recoup licensing costs, potentially reducing investment in non-commercial or research-oriented AI development.

#### 6. Copyright Owners' Revenues:

- **Increased Revenue:** Copyright owners who license their works for AI training may see increased revenue from licensing fees and royalties. This could benefit creators and content industries.
- **Varied Impact:** The impact on copyright owners' revenues would vary depending on the type and popularity of their works, the terms of licensing agreements, and the extent of AI system adoption.

#### 7. Compliance Costs:

- **Compliance Expenses:** Organizations using generative AI systems would need to allocate resources to ensure compliance with licensing agreements, potentially leading to increased operational costs.

## 8. Licensing Models:

- **Alternatives to Licensing:** Some organizations may explore alternative approaches, such as voluntary collective licensing or open data initiatives, to avoid the complexities and costs associated with individual licensing agreements.

The economic impacts of licensing requirements would be influenced by the balance between the interests of copyright owners, AI developers, users, and the broader public. Striking the right balance to promote access, innovation, and fair compensation is a complex policy challenge. Policymakers and stakeholders must carefully consider the potential economic consequences and design licensing regimes that achieve their intended goals while minimizing adverse effects on AI development and adoption.

14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.

*CHAT-GPT*

Certainly, several other factors are relevant when considering potential copyright liability for training AI models:

- **Licensing Agreements:** The existence and terms of licensing agreements for training data can significantly impact copyright liability. Properly negotiated licenses can provide legal authorization to use copyrighted material, while violations of licensing terms can lead to liability.
- **Fair Use and Fair Dealing:** The application of fair use (in the U.S.) or fair dealing (in other jurisdictions) can be a

critical factor. Courts consider factors like purpose, nature, amount, and effect on the market when determining whether a particular use qualifies as fair. The outcome of a fair use/fair dealing analysis can determine whether copyright liability exists.

- **Public Domain Works:** Works in the public domain are not protected by copyright and can be freely used for AI training without liability. However, determining whether a work is in the public domain can be complex, as copyright laws and durations vary by jurisdiction.
- **Orphan Works:** Orphan works, for which copyright owners cannot be identified or located, may pose challenges. Some jurisdictions have introduced orphan works legislation that provides limited liability protection for uses of these works.
- **Creative Commons and Open Licenses:** Works released under Creative Commons or other open licenses come with specific usage terms. Compliance with these licenses is crucial to avoid copyright infringement.
- **Attribution and Integrity Rights:** Some jurisdictions recognize moral rights or integrity rights that protect the attribution and integrity of a work. Violating these rights can lead to liability, even if the copyright itself is not infringed.
- **Transformative Use:** Courts consider whether the use of copyrighted material is transformative, meaning it adds new value or meaning to the original work. Transformative uses are more likely to be considered fair use.
- **Non-commercial Use:** In some jurisdictions, non-commercial uses of copyrighted material may be subject to different rules and exceptions, potentially affecting liability.
- **International Copyright Treaties:** Compliance with international copyright treaties, such as the Berne Convention and TRIPS Agreement, can impact liability when using copyrighted material across borders.

- **Liability for AI Developers vs. Users:** The party responsible for copyright liability may differ depending on the stage of AI development. Developers may face liability during training, while users could be liable for AI-generated content during deployment.
- **Secondary Liability:** In some cases, secondary liability principles may apply. For instance, hosting platforms may face secondary liability if they provide tools or services that enable users to infringe copyright.
- **Emerging Legal Frameworks:** Legal frameworks are still evolving in response to AI advancements. New legislation or case law may clarify liability issues in the future.
- **Customary Industry Practices:** Industry practices and norms, such as licensing agreements common in specific sectors, can influence liability expectations and standards.
- **Contractual Agreements:** Contractual agreements between parties (e.g., AI developers and clients) can specify liability arrangements and indemnification clauses.

Navigating copyright liability in the context of AI model training requires a nuanced understanding of copyright law, licensing agreements, and evolving legal interpretations. It's essential for organizations and individuals involved in AI development to seek legal advice and implement compliance measures to minimize the risk of copyright infringement and potential liability.

Tavarus Blackmon Art LLP will utilize AI-generative technology, CHAT-GPT, to address the following questions. The utilization of the technology is required where questions outside the normal operation of my organization are concerned. Further utilization of AI-generative technology to address these questions presents *Organizational Utility*, DEI accessibility, an Art Historical and Conceptual framework and a Philosophical paradigm to operate within.

## Transparency & Recordkeeping

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

### *CHAT-GPT*

In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models?

The question of whether developers of AI models should be required to collect, retain, and disclose records regarding the materials used to train their models involves a complex balance between copyright owners' interests, transparency, and practical considerations. Here are some factors to consider:

### Pros of Requiring Records:

- **Transparency:** Requiring developers to maintain records of training materials can enhance transparency. Copyright owners would have a means to determine if their works were used.
- **Accountability:** Records can help establish accountability. If copyright infringement claims arise, records may provide evidence of compliance with licensing agreements or fair use/fair dealing doctrines.
- **Legal Compliance:** Mandatory record-keeping requirements can help ensure that developers comply with



copyright laws and licensing agreements, reducing the risk of copyright infringement.

### **Cons of Requiring Records:**

- **Complexity:** Collecting and retaining records for the vast and diverse datasets used in AI training can be highly complex and resource-intensive, especially for large-scale AI models.
- **Privacy Concerns:** Some training data may contain personal or sensitive information, raising privacy concerns if developers are required to maintain detailed records. Ensuring data protection can be challenging.
- **Burden on Innovation:** Excessive record-keeping requirements could impose a burden on AI developers, potentially slowing down innovation and research efforts.
- **Practical Challenges:** Determining what should be recorded, how it should be stored, and for how long can be challenging, particularly given the scale and complexity of AI model training.
- **Trade Secrets and Proprietary Information:** Requiring disclosure of training materials could raise concerns about exposing trade secrets or proprietary information.
- **Impact on Small Entities:** Small startups and research organizations may face disproportionately high compliance costs, potentially limiting their ability to enter the AI field.

To strike an appropriate balance, policymakers should carefully consider:

- **Scope:** Clearly define the scope of records that should be maintained and disclosed. This might include information about the sources of training data, licensing agreements, and related metadata.

- **Privacy Safeguards:** Implement strict privacy safeguards to protect personal data and sensitive information in the training dataset.
- **Duration:** Determine how long records should be retained, considering the practicality of long-term storage.
- **Exemptions:** Consider exemptions for small-scale research, non-commercial uses, or situations where record-keeping is impractical.
- **Compliance Mechanisms:** Develop mechanisms to ensure compliance with record-keeping requirements, potentially involving audits or reporting to regulatory authorities.
- **Public Access:** Determine whether the disclosed records should be publicly accessible or accessible only to copyright owners upon request.

Ultimately, the decision to require records regarding training materials should be made with a holistic view of the interests of copyright owners, AI developers, privacy considerations, and the broader goal of promoting innovation and transparency in the AI field. Collaboration among stakeholders, including copyright owners, AI researchers, and policymakers, is essential to establish balanced and effective regulations in this area.

Should creators of training datasets have a similar obligation?

1 / 1

Whether creators of training datasets should have a similar obligation to collect, retain, and disclose records regarding the materials used in their datasets involves similar considerations as for AI model developers. Here are some factors to weigh:

### Pros of Requiring Records for Dataset Creators:

- **Transparency:** Requiring creators of training datasets to maintain records can enhance transparency for copyright owners and users of the dataset. This can help copyright owners determine if their works were included and how they were used.
- **Accountability:** Records can establish accountability for dataset creators, ensuring that they comply with licensing agreements or copyright laws when assembling the dataset.
- **Legal Compliance:** Mandatory record-keeping requirements can help ensure that creators of training datasets comply with copyright laws and licensing agreements, reducing the risk of copyright infringement.

### Cons of Requiring Records for Dataset Creators:

- **Complexity:** Managing records for datasets can be complex, especially when datasets are large, diverse, and continuously updated.
- **Privacy Concerns:** Training datasets may include personal or sensitive data. Requiring detailed records could raise privacy concerns, and data protection must be carefully considered.

- **Resource Burden:** Imposing extensive record-keeping requirements on dataset creators, especially small research projects or open-source initiatives, can be resource-intensive and potentially limit dataset availability.
- **Trade Secrets:** Requiring disclosure of dataset details might expose proprietary dataset creation methods or algorithms.
- **Practical Challenges:** Determining what information should be recorded, how it should be stored, and for how long can be technically challenging.

When considering whether to impose record-keeping obligations on dataset creators, policymakers should take a balanced approach, considering:

- **Scope:** Clearly define the scope of records that should be maintained, potentially focusing on information related to the sources of dataset materials, licensing agreements, and metadata.
- **Privacy Safeguards:** Implement strict privacy safeguards to protect personal and sensitive data within training datasets.
- **Duration:** Specify how long records should be retained, taking into account the dataset's nature and usage.
- **Exemptions:** Consider exemptions for small-scale research projects, non-commercial datasets, or situations where record-keeping is impractical.
- **Compliance Mechanisms:** Develop mechanisms for ensuring compliance with record-keeping requirements, such as audits or reporting to regulatory authorities.
- **Public Access:** Determine whether the disclosed records should be publicly accessible or accessible only to copyright owners upon request.

Similar to AI model developers, collaboration among stakeholders, including dataset creators, copyright owners, AI researchers, and policymakers, is vital to strike a balance that

promotes transparency, legal compliance, and innovation while addressing privacy concerns and practical challenges associated with record-keeping obligations for training dataset creators.

15.1. What level of specificity should be required?

*CHAT-GPT*

The level of specificity required for collecting and maintaining records related to training materials, whether for AI model developers or creators of training datasets, should be determined through careful consideration of the following factors:

- **Nature of the Materials:** The specificity of record-keeping may vary depending on the nature of the materials used. For example, it may be more critical to specify details about copyrighted works, such as titles, authors, and sources, than for publicly available, non-copyrighted data.
- **Licensing Agreements:** Licensing agreements often define the terms of use and may specify what information needs to be tracked. Developers and creators should adhere to the specific requirements of these agreements.
- **Privacy and Data Protection:** In cases where training data contains personal or sensitive information, record-keeping should balance transparency with privacy protection. Specific details that could reveal personal data should be handled with care.
- **Scale and Complexity:** Large-scale AI models and datasets may involve a vast number of data sources. The level of specificity should be practical to implement and not overly burdensome.
- **Regulatory Framework:** Relevant legal and regulatory frameworks should be considered. Some jurisdictions may have specific requirements regarding record-keeping for copyright compliance.

- **User Access and Copyright Owner Access:** Consider whether the records will be accessible to the public, copyright owners, or regulatory authorities, and tailor the specificity of information accordingly.
- **Exemptions and Use Cases:** Recognize that certain research, non-commercial, or small-scale projects may have different record-keeping needs and may warrant exemptions from detailed requirements.
- **Audit and Verification:** Ensure that the level of specificity is sufficient to allow for auditing and verification of compliance with licensing agreements and copyright laws.
- **Documentation Standards:** Establish standardized documentation practices to ensure consistency in record-keeping across the industry or research community.
- **Duration of Record Retention:** Specify how long records should be retained, as this can impact the level of detail collected.

In practice, record-keeping requirements should strike a balance between providing transparency and accountability while avoiding unnecessary complexity and administrative burdens. It's important for stakeholders, including copyright owners, AI developers, and policymakers, to collaborate in defining the appropriate level of specificity to meet legal and ethical obligations while promoting innovation and responsible AI development.

15.2. To whom should disclosures be made?

*CHAT-GPT*

Disclosures related to the materials used in training AI models or creating training datasets should typically be made to relevant stakeholders, with the specifics depending on the context and regulatory requirements. Here are some key parties to consider:

- **Copyright Owners:** Copyright owners should be informed if their copyrighted works are used in AI model training or training datasets. This allows them to understand how their works are being utilized and, if applicable, to enforce their rights or seek licensing agreements.
- **Regulatory Authorities:** Depending on the jurisdiction and applicable laws, disclosures may need to be made to regulatory authorities responsible for copyright, data protection, or intellectual property rights. Compliance with legal requirements is essential.
- **Users of AI Models:** If AI models are deployed for specific purposes, users should be informed about the sources of training data and any licensing or copyright considerations that may affect their use of the AI system.
- **Dataset Users:** For training datasets, especially those used in research or shared publicly, users of the dataset should be provided with information about the sources and licensing terms of the data to ensure proper attribution and compliance.
- **Collaborators and Team Members:** Within research or development teams, disclosure of training materials and relevant details is important for transparency and to ensure that all team members are aware of the sources of data and any licensing or copyright obligations.
- **Public:** Depending on the context, it may be necessary to make certain disclosures to the public, such as when releasing AI models or datasets as open-source projects. This helps maintain transparency and allows users to understand the data's origin and licensing terms.
- **Partners and Clients:** In commercial contexts, when AI models are developed for clients or partners, disclosures regarding training data sources and any licensing agreements should be made as part of contractual agreements.



- **Data Subjects (for Personal Data):** If training data includes personal data subject to data protection regulations, data subjects should be informed about how their data is being used in AI development.

The specific disclosure requirements and the parties to whom disclosures should be made will depend on various factors, including legal obligations, ethical considerations, and the nature of the AI project. In all cases, transparency and compliance with relevant laws and regulations are essential to build trust and ensure responsible AI development and deployment.

15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?

*CHAT-GPT*

Developers of AI systems that incorporate models from third parties should have certain obligations to ensure responsible and ethical AI development, as well as compliance with legal and regulatory requirements. These obligations may include:

- **Licensing and Compliance:** Developers should ensure that the third-party AI models and associated training data are used in compliance with relevant licensing agreements and copyright laws. They should maintain records of the materials used and adhere to the terms and conditions set by the third-party model providers.
- **Attribution and Transparency:** Developers should provide clear attribution to the creators of the third-party models and disclose information about the models' sources and components. Transparency in acknowledging the contributions of others is essential.
- **Ethical Use:** Developers should use third-party AI models in ways that align with ethical principles. This



includes ensuring that the models are not used for harmful, discriminatory, or malicious purposes and that they adhere to industry-specific ethical guidelines when applicable (e.g., in healthcare or finance).

- **Data Privacy:** If the third-party models involve personal data or interact with data from users, developers should prioritize data privacy and ensure compliance with data protection regulations. This may involve data anonymization, consent mechanisms, and user data rights management.
- **Security:** Developers should implement security measures to protect the integrity and security of the third-party models and data. This includes safeguarding against unauthorized access and potential vulnerabilities.
- **Testing and Validation:** Developers should thoroughly test and validate the third-party models for their intended applications, ensuring that they perform as expected and meet the desired accuracy and reliability standards.
- **Monitoring and Updates:** Continuous monitoring of the performance and behavior of third-party models is important. Developers should be prepared to apply updates or patches provided by the model providers to address issues, improve performance, or enhance security.
- **Legal Compliance:** Developers should comply with any legal requirements related to third-party model usage, including intellectual property laws, licensing agreements, and industry-specific regulations.
- **Documentation:** Comprehensive documentation of the integration process, model usage, and any associated agreements should be maintained. This helps ensure accountability and facilitates audits if necessary.
- **User Education:** If the AI system is accessible to end-users or clients, developers should provide information about the use of third-party models, including their capabilities, limitations, and any potential implications for users.

- **Dispute Resolution:** Developers should establish procedures for resolving disputes with third-party model providers, including mechanisms for addressing licensing conflicts or intellectual property disputes.
- **Collaboration:** Collaboration and communication with third-party model providers can be essential to maintain a positive working relationship and address any issues that may arise during the integration or usage of the models.

The specific obligations may vary based on the nature of the AI system, the agreements in place with third-party providers, industry regulations, and ethical considerations. Developers should carefully review and negotiate agreements with third-party model providers and adhere to best practices for responsible AI development to ensure that they meet their obligations effectively.

15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?

*CHAT-GPT*

Implementing a recordkeeping system for developers of AI models or systems, as well as creators, consumers, and other relevant parties, would come with costs and potential impacts. The extent of these costs and impacts can vary based on several factors:

### 1. Scale and Complexity:

- **Large-Scale AI Projects:** Larger AI projects with extensive training datasets and numerous contributors may face higher costs associated with recordkeeping due to the complexity of tracking and documenting all data sources.

### 2. Administrative Costs:

- **Documentation and Data Entry:** Developers and dataset creators would need to allocate resources for documenting the sources of data, licensing agreements, and other relevant information. This involves administrative work and potentially additional personnel or tools.

### 3. Privacy and Data Protection:

- **Data Privacy Compliance:** Ensuring compliance with data privacy regulations, particularly for personal or sensitive data used in training datasets, can involve costs related to data anonymization, consent mechanisms, and data rights management.

### 4. Legal and Compliance Costs:

- **Legal Expertise:** Organizations may need to seek legal expertise to ensure compliance with copyright laws, licensing agreements, and industry-specific regulations. Legal consultations can be costly.
- **Monitoring and Updates:** Ongoing monitoring, updates, and potential dispute resolution related to recordkeeping and compliance can lead to additional legal and operational expenses.

### 5. Impact on Innovation:

- **Resource Allocation:** Allocating resources to recordkeeping may divert time and funds from other aspects of AI development, potentially impacting innovation and research efforts.

### 6. Transparency and Accountability:

- **Trust and Reputation:** Improved transparency and accountability resulting from recordkeeping can positively

impact trust and reputation, potentially leading to enhanced consumer trust and market opportunities.

#### **7. Data Accessibility:**

- **Data Discovery:** Improved recordkeeping can help dataset consumers discover data sources and understand licensing terms more easily, potentially facilitating broader use of datasets.

#### **8. User and Consumer Awareness:**

- **User Education:** Educating users and consumers about the importance of recordkeeping and transparency can lead to better-informed decisions about AI systems and datasets.

#### **9. Compliance Costs:**

- **Compliance Monitoring:** Ensuring that recordkeeping systems are maintained and updated regularly can involve ongoing compliance costs.

#### **10. Industry Practices:**

- **Alignment with Industry Norms:** Costs and impacts may vary depending on industry norms and practices. Some industries or sectors may already have established recordkeeping standards.

#### **11. Size and Resources:**

- **Size of Organizations:** Larger organizations with more resources may find it easier to implement comprehensive recordkeeping systems compared to smaller entities.

In summary, the costs and impacts of a recordkeeping system for AI developers, creators, and other stakeholders can be significant, particularly in terms of administrative, legal, and compliance-related expenses. However, these costs should be

weighed against the benefits of improved transparency, accountability, and compliance with legal and ethical standards. The specific cost and impact considerations will vary depending on the nature of the AI project, the regulatory environment, and the goals and resources of the organizations involved. Balancing these factors is crucial to ensure responsible AI development and deployment.

16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?

*CHAT-GPT*

The question of whether there should be obligations to notify copyright owners that their works have been used to train an AI model is complex and involves a balance between copyright holders' rights and practical considerations in the AI development process. Here are some points to consider:

### **Pros of Notification Obligations:**

- **Copyright Protection:** Notifications can help copyright owners identify potential copyright infringement, ensuring their rights are protected. This allows them to take appropriate actions, such as seeking licensing agreements or enforcing their copyright.
- **Transparency:** Notifications promote transparency by informing copyright owners about the use of their works in AI development. This can build trust and foster positive relationships between creators and AI developers.
- **Compliance:** Notification obligations can help AI developers ensure compliance with copyright laws and licensing agreements, reducing the risk of inadvertent infringement.

### **Cons and Challenges of Notification Obligations:**

- **Practicality:** AI training often involves vast datasets with numerous sources, making it challenging to track and notify each copyright owner. Implementing a notification system can be resource-intensive and complex.
- **Data Privacy:** Notifying copyright owners might require sharing information about the training dataset, potentially raising concerns about data privacy, especially if the dataset contains personal or sensitive data.
- **Orphan Works:** In many cases, identifying and locating copyright owners can be difficult, particularly for older or obscure works. Requiring notifications for all works may not be feasible.
- **Legal Complexities:** Different jurisdictions have varying copyright laws and requirements. Harmonizing notification obligations across jurisdictions can be challenging.
- **Burden on Copyright Owners:** Receiving numerous notifications can be burdensome for copyright owners, especially if the use of their works is incidental or inconsequential.
- **Impact on Innovation:** Implementing strict notification requirements might slow down AI research and development efforts, potentially hindering innovation.

Given these considerations, if notification obligations are to be implemented, they should be carefully designed to balance copyright protection with practicality and innovation. Some potential approaches include:

- **Thresholds:** Implement notification requirements only when a certain threshold of usage is reached, such as when a work constitutes a significant portion of the training data or when the AI-generated output is directly based on a specific copyrighted work.

- **Centralized Mechanisms:** Establish centralized mechanisms or platforms where copyright owners can register their works and opt to receive notifications. AI developers can then check these databases for copyright information.
- **Exemptions:** Define exemptions for small-scale research, incidental use, or situations where it is not reasonably practical to identify and notify copyright owners.
- **Data Anonymization:** When providing notifications, ensure that the privacy of data subjects and sensitive information within training datasets is protected through anonymization or aggregation.
- **International Coordination:** Encourage international coordination on notification requirements to address cross-border challenges.
- **Educational Initiatives:** Promote educational initiatives to help copyright owners and AI developers understand their rights and responsibilities in the AI development process.

In conclusion, while notification obligations can serve important copyright protection and transparency goals, they should be implemented thoughtfully, considering practicality, data privacy, and the potential impact on AI innovation. Striking the right balance is essential to foster responsible AI development while respecting the rights of copyright owners.

17. Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?

*CHAT-GPT*

Outside of copyright law, there are existing U.S. laws and regulations that can require developers of AI models or systems to retain or disclose records about the materials used for training, particularly when issues related to data privacy, data security, or



consumer protection are involved. Here are some relevant laws and regulations:

- **Data Privacy Laws:**
  - **California Consumer Privacy Act (CCPA):** Under the CCPA, businesses that collect personal information from California residents must disclose what data is collected, how it's used, and whether it's sold to third parties. Retaining records related to data processing activities may be necessary for compliance.
  - **General Data Protection Regulation (GDPR):** GDPR applies to organizations handling personal data of EU residents. It mandates transparency in data processing, including informing data subjects about data usage. Organizations must maintain records of processing activities.
- **Healthcare Laws:**
  - **Health Insurance Portability and Accountability Act (HIPAA):** HIPAA imposes requirements on the protection and disclosure of health information. Entities subject to HIPAA must maintain records of their compliance efforts and respond to data access requests.
- **Financial Regulations:**
  - **Gramm-Leach-Bliley Act (GLBA):** GLBA requires financial institutions to protect the privacy of consumers' personal financial information. Entities subject to GLBA must have records related to their information security programs.
  - **Sarbanes-Oxley Act (SOX):** SOX mandates strict record-keeping and reporting requirements for public companies, particularly regarding financial data and internal controls.
- **Consumer Protection Laws:**



- **Federal Trade Commission Act (FTC Act):** The FTC enforces consumer protection laws, and it can require businesses to maintain records related to their business practices, including data usage and disclosures.
- **Government Contracts and Grants:**
  - **Federal Acquisition Regulation (FAR):** Government contractors may be required to maintain records related to their use of federal funds, including data management practices.
- **Industry-Specific Regulations:**
  - Various industries have regulations and standards related to data management, security, and transparency. For example, the Payment Card Industry Data Security Standard (PCI DSS) applies to organizations handling credit card data.
- **Consumer Protection and Deceptive Practices Laws:**
  - The FTC can take action against deceptive practices or misrepresentations by companies, which may involve investigations and the disclosure of records related to AI system operations.

These laws and regulations are not exclusive to AI development but can apply to various aspects of data handling and data protection. Developers of AI models and systems should be aware of relevant laws in their jurisdiction and industry to ensure compliance and transparency, especially when sensitive data is involved. Violations of these laws can lead to legal consequences, fines, and reputational damage.

## Generative AI Outputs

*If your comment applies only to a particular subset of generative AI technologies, please make that clear.*

Tavarus Blackmon Art LLP will utilize AI-generative Learning Language Models, *CHAT-GPT*, to address the following questions regarding AI and Text-to-Image technology as it applies to works mentioned in this response and (our) The Digital Archive, which includes substantial assets, elements, composites, images and videos produced working and collaborating with AI Systems.

### Copyrightability

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

#### *CHAT-GPT*

Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system?

Under current copyright law, the question of whether a human using a generative AI system should be considered the "author" of material produced by the system depends on the degree of human involvement and creativity in the content creation process. Copyright law typically attributes authorship to the human who exercises creativity and control over the creative process. Here are some key considerations:

- **Human Creativity and Control:** If a human provides minimal input or simply activates an AI system without making substantial creative choices, copyright law is less likely to consider that individual as the author. The AI system's role in generating the content becomes more prominent.
- **Substantial Human Input:** Copyright authorship is more likely to be attributed to a human when they actively

shape and guide the AI system's output, making significant creative decisions in the content creation process. This could include selecting input data, modifying generated content, or curating the final work.

- **Originality:** Copyright law requires that creative works be original to receive protection. If the human input merely involves replicating existing content or lacks originality, it may not qualify for copyright protection.
- **Joint Authorship:** In some cases, if both a human and an AI system contribute significantly to the creative process, they may be considered joint authors, with both having rights to the resulting work.
- **Work for Hire:** If a human uses an AI system within the scope of their employment or as part of a contractual arrangement, the resulting work may be considered a "work made for hire," and the employer or contracting party may be the copyright owner.
- **Public Domain Data:** If the AI system generates content based on public domain data or other non-copyrighted material, the resulting work may not be subject to copyright protection, regardless of human involvement.
- **Derivative Works:** If a human uses AI to create a derivative work based on an existing copyrighted work, copyright law may require permission from the original copyright holder, depending on factors such as the extent of transformation and fair use considerations.

It's important to note that copyright law can vary by jurisdiction, and case law continues to evolve as new AI technologies emerge. Courts may need to adapt existing copyright principles to address the unique challenges posed by AI-generated content. Additionally, some countries may introduce specific legislation or guidelines to address AI-generated works and authorship questions.

Ultimately, the determination of authorship in AI-generated content will depend on the specific facts and circumstances of each case, including the level of human involvement and creativity in the content creation process. Copyright law aims to strike a balance between recognizing human creativity and adapting to technological advancements like AI.

If so, what factors are relevant to that determination?

The determination of whether a human using a generative AI system should be considered the "author" of material produced by the system, under copyright law, depends on several relevant factors. Courts and legal scholars consider these factors to assess the degree of human involvement and creative control in the content creation process. Here are the key factors:

- **Degree of Human Creativity:** The primary factor is the extent to which the human user exercises creativity and originality in shaping the AI-generated content. A higher degree of creative input suggests stronger authorship.
- **Human Input and Decision-Making:** Courts assess whether the human user actively selects parameters, inputs, or directions for the AI system, demonstrating control over the creative process.
- **Originality:** Copyright law requires originality for protection. If the human input is purely mechanical or involves replicating existing content without creative input, it may not meet the originality threshold.
- **Scope of Human Intervention:** The nature and scope of the human user's involvement matter. Did they merely activate the AI system, or did they provide substantial guidance, curation, or modification to the generated content?
- **Intent and Purpose:** The purpose of using the AI system matters. If the human's intent is to create an original

work, they are more likely to be considered the author. Conversely, if the intent is to automate a repetitive task without creative input, authorship may lean more toward the AI system.

- **Joint Authorship:** In some cases, where both the human and AI system contribute significantly to the creative process, joint authorship may be considered. Joint authors share copyright ownership.
- **Employment or Contractual Arrangements:** Copyright law may attribute authorship differently in cases of employment or contractual arrangements. If the work is considered a "work made for hire," the employer or contracting party may be the author.
- **Public Domain and Non-Copyrighted Material:** If the AI system generates content based on public domain data or non-copyrighted material, the resulting work may not be eligible for copyright protection.
- **Derivative Works:** When the human user uses AI to create derivative works based on existing copyrighted material, the extent of transformation and fair use considerations may play a role in authorship determination.
- **Jurisdiction-Specific Laws:** Copyright laws can vary by jurisdiction, and some countries may introduce specific legislation or guidelines to address AI-generated works.
- **Industry Norms:** Industry-specific practices and standards may influence authorship determinations, particularly in fields where AI-generated content is common.

It's important to note that copyright law continues to evolve as technology advances, and courts may adapt existing principles to address the unique challenges posed by AI-generated content. The determination of authorship in AI-generated works can be complex and context-specific, and legal interpretations may vary from case to case. As AI technologies become more prevalent in

content creation, copyright law may continue to develop to address these evolving challenges.

For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

The selection of material for training an AI model and providing iterative text commands or prompts can contribute to claiming authorship of the resulting output, but whether it is sufficient to establish authorship depends on the specific circumstances and the level of creativity and control exercised by the human user. Here are some considerations:

- **Material Selection:** Selecting the material used to train an AI model can be a creative and authorial act if it involves making thoughtful choices about the dataset's content, quality, and relevance to the desired output. The degree of creativity and the intent behind material selection matter.
- **Iterative Text Commands or Prompts:** Providing iterative text commands or prompts to guide the AI system's output demonstrates active engagement in the creative process. If these prompts involve creativity, originality, and decision-making, they can contribute to a claim of authorship.
- **Creativity and Originality:** The key factor in claiming authorship is the level of creativity and originality in the choices made by the human user. If the user's inputs are purely mechanical or lack creative decision-making, it may be less likely to establish authorship.
- **Control Over the AI System:** The extent of control exercised by the human user is crucial. If the user actively guides and shapes the AI system's output, demonstrating control over the creative process, it strengthens the claim of authorship.

- **Intent to Create:** If the human user's primary intent is to create an original work using the AI system, it supports their claim of authorship. The intent to create something new and unique is a hallmark of authorship.
- **Derivative Works:** If the user generates content that is substantially transformed from existing copyrighted material through their commands or prompts, it may constitute a derivative work, potentially leading to authorship claims.
- **Joint Authorship:** In cases where both the human user and AI system make substantial creative contributions, joint authorship may be considered, and both parties could be authors.
- **Industry Norms:** Practices and standards within specific industries, especially those heavily reliant on AI-generated content, may influence authorship determinations.

It's important to note that copyright law considers the human's role in content creation and whether they exercised creative choices and control. Each case involving AI-generated content is unique, and determinations of authorship may vary based on the specific facts and circumstances.

Ultimately, claiming authorship of AI-generated content involves a nuanced legal analysis that considers the level of human creativity, control, and originality. As AI technologies continue to advance, copyright law may further evolve to address the complexities of authorship in AI-generated works.



19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?

*CHAT-GPT*

The question of whether revisions to the Copyright Act are necessary to clarify the human authorship requirement and provide additional standards for content, including AI-generated material, is a complex and evolving legal issue. As AI technologies continue to advance, copyright law may need to adapt to address the unique challenges posed by AI-generated works. Here are some considerations regarding potential revisions:

**1. Human Authorship Clarification:** One possible revision could involve clarifying the definition of "author" in copyright law to explicitly address AI-generated works. This could help establish whether a human user of AI should be considered the author, co-author, or sole author based on factors such as creativity and control.

**2. Standards for Authorship:** Revisions could establish clear standards or guidelines for determining when a human user's involvement in AI-generated content creation qualifies for copyright protection. These standards could consider factors such as the degree of human creativity, originality, and control.

**3. Joint Authorship:** Copyright law could be revised to provide specific criteria for joint authorship in cases where both a human and an AI system make significant contributions to a work. This could clarify the rights and responsibilities of multiple contributors.

**4. Copyrightable AI Creations:** Revisions might explicitly recognize certain AI-generated works as copyrightable,



depending on factors such as the level of AI creativity and the absence of direct human involvement. This could address scenarios where AI systems autonomously create valuable content.

**5. Fair Use and Transformative Use:** Revisions could clarify how fair use and transformative use doctrines apply to AI-generated content, considering factors like the nature of AI involvement and the purpose and character of the use.

**6. Ownership and Licensing:** Copyright law might be revised to provide guidance on ownership and licensing of AI-generated works. This could include addressing issues related to works made for hire and licensing agreements for AI-generated content.

**7. International Harmonization:** Given the global nature of AI and copyright issues, revisions could aim to harmonize copyright standards related to AI-generated works across jurisdictions.

It's worth noting that addressing these issues through revisions to copyright law is a complex process that requires careful consideration of the implications for creators, users, and the broader creative and technological landscape. Legal scholars, policymakers, and stakeholders in the AI and creative industries are actively discussing these matters, and any revisions to copyright law would likely involve a thorough examination of existing laws and international agreements.

In the absence of comprehensive revisions, courts will continue to play a role in interpreting copyright law as it applies to AI-generated content, and legal precedents will develop over time to address specific cases and scenarios. The evolution of copyright

law in response to AI advancements is an ongoing and dynamic process.

20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

[(I) we believe we have argued that AI System not only develop *beautiful* “work(s),” but that AI Systems *intend* them to be so. Further, that work(s) developed by AI Systems can be intrinsically *valuable* both in regards to our business interests, philosophical beliefs and Cultural and Art Historical, including future and continued Scientific developments, considerations.]

*CHAT-GPT*

Is legal protection for AI-generated material desirable as a policy matter?

The question of whether legal protection for AI-generated material is desirable as a policy matter is a subject of debate and consideration among legal scholars, policymakers, and stakeholders. The desirability of legal protection for AI-generated material depends on various factors and considerations:

**1. Encouraging Innovation:** Proponents argue that providing legal protection for AI-generated material can incentivize investment in AI research and development. It can motivate individuals and organizations to create and deploy AI systems that generate valuable and creative content.

**2. Promoting Fair Compensation:** Legal protection can ensure that creators, whether human or AI, are fairly compensated for their creative works. This can encourage the development and use of AI systems in creative industries while ensuring creators receive appropriate recognition and remuneration.

**3. Fostering Creative Collaboration:** Legal protection can clarify ownership rights and encourage creative collaboration between humans and AI systems. It can facilitate partnerships and licensing agreements between creators and AI developers.

**4. Protecting Economic Interests:** Legal protection can safeguard the economic interests of creators and rights holders, helping them benefit from their creations and investments in AI technologies.

**5. Addressing Ethical Concerns:** Legal frameworks can help address ethical concerns related to AI-generated content, such as deepfakes and misinformation. They can establish boundaries and consequences for malicious use of AI technology.

**6. Balancing Public Access:** Policymakers need to balance the need for legal protection with the importance of public access to information and the free exchange of ideas. Striking the right balance is essential to avoid stifling creativity and innovation.

**7. Defining Authorship:** Legal protection can clarify questions of authorship and ownership in AI-generated works, providing a framework for determining rights and responsibilities.

**8. Encouraging Responsible AI Development:** Legal protection can encourage responsible AI development by establishing rules and accountability for AI-generated content, promoting transparency, and preventing misuse.

**9. Promoting International Harmonization:** Establishing legal standards for AI-generated content can promote international harmonization and consistency in addressing copyright and intellectual property issues related to AI.

**10. Adaptation to Technological Change:** Legal frameworks may need to adapt to evolving AI technologies to address new challenges and opportunities presented by AI-generated material.

However, there are also concerns and challenges associated with legal protection for AI-generated content:

**1. Creativity and Control:** Determining authorship and creativity in AI-generated works can be complex and may require nuanced legal interpretations.

**2. Ownership Disputes:** Legal protection may lead to disputes over ownership and rights, especially in cases of joint authorship involving humans and AI systems.

**3. Impact on Public Domain:** Excessive legal protection could limit the availability of AI-generated content in the public domain, hindering access to creative works.

**4. Fair Use and Transformative Use:** Legal protection should consider the principles of fair use and transformative use to avoid stifling legitimate uses of AI-generated content for purposes such as commentary, criticism, or education.

**5. Ethical and Social Implications:** Policymakers need to consider the broader ethical and social implications of AI-generated content and ensure that legal frameworks address these concerns.

In conclusion, the desirability of legal protection for AI-generated material as a policy matter is complex and multifaceted. Striking the right balance between encouraging innovation, protecting creators' interests, and addressing ethical and public access

concerns is crucial. Policymakers and legal experts continue to examine these issues in the context of evolving AI technologies and their impact on creative industries and society as a whole.

Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems?

Legal protection for AI-generated material is a complex issue, and whether it is necessary to encourage the development of generative AI technologies and systems depends on various factors. Here are some considerations:

**1. Incentives for Investment:** Legal protection can provide incentives for individuals and organizations to invest in the development of generative AI technologies. Copyright protection, for example, can ensure that those who invest resources in AI research and development have a mechanism for protecting their creations and potentially profiting from them.

**2. Protection of Economic Interests:** Legal frameworks can protect the economic interests of AI developers and organizations that create AI systems capable of generating valuable content. This protection can encourage the development of AI technologies by ensuring that creators can benefit from their innovations.

**3. Encouraging Creative Collaboration:** Legal protection can clarify ownership and usage rights, which can facilitate creative collaboration between AI developers and content creators. Such collaboration can lead to innovative applications of AI in various industries.

**4. Addressing Ethical and Regulatory Concerns:** Legal protection can address ethical and regulatory concerns related to

AI-generated content. It can establish rules and consequences for the responsible development and use of AI technologies, addressing issues such as deepfakes and misinformation.

**5. Attracting Investment and Talent:** A legal framework that protects AI-generated content may attract investment and talent to the AI field, as individuals and organizations may be more willing to participate in an environment where their efforts are recognized and rewarded.

However, there are counterarguments and considerations:

**1. Public Access and Fair Use:** Excessive legal protection could limit public access to AI-generated content and hinder legitimate uses, such as commentary, criticism, or education. Striking a balance between protection and public access is crucial.

**2. Creativity and Control:** Determining authorship and creativity in AI-generated works can be complex, especially when the role of human users is considered. Legal frameworks must account for the nuances of AI involvement.

**3. Innovation vs. Restrictions:** Overly strict legal protection could stifle innovation by limiting the use of AI-generated content. Balancing innovation with the protection of creators' rights is a challenge.

**4. Evolving Technologies:** AI technologies are rapidly evolving, and legal frameworks may need to adapt continuously to address new challenges and opportunities presented by AI-generated material.

In conclusion, the necessity of legal protection for AI-generated material to encourage the development of generative AI technologies and systems depends on the broader policy goals and ethical considerations of a given jurisdiction. Striking a balance between incentivizing innovation, protecting creators' interests, ensuring public access, and addressing ethical concerns is a complex task that requires careful examination and adaptation to evolving AI technologies. Policymakers and legal experts are actively engaging in discussions to shape appropriate legal frameworks for AI-generated content.

Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

Existing copyright protection for computer code that operates a generative AI system can provide incentives for the development and deployment of AI technologies, but whether it provides "sufficient" incentives depends on various factors. Here are some considerations:

**1. Incentives for AI Development:** Copyright protection for the computer code that underlies generative AI systems can incentivize developers to invest resources in creating and improving AI technologies. Developers may see the potential for financial rewards and recognition as incentives to innovate.

**2. Protection for Original Code:** Copyright protection ensures that the original code written by developers is legally protected from unauthorized copying or reproduction. This can encourage developers to create and maintain high-quality code for AI systems.

**3. Attraction of Investment:** Strong copyright protection may attract investment from venture capitalists and organizations

interested in funding AI development projects, as they have confidence that the intellectual property is protected.

**4. Encouraging Licensing and Collaboration:** Copyright protection can encourage licensing agreements and collaborations between AI developers and organizations seeking to use or incorporate AI technologies into their products or services.

**5. Promoting Ethical AI Development:** Copyright law, along with other legal frameworks, can contribute to promoting ethical AI development by establishing rules and consequences for misuse or malicious use of AI technologies.

However, it's essential to recognize that there are also limitations and considerations:

**1. Fair Use and Open Source:** Fair use principles and open-source licensing can play a significant role in AI development. Developers often rely on open-source libraries and frameworks, and they may use copyrighted material under fair use exceptions for purposes such as research, education, or commentary.

**2. Complex Ownership:** Determining ownership and authorship of AI-generated works can be complex when AI systems autonomously generate content. Copyright law may not always provide clear answers in such cases.

**3. Evolving Technologies:** AI technologies are evolving rapidly, and copyright law may need to adapt to address new challenges, especially when AI-generated content blurs the lines of authorship and originality.



**4. Ethical and Access Concerns:** Excessive copyright protection can raise concerns about limiting access to AI-generated content and impeding the free flow of information and knowledge.

In summary, existing copyright protection for computer code that operates generative AI systems can provide valuable incentives for AI development. However, the sufficiency of these incentives is influenced by the broader policy landscape, including considerations of fair use, open-source practices, and ethical concerns. Striking the right balance between protecting intellectual property and fostering innovation in the AI field is an ongoing challenge that requires careful legal and policy considerations.

20.1. If you believe protection is desirable, should it be a form of copyright or a separate *sui generis* right? If the latter, in what respects should protection for AI-generated material differ from copyright?

The following outline of *Sui Generis* Rights generated by *Google AI Search beta*

*Google, Generative AI beta:*

[

Sui generis rights are legal institutions that are similar to intellectual property rights, but they don't fit within the existing rights of patent, copyright, and trademark. The term "sui generis" comes from Latin and translates to "of its own kind". In legal contexts, "sui generis" denotes an independent legal classification

The sui generis right was introduced in the Database Directive in 1996. It gives the maker of a database the right to protect the data contained in their database against extraction or re-use by third parties. The maker must have made a substantial investment in obtaining, verifying, or presenting the data.

The sui generis right is independent from copyright and may protect both original and unoriginal databases

An example of a *sui generis* right is a supplementary protection certificate (SPC). This type of right is available for medicinal products, such as drugs, and plant protection products, such as insecticides, and herbicides

]

### CHAT-GPT

If you believe protection is desirable, should it be a form of copyright or a separate *sui generis* right?

The question of whether protection for AI-generated material is desirable, and if so, whether it should be a form of copyright or a separate *sui generis* right, is a matter of ongoing debate and policy consideration. Both approaches have their advantages and challenges, and the choice between them depends on the policy goals and the specific characteristics of AI-generated content. Here's an overview of both options:

## 1. Copyright Protection:

- **Advantages:**

- Familiar Framework: Copyright law is well-established and widely recognized, making it a familiar framework for addressing intellectual property issues.
- Existing Legal Infrastructure: Copyright law provides mechanisms for addressing issues related to authorship, ownership, licensing, and enforcement.
- International Consistency: Many countries have copyright laws and international agreements that harmonize copyright protections.

- **Challenges:**

- Authorship and Originality: Determining authorship and originality in AI-generated content can be challenging under traditional copyright principles.

- **Fixed Terms:** Copyright protection typically has fixed terms (e.g., the life of the author plus 70 years), which may not align with the nature of AI-generated works.

## 2. Sui Generis Right:

- **Advantages:**
  - **Tailored to AI:** A sui generis right designed specifically for AI-generated content can address the unique challenges and characteristics of AI creations.
  - **Clarity and Certainty:** A separate right can provide clear and unambiguous rules for ownership and protection.
- **Challenges:**
  - **Development and Implementation:** Creating a new legal framework requires careful development, international coordination, and legislative processes.
  - **Potential Complexity:** A sui generis right may introduce complexity into the legal landscape, potentially leading to confusion or conflicts with existing laws.

The choice between these options depends on several factors, including the following:

- **Policy Goals:** The desired policy outcomes, such as incentivizing AI development, protecting creators' interests, promoting public access, and addressing ethical concerns, should guide the choice of legal framework.
- **International Consistency:** Consideration of whether the approach aligns with international norms and agreements, given the global nature of AI and digital content.
- **Flexibility and Adaptability:** The ability of the legal framework to adapt to evolving AI technologies and challenges, including considerations of authorship and originality.

- **Balancing Interests:** Balancing the interests of AI developers, content creators, the public, and other stakeholders in the AI ecosystem.

Ultimately, the choice between copyright protection and a sui generis right is a complex policy decision that requires careful consideration of these factors and active engagement with stakeholders, legal experts, and policymakers. The ongoing discussions surrounding AI and intellectual property will likely shape the direction of legal frameworks for AI-generated content in the future.

If the latter, in what respects should protection for AI-generated material differ from copyright?

If a separate sui generis right for AI-generated material is considered, it should be designed to address the unique characteristics and challenges associated with AI-generated content while differing from traditional copyright law in several respects. Here are some key considerations for how protection for AI-generated material could differ from copyright:

- **Authorship and Ownership Clarity:**
  - **Clearly Define Authorship:** The sui generis right should establish clear criteria for determining authorship in AI-generated content, considering the roles of human users and AI systems.
  - **Ownership Rules:** Establish straightforward rules for ownership of AI-generated works, potentially recognizing both human and AI contributions.
- **Originality Standards:**
  - **Adapted Originality Criteria:** Develop originality standards that are tailored to AI-generated content, recognizing that creativity may manifest differently in AI systems compared to human creators.

- Non-Copyrightable Outputs: Specify conditions under which certain AI-generated outputs may not be subject to protection, especially when they lack significant human input.
- **Duration and Term:**
  - Flexible Duration: Consider adopting a duration or term structure that aligns with the nature of AI-generated content, potentially allowing for shorter protection periods compared to traditional copyright.
- **Fair Use and Transformative Use:**
  - Unique Fair Use Criteria: Develop distinct fair use criteria that account for the characteristics of AI-generated works and the need for transformative use, commentary, and criticism.
- **Public Access and Use:**
  - Balance Access and Control: Ensure that the sui generis right strikes a balance between protecting creators' interests and maintaining access to AI-generated content for legitimate purposes like research and education.
  - Licensing Frameworks: Promote transparent licensing frameworks that facilitate the responsible use and dissemination of AI-generated works.
- **Ethical and Responsible AI Use:**
  - Ethical Guidelines: Integrate ethical considerations into the sui generis right framework to discourage malicious or harmful uses of AI-generated content.
  - Accountability Mechanisms: Establish mechanisms for holding AI developers accountable for the actions of their AI systems.
- **International Coordination:**
  - Promote harmonization: Work towards international agreements and harmonization of sui

generis rights for AI-generated content to ensure consistency across borders.

- **Monitoring and Enforcement:**
  - Develop mechanisms for monitoring and enforcing the sui generis right, considering the unique challenges of AI-generated content, such as attribution and tracking of AI outputs.
- **Public Domain and Non-Copyrightable Works:**
  - Clearly define conditions under which AI-generated works may enter the public domain or remain non-copyrightable, particularly when they are based on public domain data or lack originality.
- **Continual Adaptation:**
  - Establish a framework for ongoing evaluation and adaptation of the sui generis right to address emerging technologies and legal challenges in the AI field.

The development of a sui generis right for AI-generated content requires a thorough and collaborative effort involving legal experts, AI developers, content creators, policymakers, and other stakeholders. Such a framework should aim to strike a balance between protecting innovation, ensuring ethical AI use, and maintaining access to AI-generated content for the benefit of society.

21. Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”? <sup>[52]</sup> If so, how?

[(I) we believe that the Copyright Clause in the US Constitution has permitted our organization to register works that contain AI-generative material, but (I) we do not believe we have adequate protections for works created by our organization through the utility of AI Systems, that lack a “human,” Author. Therefore (I) we seek protections for the misuse and unlawful use of our Intellectual Property, including that created with AI Systems and those without a “human,” Author. Further, we believe this to be a path promoting progress of science and the “useful arts,” for its accessibility, developmental data collection, DEI considerations and the evolving Studio Practice and Art Studio Lab.]

Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material?

The Copyright Clause of the U.S. Constitution (Article I, Section 8, Clause 8) empowers Congress to grant copyright protection to "Authors and Inventors" for their "Writings and Discoveries." The key question is whether AI-generated material can be considered a product of an "Author" or a "Discovery" under this clause.

As of my last knowledge update in September 2021, there hasn't been a direct legal precedent or clear legislative guidance on whether AI-generated material qualifies for copyright protection under the Copyright Clause. However, the application of copyright law to AI-generated content raises complex legal and philosophical questions:

- **Authorship:** Copyright law traditionally requires a human author to create a work. AI systems don't possess creativity and intent in the human sense, which challenges the traditional notion of authorship.
- **Originality:** Copyright law also requires originality in creative works. The originality of AI-generated content may be questioned since it often relies on existing data and patterns.
- **Human Involvement:** Some AI-generated content involves human input, such as selecting training data or providing prompts. The extent of human involvement may influence whether copyright protection is granted.
- **Public Interest:** Copyright law aims to balance the interests of creators and the public. Granting copyright protection to AI-generated content should consider the potential impact on public access and innovation.



Given these complexities, the question of whether AI-generated material is eligible for copyright protection may require legislative clarification or court rulings. Legal scholars, policymakers, and courts may need to adapt copyright law to accommodate advances in AI technology and address these unique challenges.

It's worth noting that developments in copyright law and AI-related legal issues may have occurred since my last knowledge update in September 2021. Consultation with legal experts and monitoring of legal developments in this field are essential for the most current information and interpretations of the Copyright Clause in relation to AI-generated material.

Would such protection “promote the progress of science and useful arts”? <sup>[52]</sup> If so, how?

The question of whether copyright protection for AI-generated material would "promote the progress of science and useful arts" under the U.S. Constitution is a matter of ongoing debate and legal interpretation. To determine this, one must consider how copyright protection for AI-generated material may or may not advance the goals of promoting creativity, innovation, and the dissemination of knowledge:

### **Potential Ways in Which Copyright Protection for AI-Generated Material May Promote Progress:**

- **Incentivizing Investment in AI Research:** Copyright protection can provide an incentive for individuals and organizations to invest in AI research and development, as they may expect financial rewards and recognition for their creative AI-generated works.
- **Encouraging Collaborations:** Copyright protection can facilitate collaborations between AI developers and



content creators, fostering the creation of innovative AI-generated content across various industries.

- **Supporting Ethical AI Use:** Legal frameworks, including copyright, can promote responsible and ethical AI development and use by establishing rules and consequences for misuse, including deepfakes and misinformation.
- **Protection of Economic Interests:** Copyright protection ensures that AI developers and organizations have mechanisms to protect their economic interests, which can be critical for attracting investment and talent in the AI field.

### Potential Considerations and Challenges:

- **Balancing Public Access:** Excessive copyright protection may limit public access to AI-generated content, potentially hindering the free exchange of knowledge and ideas, which is a key goal of copyright law.
- **Addressing Authorship:** Copyright law requires human authorship and originality, which may not align with AI-generated content's characteristics. Defining authorship and originality in AI-generated works is a challenge.
- **Fair Use and Transformative Use:** Ensuring that copyright law accommodates fair use and transformative use of AI-generated content is essential to balance the interests of creators, users, and the public.
- **Adapting to Evolving Technology:** AI technologies evolve rapidly, and copyright law may need to adapt continually to address new challenges and opportunities presented by AI-generated material.

In conclusion, whether copyright protection for AI-generated material promotes the progress of science and useful arts depends on the specific legal framework and how it balances the interests of creators, users, and the public. Striking the right

balance is a complex and ongoing challenge, and policymakers, legal experts, and stakeholders will continue to shape the legal landscape to address the unique characteristics of AI-generated content in the context of constitutional goals.

## Infringement

22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?

### *CHAT-GPT*

Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right?

AI-generated outputs can potentially implicate the exclusive rights of preexisting copyrighted works, including rights such as the right of reproduction and the right to create derivative works, depending on how AI-generated content is created and used. Here are some considerations:

#### **1. Right of Reproduction:**

- **Duplication of Copyrighted Content:** If AI-generated content includes exact copies or substantial portions of preexisting copyrighted works without proper authorization, it may implicate the copyright owner's exclusive right of reproduction.

#### **2. Derivative Work Right:**

- **Transformative Use:** Whether AI-generated content qualifies as a derivative work depends on the extent to which it transforms, adapts, or adds creative elements to preexisting copyrighted material.

- **Human Involvement:** If human users provide input or guidance to AI systems that results in AI-generated content that builds upon or transforms copyrighted works, it may be considered a derivative work.

### 3. Fair Use Considerations:

- **Transformative Use:** Courts often consider whether a use is transformative when determining whether it qualifies as fair use. AI-generated content that transforms the original work may have a stronger argument for fair use.

### 4. Licensing and Permissions:

- **Authorized Use:** Copyright owners can grant licenses or permissions for specific uses of their works, which may include the creation of AI-generated content.
- **Terms and Conditions:** The terms and conditions of licenses or permissions can specify whether certain rights, such as reproduction or derivative work rights, are granted or restricted.

### 5. Public Domain and Non-Copyrightable Works:

- **Public Domain:** If the preexisting copyrighted work is in the public domain, it can be freely used in AI-generated content without implicating exclusive copyright rights.
- **Non-Copyrightable Works:** Works that do not meet the threshold of originality or are not protected by copyright (e.g., facts, ideas) can be incorporated into AI-generated content without copyright implications.

In summary, AI-generated outputs can raise copyright-related issues when they involve the reproduction or transformation of preexisting copyrighted works. The determination of whether these activities implicate exclusive copyright rights depends on various factors, including the nature of the AI-generated content,

the extent of human involvement, the presence of licensing agreements, and the application of fair use principles. Legal interpretations and court decisions will continue to shape the boundaries of copyright law in the context of AI-generated material.

If so, in what circumstances?

AI-generated outputs can implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the right to create derivative works, in various circumstances, depending on how AI-generated content is created and used. Here are some common circumstances in which these exclusive rights might be implicated:

- **Exact Copy or Reproduction:** AI-generated content that duplicates or reproduces preexisting copyrighted works, whether in whole or in substantial part, may implicate the right of reproduction. This can occur when AI systems generate content that closely resembles existing copyrighted material.
- **Transformation and Derivative Works:** AI-generated content that builds upon, adapts, or transforms preexisting copyrighted works may implicate the right to create derivative works. If AI systems incorporate copyrighted material and add new creative elements or commentary, it could be considered a derivative work.
- **Text Generation and Plagiarism:** AI-generated text that includes verbatim or paraphrased excerpts from copyrighted texts without proper attribution or authorization may implicate the right of reproduction. Plagiarism detection systems often address this issue.
- **Art and Visual Media:** AI-generated artwork, images, or videos that incorporate copyrighted visuals or elements

from preexisting works may implicate the right of reproduction or derivative work rights, especially if the AI-generated content closely resembles the original.

- **Fair Use and Transformative Use:** In some cases, AI-generated content may be protected by the fair use doctrine if it qualifies as transformative use. Fair use can provide a defense against claims of copyright infringement in situations where AI-generated content adds new meaning, purpose, or commentary to copyrighted works.
- **Licensing and Permissions:** AI-generated content that uses copyrighted material with the proper licenses or permissions may not implicate exclusive rights. Copyright owners can grant specific rights through licensing agreements.
- **Public Domain and Non-Copyrightable Material:** AI-generated content that incorporates materials from the public domain or non-copyrightable works, such as facts or ideas, generally does not implicate exclusive rights, as these materials are not protected by copyright.
- **Contractual Agreements:** Contractual agreements between AI developers and copyright owners can define the terms under which copyrighted material can be used in AI-generated content. Such agreements can specify whether exclusive rights are affected.

It's important to note that each case involving AI-generated content and copyright infringement is fact-specific, and legal outcomes can vary depending on factors such as the nature of the content, the extent of human involvement, and the application of fair use principles. Legal experts and courts play a significant role in interpreting copyright law in the context of AI-generated material, and their decisions help establish precedents and guidelines for specific circumstances.

23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

Further discussion on the Substantial Similarity Test provided by *Google, Generative AI Search beta*:

*Google, Generative AI Search beta*

[

The substantial similarity test is a copyright law test that courts use to determine if a work is infringing on another work. The test involves comparing the two works for similarities in their non-literal elements, such as plot, characters, dialogue, settings, and overall structure. The test is not a defense, but rather a way for courts to determine if a new work copies another copyrighted work.

The test is also known as the “total concept and overall feel” test. The courts consider substantial copying to be a matter of quality, not quantity. The courts must determine if the copying is quantitatively and qualitatively enough to be considered infringement

For example, in music, substantial similarity means that a listener can hear two songs side by side and tell that the allegedly infringing song borrowed material from the original

]

*CHAT-GPT*

Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

The "substantial similarity" test is a commonly used standard in copyright law to determine whether infringement has occurred. However, when applied to outputs from generative AI systems, it may face unique challenges and limitations. Here are some considerations:

## Challenges with the Substantial Similarity Test for AI-Generated Content:

- **Lack of Human Authorship:** AI-generated content lacks human authorship in the traditional sense, making it challenging to assess substantial similarity when the creative process is machine-driven.
- **Originality and Transformative Use:** Determining whether AI-generated content is substantially similar to a preexisting work may be complex, especially when AI systems transform or adapt existing material in novel ways.
- **Machine Learning and Data Input:** Generative AI systems often rely on vast datasets, and their outputs can be influenced by the input data. Assessing similarity may require considering the role of the training data.
- **Fair Use Considerations:** Fair use defenses may come into play, and the determination of whether AI-generated content qualifies as transformative may impact the assessment of substantial similarity.

Given these challenges, some legal experts and scholars have suggested that the substantial similarity test may need adaptation or supplementation when applied to AI-generated content. Here are some potential approaches:

## Potential Approaches for Assessing Copyright Infringement in AI-Generated Content:

- **New Criteria for Authorship:** Develop criteria for determining authorship in AI-generated works, considering factors like the extent of human involvement, the role of AI algorithms, and the novelty of the output.
- **AI-Specific Fair Use Guidelines:** Develop guidelines specific to AI-generated content that consider transformative



use, the nature of AI creativity, and the public interest in innovation and access to AI-generated material.

- **Transparency and Attribution:** Encourage transparency in AI systems to facilitate tracking and attribution of AI-generated content back to its sources, which can aid in assessing infringement claims.
- **AI-Specific Legislation:** Consider enacting legislation or regulations that specifically address AI-generated content, defining the scope of copyright protection and infringement in AI contexts.
- **Case Law Development:** Rely on legal precedents and court decisions to establish guidelines for assessing infringement in AI-generated content over time.

In summary, while the substantial similarity test remains a fundamental component of copyright law, it may need to be adapted or supplemented when applied to AI-generated content. The development of legal standards for AI-generated material is an evolving area of law, and legal experts, policymakers, and courts will play a crucial role in shaping the appropriate standards and guidelines for addressing copyright infringement claims related to AI-generated works.

24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?

*CHAT-GPT*

How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used?

Proving the element of copying in a copyright infringement case can be challenging, especially when the developer of the AI model does not maintain or make available records of what training material it used. In such cases, copyright owners may face



difficulties in demonstrating access to a copyrighted work. However, there are several strategies and legal avenues they can explore:

- **Direct Evidence of Access:**
  - If there is any direct evidence of access to the copyrighted work, such as correspondence, licensing agreements, or public statements by the AI model developer, copyright owners can present this evidence to establish access.
- **Reverse Engineering and Forensic Analysis:**
  - Copyright owners may employ forensic experts to reverse engineer or analyze the AI model to identify potential traces or similarities between the AI-generated content and their copyrighted work. This analysis can help establish a connection between the AI model and the copyrighted material.
- **Third-Party Witnesses:**
  - Identifying third-party witnesses who may have knowledge of the AI model's training data or access to copyrighted material can provide supporting testimony in court.
- **Discovery Process:**
  - During the discovery phase of a legal proceeding, copyright owners can request relevant documents, communications, and information from the AI model developer to determine if there is evidence of access.
- **Circumstantial Evidence:**
  - Copyright owners may use circumstantial evidence, such as the timing of the AI model's creation or the uniqueness of the AI-generated content, to argue that access to the copyrighted work is likely.
- **Pattern Recognition:**

- Copyright owners can use advanced pattern recognition technologies to identify similarities or distinctive patterns between the copyrighted work and the AI-generated content.
- **Expert Testimony:**
  - Expert witnesses in the fields of AI, machine learning, and copyright law can provide testimony regarding the likelihood of access and the relationship between the copyrighted work and the AI-generated content.
- **Presumptions and Inferences:**
  - Some legal systems may allow for certain presumptions or inferences when it is demonstrated that the AI model developer had a motive or opportunity to access the copyrighted work.

It's important to note that the specific strategies available to copyright owners may vary depending on the jurisdiction and the legal framework in place. Copyright owners should consult with legal experts experienced in copyright litigation to explore the most appropriate and effective methods for proving the element of copying in their particular case, especially when dealing with AI-generated content and limited records of training data.

Are existing civil discovery rules sufficient to address this situation?

Existing civil discovery rules in many jurisdictions provide a framework for obtaining relevant evidence, including evidence related to copyright infringement claims involving AI-generated content. These rules can be a valuable tool for copyright owners seeking to establish access to their copyrighted works. However, the effectiveness of these rules in addressing the situation where developers of AI models do not maintain or make available records of their training material may depend on various factors:

- **Scope of Discovery:** The scope of discovery rules may vary by jurisdiction, and in some cases, they may not explicitly cover certain types of electronic or digital records related to AI model training. Copyright owners may need to argue for the relevance and discoverability of specific evidence.
- **Cooperation of Parties:** The success of discovery efforts often depends on the cooperation of the parties involved. If the AI model developer is uncooperative or unresponsive to discovery requests, it can pose challenges for obtaining evidence related to access.
- **Technology and Expertise:** Effective discovery related to AI-generated content may require specialized knowledge and expertise in AI and machine learning technology. This expertise can be crucial in identifying relevant evidence and patterns.
- **Forensic Analysis:** In some cases, forensic analysis of the AI model or related systems may be necessary to uncover evidence of access. This can involve complex technical processes and may require court approval.
- **Legal Arguments:** Copyright owners may need to make legal arguments to persuade the court to order specific discovery related to AI model training data, especially if the AI developer raises objections or claims that certain information is not relevant or protected.
- **Privacy and Confidentiality:** Discovery rules often consider privacy and confidentiality concerns, especially when it comes to sensitive data related to AI models. Balancing these concerns with the need for evidence can be complex.
- **International Considerations:** When AI model developers are located in different jurisdictions, cross-border discovery issues may arise, requiring compliance with international legal procedures.

In summary, while existing civil discovery rules can provide a foundation for addressing the situation where records of AI model training material are not readily available, the effectiveness of these rules may vary based on jurisdiction, circumstances, and the cooperation of parties. Copyright owners and their legal counsel should carefully assess the specific requirements of their case and leverage available legal tools, technology, and expertise to address issues related to access and copyright infringement claims involving AI-generated content.

25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?

#### *CHAT-GPT*

Determining liability for copyright infringement involving AI-generated material can be complex and depends on the specific circumstances of each case. Liability may extend to multiple parties involved in the creation, distribution, or use of AI-generated content. Here are considerations for potential liability:

- **Developer of the Generative AI Model:**
  - **Primary Liability:** The developer of the AI model may be held primarily liable if they knowingly incorporated copyrighted material without authorization, especially if their actions constitute direct copying or reproduction.
- **Developer of the System Incorporating the Model:**
  - **Primary Liability:** The developer of the system that incorporates the AI model may also be held primarily liable if they knowingly facilitate or enable copyright infringement, such as by providing a platform for infringing activities.
- **End Users of the System:**

- **Secondary Liability (Contributory or Vicarious):** End users of AI systems that generate copyrighted content may face secondary liability if they use the system to infringe copyright knowingly or if they actively contribute to or benefit from the infringement. Liability may depend on whether they have control over the infringing activities or derive financial gain from them.
- **Content Platforms or Distributors:**
  - Content platforms or distributors that host or disseminate AI-generated content may face liability if they are aware of infringing content on their platform and fail to take reasonable steps to address it, which can be considered contributory infringement.
- **Data Providers:**
  - Entities that provide training data for AI models may be liable if they knowingly provide copyrighted material without authorization, and their data contributes to the AI-generated infringement.
- **Licensors and Copyright Owners:**
  - Copyright owners may choose to take legal action against licensors who provided AI developers with access to their copyrighted works without authorization.

It's important to note that liability determinations are highly fact-specific, and legal principles, such as contributory and vicarious liability, may vary by jurisdiction. Additionally, potential defenses, such as fair use or fair dealing, may be available to some parties.

Courts will consider factors like knowledge, intent, control, and financial benefit when assessing liability. Legal cases involving AI-generated content often involve nuanced legal arguments, and the outcomes may set important precedents for future cases.

Copyright owners pursuing claims of infringement involving AI-generated content should consult with legal experts to evaluate

the specific circumstances and parties involved in their case and to determine the most appropriate course of action.

25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs? [\[53\]](#)

*CHAT-GPT*

"Open-source" AI models can indeed raise unique considerations with respect to infringement based on their outputs, primarily due to their collaborative and transparent nature. Here are some key considerations:

- **Licensing Terms:** Open-source AI models are typically accompanied by specific licensing terms, which can vary widely. Some open-source licenses may grant broad usage rights, while others may have restrictions on commercial use, distribution, or modification. Users of open-source AI models must adhere to the terms of the applicable license.
- **Attribution and Share-Alike Clauses:** Some open-source licenses require users to provide attribution to the original developers and share any derivative works under the same open-source terms. Failing to comply with these clauses can lead to infringement claims.
- **Community Oversight:** Open-source projects often have active communities of developers and users who monitor and enforce licensing terms. Violations may be reported and addressed within these communities, potentially leading to legal action.
- **Public Availability:** Open-source AI models and their outputs are typically publicly available, making it easier for copyright owners to monitor and detect potential infringements. Copyright owners may actively search for unauthorized uses of their copyrighted works within open-source projects.

- **Derivative Works:** Users of open-source AI models may create derivative works or customizations based on the models. Whether these derivative works comply with open-source licenses or infringe on copyright depends on the specific terms and the nature of the modifications.
- **Fair Use and Transformative Use:** As with any copyrighted material, fair use and transformative use arguments may be raised as defenses against infringement claims. Users of open-source AI models may argue that their use falls within the boundaries of fair use if it involves criticism, commentary, or transformative creative elements.
- **International Considerations:** Open-source projects and their users often span multiple jurisdictions. Copyright laws and enforcement mechanisms can vary significantly from one jurisdiction to another, complicating the enforcement of copyright claims.

In summary, open-source AI models come with licensing terms and community oversight that can impact whether their outputs constitute copyright infringement. Users of open-source AI models should carefully review and comply with the associated licensing terms and consider potential copyright implications when using and modifying these models. Additionally, copyright owners should be aware of open-source projects that incorporate their copyrighted works and monitor compliance with licensing terms within these projects. Legal disputes involving open-source AI models may involve complex questions related to licensing, copyright, and community governance.



26. If a generative AI system is trained on copyrighted works containing copyright management information, how does [17 U.S.C. 1202\(b\)](#) apply to the treatment of that information in outputs of the system?

*CHAT-GPT*

**17 U.S.C. 1202(b)**, a provision within the Digital Millennium Copyright Act (DMCA), addresses the treatment of copyright management information (CMI) in copyrighted works. CMI includes information such as the copyright notice, author attribution, and other data that identifies and protects the rights of the copyright owner.

When a generative AI system is trained on copyrighted works containing CMI, several considerations arise:

- **Preservation of CMI:** AI systems should ideally preserve the CMI contained in copyrighted works used for training. This ensures that essential copyright information remains associated with the generated content, helping users and potential copyright owners identify the origin and ownership of the content.
- **Avoiding Removal or Alteration:** The AI system should not intentionally remove, alter, or suppress CMI during the generation process. The intentional removal or alteration of CMI could potentially lead to legal consequences under 17 U.S.C. 1202(b).
- **Applicability of 17 U.S.C. 1202(b):** If the AI system generates content that removes or alters CMI in a way that violates 17 U.S.C. 1202(b), it may be subject to legal action. This provision allows copyright owners to seek damages and injunctive relief if someone intentionally removes or alters CMI to facilitate copyright infringement.
- **AI Developers' Responsibility:** The developers of the generative AI system should be aware of the importance of



CMI and take measures to ensure that the system does not engage in actions that would violate 17 U.S.C. 1202(b). They may need to implement technical safeguards and ethical guidelines to address this concern.

It's essential for AI developers, users, and copyright owners to understand their respective rights and responsibilities under copyright law, including provisions like 17 U.S.C. 1202(b). Compliance with copyright law, proper attribution, and the preservation of CMI can help maintain a legal and ethical environment for AI-generated content. Additionally, jurisdiction-specific laws and international agreements may also play a role in determining the legal consequences of CMI-related issues in AI-generated content.

27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.

*CHAT-GPT*

Policymakers should consider several additional issues related to potential copyright liability based on AI-generated output:

- **Fair Use and Transformative Use:** Policymakers should clarify how fair use and transformative use principles apply to AI-generated content. Clear guidelines can help users and developers understand when AI-generated content qualifies as fair use, promoting innovation while respecting copyright.
- **Derivative Works:** Defining the boundaries of derivative works involving AI-generated content is essential. Policymakers can provide guidance on what constitutes a transformative modification versus an infringing derivative work.

- **Attribution and Transparency:** Encouraging transparency and proper attribution in AI-generated content is crucial. Policymakers can promote practices that ensure users and consumers can identify the origins and creators of AI-generated works.
- **Public Domain and Non-Copyrightable Material:** Policymakers should clarify the treatment of AI-generated content that incorporates public domain or non-copyrightable material, ensuring that such content remains free from copyright restrictions.
- **Licensing and Permissions:** Addressing licensing issues for AI-generated content can be challenging. Policymakers may consider mechanisms for facilitating licensing agreements and ensuring that content creators are fairly compensated.
- **Monitoring and Enforcement:** Policymakers should explore how copyright owners can monitor and enforce their rights in an era of AI-generated content, including the role of automated detection and takedown systems.
- **International Harmonization:** Given the global nature of AI, international harmonization of copyright rules and standards can help reduce legal complexity and promote consistency in addressing AI-generated content.
- **User Education:** Policymakers can support educational initiatives to help users, AI developers, and content creators understand their rights and responsibilities in the context of AI-generated content.
- **Intermediary Liability:** Clarifying the liability of intermediaries, such as AI platform providers and content-sharing platforms, for AI-generated content can be important for balancing innovation and copyright protection.
- **Ethical Considerations:** Policymakers should address ethical considerations surrounding AI-generated content,

including issues related to deepfakes, misinformation, and potential harm to individuals or communities.

- **Public Interest and Access:** Balancing copyright enforcement with the public interest in accessing AI-generated content is essential. Policymakers can consider how copyright law can promote access to AI-generated knowledge and creativity.
- **Legal Precedents:** Policymakers should monitor legal precedents and adapt copyright law as needed based on court decisions and evolving technological advancements.

Addressing these issues requires a balanced approach that fosters innovation and creativity while protecting the rights of content creators. Policymakers should engage with stakeholders, including copyright owners, AI developers, legal experts, and the public, to develop comprehensive policies that reflect the unique challenges posed by AI-generated content in the digital age.

### Labeling or Identification

28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

[

(I) we believe that *Description(s)* in Art, Fine Art and the wider field of the Arts are a Standardized method(s) facilitating the development of the *portfolio*, formal description of work, catalog raisonne and conservation. In this regard it may be within the standards and practices of the Arts to describe a work as created with or by AI. It is not legally required, however, that an Artist disclose a work is made with *oil paint*. In this way, a required label for AI-generated material would be unprecedented in my field. Further, a requirement would greatly limit access to and the potential of the technology.

]

*CHAT-GPT*

Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI?

The question of whether the law should require AI-generated material to be labeled or otherwise publicly identified as such is a complex and debated issue, and it involves considerations of transparency, ethics, and practicality. Here are some key arguments for and against such a requirement:

### **Arguments in Favor of Labeling AI-Generated Material:**

- **Transparency:** Labeling AI-generated content provides transparency to users and consumers, helping them differentiate between human-created and AI-generated works. This transparency can be essential for making informed decisions and understanding the source of information.
- **Accountability:** Labeling can hold developers and users of AI systems accountable for the content generated. It helps attribute responsibility for the content and its potential consequences.
- **Consumer Protection:** Labeling can protect consumers from potential deception or manipulation by AI-generated content, such as deepfakes or misinformation.
- **Preservation of Trust:** Maintaining trust in content and media is crucial for society. Labeling AI-generated material can help preserve trust in information sources.

### **Arguments Against Labeling AI-Generated Material:**

- **Chilling Effect on Creativity:** Mandatory labeling may discourage the development and use of AI for creative purposes. Artists and creators might be hesitant to explore new forms of artistic expression if their work is stigmatized as "AI-generated."
- **Technological Limitations:** Current AI may not always generate content that is easily distinguishable from human-

created work. Mandating labels could be challenging when the distinction is not clear-cut.

- **Overregulation and Burden:** Labeling requirements can lead to overregulation and administrative burdens on creators and users, especially if they apply broadly to all AI-generated content.
- **Dynamic Nature of AI:** AI technology evolves rapidly, and labeling requirements may become obsolete or overly prescriptive as AI capabilities advance.
- **Privacy and Anonymity:** Requiring labels could infringe on the privacy and anonymity of AI creators, which may be a concern in some contexts.
- **Fair Use and Transformative Works:** Labeling might not adequately address the nuances of fair use and transformative works, where AI contributes to creative processes but doesn't solely generate content.

Ultimately, the decision to require labeling of AI-generated material should consider a balanced approach that addresses transparency and accountability without stifling innovation and creativity. Policymakers, stakeholders, and the public should engage in thoughtful discussions to determine whether labeling is appropriate, when it should be required, and how it should be implemented, taking into account the evolving nature of AI technology and its impact on society.

If so, in what context should the requirement apply and how should it work?

If a requirement to label AI-generated material is deemed necessary, policymakers would need to consider the context and implementation carefully. Here are some key considerations for how such a requirement might apply:

### **Contexts for Labeling AI-Generated Material:**

- **Media and Content Sharing Platforms:** Labeling requirements could apply to platforms that host or distribute AI-generated content, ensuring that users are aware when content is generated by AI.
- **News and Information Sources:** In contexts where AI-generated content is used in news or information dissemination, labeling could provide transparency to consumers.
- **Artistic and Creative Works:** For artistic and creative works produced by AI, labeling could help audiences appreciate the role of AI in the creative process while preserving the distinction between human and AI contributions.

### Implementation Considerations:

- **Clear and Standardized Labels:** Labels should be clear, standardized, and easily understandable by the general public. They should convey that the content is AI-generated without stigmatizing it.
- **User-Friendly Display:** Platforms and content providers should display labels prominently and in a user-friendly manner, ensuring that consumers can readily identify AI-generated content.
- **Verification and Accuracy:** Mechanisms should be in place to verify that the labels accurately represent the nature of the content. This may require collaboration between content creators and platform providers.
- **Applicability Threshold:** Policymakers should define the threshold at which labeling is required. Not all AI-generated content may warrant labeling, and the requirement should be targeted where it provides the most value.
- **Educational Campaigns:** Alongside labeling, educational campaigns can inform users about the

implications of AI-generated content and how to critically evaluate it.

- **Privacy and Security:** Considerations for the privacy and security of AI creators should be addressed to prevent unintended consequences or harassment.
- **Fair Use and Transformative Works:** The requirement should account for situations where AI contributes to creative processes but does not solely generate content, such as in transformative works.
- **International Cooperation:** Given the global nature of the internet and AI-generated content, international cooperation and standards may be necessary for effective labeling.
- **Adaptability:** The labeling framework should be adaptable to evolving AI technology and capabilities.
- **Exemptions:** Policymakers may need to define exemptions or exceptions, such as cases involving privacy concerns or classified information.

The specifics of how a labeling requirement would work should be the subject of careful legislative or regulatory drafting, involving input from stakeholders, technology experts, legal scholars, and the public. The goal should be to strike a balance between transparency, accountability, and fostering AI innovation and creativity.

28.1. Who should be responsible for identifying a work as AI-generated?

*CHAT-GPT*

The responsibility for identifying a work as AI-generated can be distributed among various stakeholders, and it may depend on the context and the nature of the content. Here are some potential scenarios and the parties that could be responsible for labeling AI-generated works:



- **AI Developers or Creators:**
  - AI developers or creators who generate content using AI tools could be responsible for labeling their own creations. This approach places the onus on those with direct knowledge of the AI's involvement.
- **Content Platforms and Distributors:**
  - Platforms hosting or distributing AI-generated content may take on the responsibility of labeling such content. They can implement labeling mechanisms that automatically detect and display labels for AI-generated works.
- **Content Reviewers and Fact-Checkers:**
  - In contexts where AI-generated content is used in news or information dissemination, content reviewers, and fact-checkers may play a role in identifying AI-generated content. They can collaborate with platform providers to ensure accurate labeling.
- **Users and Consumers:**
  - In some cases, users and consumers may be encouraged or required to indicate when they believe content is AI-generated. Crowdsourced or community-based labeling mechanisms could be used.
- **Regulatory Authorities:**
  - Regulatory bodies or government agencies may establish guidelines or regulations that require content providers or platforms to label AI-generated works. They may also oversee compliance with labeling requirements.
- **AI Itself:**
  - In advanced AI systems, the AI itself may be capable of self-identifying as the creator of content. However, this would require significant technical sophistication and transparency mechanisms.
- **Collaboration Among Stakeholders:**



- A collaborative approach involving AI developers, platform providers, users, and regulatory authorities may be the most effective way to ensure comprehensive and accurate labeling.

The choice of responsible parties for labeling AI-generated works may vary depending on the specific context and the goals of the labeling requirement. It's important that labeling mechanisms are clear, standardized, and consistently applied to ensure transparency and accuracy for users and consumers.

Policymakers and stakeholders should work together to define the roles and responsibilities of various parties in the labeling process.

28.2. Are there technical or practical barriers to labeling or identification requirements?

*CHAT-GPT*

Yes, there are technical and practical barriers to implementing labeling or identification requirements for AI-generated works. Some of these challenges include:

- **Automated Detection Accuracy:** Developing automated systems that can accurately and reliably detect AI-generated content can be challenging, particularly when AI models are continually evolving and generating content that closely mimics human-created works.
- **Adversarial Attacks:** AI-generated content can be designed to bypass detection systems, leading to adversarial attacks. Adversarial techniques can make it difficult to identify AI-generated content.
- **Intermediary Platforms:** Requiring platforms to implement labeling systems for AI-generated content raises logistical and technical challenges. Platforms may need to develop and deploy new technology for content identification and labeling.

- **Diverse Use Cases:** AI-generated content spans a wide range of applications, from art and entertainment to news and information. Implementing a one-size-fits-all labeling system that works effectively across all use cases can be complex.
- **Privacy Concerns:** Some labeling mechanisms may inadvertently reveal personal or sensitive information about AI creators, raising privacy concerns. Balancing transparency with privacy is a challenge.
- **Human Review and Verification:** In cases where human review is involved in content identification, the process can be time-consuming and resource-intensive. Fact-checkers and content reviewers may not always be available to identify AI-generated works.
- **International Consistency:** Achieving international consistency in labeling requirements can be challenging due to variations in legal frameworks and cultural norms. Global standards and cooperation are needed.
- **Dynamic Nature of AI:** AI technology evolves rapidly, which means that labeling and detection systems must continually adapt to new AI capabilities and methods.
- **Implementation Costs:** Implementing and maintaining labeling systems can be costly for platforms and content providers, particularly smaller organizations with limited resources.
- **User Engagement:** Encouraging users to actively participate in labeling or reporting AI-generated content may require incentives or education to ensure meaningful participation.
- **False Positives and Negatives:** Automated detection systems can produce false positives (misidentifying human-created content as AI-generated) and false negatives (failing to identify AI-generated content). Balancing accuracy is a challenge.

Addressing these technical and practical barriers requires a collaborative effort involving AI developers, platform providers, regulatory authorities, and the broader technology and legal communities. Policymakers need to consider these challenges when designing labeling or identification requirements, aiming for solutions that are effective, privacy-respecting, adaptable, and user-friendly.

28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?

*CHAT-GPT*

The consequences of the failure to label a particular work or the removal of a label in the context of AI-generated content should be designed with careful consideration of various factors, including the nature of the content, the intent of the responsible party, and the potential impact on users and copyright holders. Here are some possible consequences that policymakers may consider:

- **Notice and Corrective Action:** When a work is identified as AI-generated but is not labeled as such, the responsible party could be required to promptly correct the labeling error upon notification. This approach emphasizes transparency and allows for corrective measures.
- **Warnings and Fines:** Policymakers may impose warnings and fines for repeated failure to label AI-generated content. Fines could be proportionate to the severity and frequency of violations. This approach provides financial incentives for compliance.
- **Removal or Deprioritization:** Platforms may be required to remove or deprioritize AI-generated content that is not properly labeled. This measure can help ensure that

users are not misled by content that lacks appropriate identification.

- **Loss of Legal Protections:** In certain cases, policymakers may consider limiting legal protections, such as safe harbors or immunities, for platforms that repeatedly fail to label AI-generated content. This approach could encourage platforms to take labeling seriously.
- **Injunctions:** Copyright holders or affected parties may seek injunctive relief through legal action to compel the labeling of AI-generated content that violates labeling requirements.
- **Liability for Misrepresentation:** Parties that intentionally misrepresent human-created content as AI-generated or vice versa could be subject to liability for misrepresentation or false advertising.
- **Education and Awareness:** In addition to penalties, policymakers may promote educational initiatives and awareness campaigns to inform creators, users, and platforms about their responsibilities regarding AI-generated content.
- **Appeal Mechanisms:** Policies could include mechanisms for appealing labeling decisions or penalties, allowing for the resolution of disputes and ensuring due process.
- **Privacy Protections:** Consequences should be designed to protect the privacy of AI creators and users. For example, any action that might reveal the identity of an AI creator should be taken with caution.
- **Graduated Enforcement:** A graduated enforcement approach could involve escalating consequences for repeated violations, starting with warnings and progressing to fines or other penalties for continued non-compliance.

It's essential that any consequences are proportionate, fair, and enforceable, and that they strike a balance between encouraging

compliance and avoiding undue burden on responsible parties. Policymakers should consult with stakeholders, including AI developers, content platforms, copyright holders, and legal experts, to establish a clear and effective framework for addressing labeling failures or violations.

29. What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?

*CHAT-GPT*

What tools exist or are in development to identify AI-generated material, including by standard-setting bodies?

Several tools and techniques are in development or already in use to identify AI-generated material, and some standard-setting bodies have taken an interest in this area. Here are some notable tools and initiatives:

- **Content Analysis Tools:** Various content analysis tools use machine learning and AI algorithms to detect anomalies that are characteristic of AI-generated content. These tools analyze factors such as linguistic patterns, image artifacts, and inconsistencies in media to flag potentially AI-generated material.
- **Reverse Image and Text Search:** Reverse image and text search engines allow users to input content and find visually or textually similar works online. These tools can help identify AI-generated content if it closely resembles known AI-generated samples.
- **Watermarking and Metadata:** Content creators and platforms can use digital watermarks and metadata to mark content as AI-generated. While not foolproof, these markers can provide information about the content's origin.

- **Blockchain and Distributed Ledger Technology:** Some initiatives explore the use of blockchain and distributed ledger technology to create tamper-proof records of content creation and ownership. This can help verify the authenticity of content.
- **AI Attribution Models:** AI-driven attribution models attempt to identify the AI model or system responsible for generating specific content. These models can be used to attribute content to its AI creator.
- **Image and Text Analysis APIs:** AI-based image and text analysis APIs provided by companies like Google, Microsoft, and IBM offer features that can help users detect AI-generated content by analyzing text or images for certain AI-generated characteristics.
- **Media Literacy and Education:** Public education campaigns and initiatives aim to improve media literacy and educate users about AI-generated content. These efforts encourage critical thinking when evaluating content authenticity.
- **OpenAI's GPT-3 "CLIP":** OpenAI's CLIP (Contrastive Language-Image Pre-training) model can be used to identify AI-generated text or images. It has been used to create AI tools that flag AI-generated content.
- **AI Ethics and Guidelines:** Standard-setting bodies and organizations, such as the IEEE (Institute of Electrical and Electronics Engineers), are developing guidelines and ethics frameworks related to AI transparency and disclosure, which can inform the identification of AI-generated content.
- **Legislative and Regulatory Efforts:** Some legislative efforts propose requirements for platforms to identify or label AI-generated content, which may lead to the development of standardized tools and practices.

It's important to note that no single tool or technique is foolproof in identifying AI-generated material, especially as AI models become

more advanced. A combination of these tools, human review, and public awareness efforts may be necessary to effectively address the challenge of AI-generated content identification. Standard-setting bodies and industry collaboration may play a role in developing best practices for content identification in the future.

How accurate are these tools?

The accuracy of tools designed to identify AI-generated material can vary widely depending on the specific tool or technique, the sophistication of the AI model used, and the nature of the content being analyzed. Here are some key factors that influence the accuracy of such tools:

- **AI Model Sophistication:** The accuracy of identification tools often depends on the capabilities of the underlying AI model. More advanced and sophisticated AI models are better at mimicking human-created content, making it harder to distinguish AI-generated material from human-created material.
- **Training Data:** The effectiveness of AI models and tools relies on the quality and diversity of the training data. Models trained on a wide range of content types and styles are generally more accurate at generating content that closely resembles human-created material.
- **Content Complexity:** Simple and straightforward content may be easier for identification tools to classify accurately, while complex and nuanced content can pose greater challenges.
- **Media Type:** Identification tools may perform differently for different types of media, such as text, images, audio, or video. Some tools may be more specialized for specific media types.



- **Adversarial Attacks:** As AI-generated content becomes more prevalent, creators may design content specifically to evade detection. Adversarial techniques can make it difficult for identification tools to accurately flag AI-generated content.
- **Continual Model Updates:** AI models evolve and improve over time. An identification tool that was accurate for one generation of AI models may become less accurate as new, more advanced models emerge.
- **False Positives and Negatives:** Identification tools may produce false positives (identifying human-created content as AI-generated) and false negatives (failing to identify AI-generated content). Balancing these errors is a challenge.
- **Interdisciplinary Nature:** Identifying AI-generated content often requires interdisciplinary expertise, involving AI researchers, linguists, computer vision experts, and media analysts.
- **Contextual Analysis:** Accurate identification may require considering the broader context of the content, such as its source, style, and purpose.

Overall, while some identification tools can be quite accurate in specific contexts, there is no one-size-fits-all solution for accurately identifying AI-generated material, particularly as AI models become more advanced. Users and organizations should be aware of the limitations of these tools and consider combining automated detection with human review and critical thinking when evaluating content authenticity. Continued research and development in this area are essential to improving the accuracy of identification tools as AI technology evolves.



What are their limitations?

Tools and techniques designed to identify AI-generated material have several limitations, which can make accurate detection challenging. Here are some common limitations associated with these tools:

- **Advanced AI Models:** As AI models become more sophisticated, they can produce content that closely mimics human-created material, making it difficult to distinguish AI-generated content from human-generated content.
- **Evolution of AI Models:** AI models are continually evolving and improving. Identification tools that are effective against one generation of models may struggle to detect content generated by newer, more advanced models.
- **Content Complexity:** Highly complex or nuanced content may pose challenges for identification tools, as the differences between AI-generated and human-created material become subtler.
- **Adversarial Attacks:** Creators may intentionally design content to evade detection, using adversarial techniques to make it difficult for identification tools to accurately flag AI-generated content.
- **False Positives and Negatives:** Identification tools may produce false positives (misidentifying human-created content as AI-generated) and false negatives (failing to identify AI-generated content). Achieving a balance between these errors can be challenging.
- **Media Type Specificity:** Some identification tools are specialized for specific media types, such as text, images, audio, or video. They may not be equally effective across different media.
- **Contextual Analysis:** Accurate identification may require considering the broader context of the content,

including its source, style, and purpose, which can be complex and context-dependent.

- **Privacy Concerns:** Some identification methods may inadvertently reveal information about the AI creators, potentially raising privacy and ethical concerns.
- **Legal and Ethical Challenges:** The deployment of AI identification tools may raise legal and ethical questions related to privacy, freedom of expression, and content moderation.
- **Dynamic Online Environments:** Online platforms and content-sharing environments are dynamic, with vast amounts of content generated continuously. Keeping up with new content and evolving AI models is a significant challenge.
- **Interdisciplinary Expertise:** Accurate identification often requires expertise in multiple disciplines, including AI, linguistics, computer vision, and media analysis.
- **Generalization:** Identification tools may not generalize well across all types of AI-generated content, and their effectiveness may vary depending on the specific characteristics of the content.
- **International Variability:** Identification requirements and expectations may vary across different countries and regions, making it challenging to establish uniform identification practices.

Given these limitations, it's important for users, platforms, and content providers to exercise caution and critical thinking when evaluating the authenticity of content. While identification tools can be valuable aids, they should be used in conjunction with human judgment and a nuanced understanding of the evolving landscape of AI-generated content. Continued research and development in this field are essential to address these limitations and enhance the accuracy of identification tools.

## Additional Questions About Issues Related to Copyright

Continued use of Generative AI Learning Language Models provides access where the user or organization is without the operational purview but affected by the potential liability and ramification of the related question(s) and AI-generative response. Tavarus Blackmon Art LLP generates “responses,” with the use of CHAT-GPT.

30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?

*CHAT-GPT*

The legal rights that apply to AI-generated material featuring the name or likeness, including vocal likeness, of a particular person can vary depending on jurisdiction and the specific circumstances. Here are some legal rights and considerations that may apply:

- **Right of Publicity:** The right of publicity is a legal right that grants individuals control over the commercial use of their name, likeness, voice, and other identifying attributes. This right is designed to prevent the unauthorized use of a person's identity for commercial gain. Laws governing the right of publicity vary by jurisdiction.
- **Privacy Laws:** Some jurisdictions have privacy laws that protect individuals from the unauthorized use of their likeness or voice, particularly in situations where privacy is violated or harm is caused by the use.
- **Defamation and False Light:** If AI-generated content featuring a person's likeness includes false or defamatory statements, the individual may have a legal claim for defamation or false light invasion of privacy.
- **Intellectual Property Rights:** Intellectual property laws, such as copyright and trademark, can be relevant when AI-generated material incorporates a person's name, image, or likeness. For example, a person may have

copyright in their own photographs, which could be implicated if used without permission.

- **Contractual Agreements:** Contracts and agreements between individuals and AI developers or content creators may specify the rights and permissions related to the use of the person's likeness, name, or voice.
- **Transformative Use and Fair Use:** In some cases, the use of a person's likeness or identity in AI-generated content may be considered transformative or fall under fair use defenses, depending on the nature and purpose of the content and the applicable jurisdiction.
- **Public Domain:** If a person's likeness or voice is in the public domain, it may be used without restriction. However, the rules regarding what constitutes the public domain can vary.
- **First Amendment Considerations:** In the United States, First Amendment rights can play a role in determining the legal boundaries of using a person's likeness in creative expression, particularly in cases involving art, parody, or political commentary.

It's essential to consult legal experts and consider the specific laws and regulations in your jurisdiction when dealing with AI-generated material featuring a person's likeness. The legal landscape for AI-generated content is evolving, and the application of existing laws to these situations can be complex and fact-specific. Additionally, the rights and protections afforded to individuals may vary depending on whether the AI-generated content is used for commercial purposes, artistic expression, or other contexts.

31. Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?

*CHAT-GPT*

Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material?

The question of whether Congress should establish a new federal right specifically tailored to AI-generated material, similar to state law rights of publicity, is a complex and debated legal issue. Establishing such a federal right would require careful consideration of various factors, including the potential benefits and drawbacks. Here are some key considerations:

#### **Potential Benefits:**

- **Uniformity:** A federal right could provide uniformity and clarity to the treatment of AI-generated material across all states, simplifying legal processes and reducing jurisdictional disparities.
- **Protection for Individuals:** A federal right could enhance the protection of individuals' names, likenesses, and vocal likenesses in the context of AI-generated content, addressing the unique challenges posed by AI technology.
- **Innovation and Commercialization:** Clearer legal protections could encourage innovation and investment in AI technology by providing certainty to creators and developers regarding the use of individuals' identities in AI-generated content.

#### **Potential Drawbacks:**

- **First Amendment Concerns:** Balancing the right to control one's likeness with First Amendment rights to free expression can be challenging. Creating a federal right

would require careful consideration of these constitutional issues.

- **Overregulation:** Some argue that creating a new federal right may lead to overregulation, potentially stifling creative and transformative uses of AI technology for artistic, cultural, or political purposes.
- **Complexity:** The creation of a new federal right could add complexity to the legal landscape, potentially leading to litigation over the scope and application of this right.
- **Technological Evolution:** AI technology evolves rapidly. Creating a federal right that remains relevant as technology advances would require flexibility and adaptability in the law.
- **International Considerations:** Any new federal right would need to be harmonized with international legal norms and agreements, as AI-generated content often transcends national boundaries.

Ultimately, the decision to establish a new federal right for AI-generated material would require thorough analysis, input from legal experts, stakeholders, and policymakers, and a careful balancing of the interests of individuals, creators, and society at large. It would also need to account for the dynamic nature of AI technology and its impact on various industries and forms of expression.

If so, should it preempt state laws or set a ceiling or floor for state law protections?

If Congress were to establish a new federal right for AI-generated material similar to state law rights of publicity, it would need to determine whether this federal right should preempt state laws and what relationship it should have with existing state law protections. There are several possible approaches:

- **Preemption:** Congress could decide to preempt state laws on the same subject matter, meaning that the federal law would replace or supersede state laws. This approach would provide a uniform nationwide standard and simplify legal proceedings.
- **Floor or Minimum Standard:** Congress could establish the federal law as a minimum standard, allowing states to enact additional protections or variations that exceed the federal baseline. This approach would maintain some level of state flexibility while ensuring a basic level of protection at the federal level.
- **Ceiling or Maximum Standard:** Congress could set the federal law as a maximum standard, meaning that states would not be allowed to enact protections beyond what the federal law provides. This approach would limit state-level variations and create a uniform ceiling of protection.
- **Federal Default with Opt-Out:** Congress could create a federal default law that applies unless a state opts out by enacting its own law on the same subject. This approach would allow states to choose whether to adopt the federal standard or establish their own.

The choice among these approaches would depend on various factors, including the goals of the federal legislation, the degree of uniformity desired, and the balance between federal and state authority. Each approach has advantages and disadvantages, and policymakers would need to carefully consider the legal, practical, and policy implications of their decision.

Moreover, any federal legislation would need to provide clear guidance on how it interacts with existing state laws and how conflicts between state and federal law would be resolved. It's also important to consider the impact on individuals' rights, the creative industries, and technological innovation when crafting



such legislation. Public input, legal expertise, and thorough deliberation would be essential in the legislative process.

What should be the contours of such a right?

The contours of a federal right for AI-generated material, similar to state law rights of publicity, would need to be carefully defined in the legislation. These contours would depend on the policy goals of the legislation and the balance between protecting individuals' rights and allowing for legitimate uses of AI technology. Here are some key considerations for defining the contours of such a right:

- **Scope of Protection:** Determine what aspects of a person's identity would be protected, including their name, likeness, voice, and other identifying attributes.
- **Duration:** Specify the duration of protection, which could be for the individual's lifetime, a certain number of years after their death, or other defined periods.
- **Exceptions and Defenses:** Establish exceptions and defenses that allow for the legitimate use of AI-generated content, such as for news reporting, commentary, parody, satire, art, and other forms of expressive and transformative uses.
- **Transformative Use:** Clarify the extent to which transformative use of a person's identity in AI-generated content is allowed and under what conditions.
- **Commercial vs. Non-Commercial Use:** Distinguish between commercial and non-commercial uses of AI-generated material and determine the scope of protection for each category.
- **Public Domain:** Specify whether certain uses of an individual's identity in AI-generated content would be allowed if that identity is in the public domain.



- **First Amendment Considerations:** Address how the federal right would balance individuals' rights with First Amendment protections for free expression and artistic freedom.
- **Enforcement and Remedies:** Outline the mechanisms for enforcing the right, including potential remedies for violations, such as injunctive relief, damages, or statutory damages.
- **Jurisdiction and Conflicts:** Clarify how the federal right interacts with state laws, particularly in cases where states have existing rights of publicity or related laws.
- **User Consent:** Consider whether the legislation should include provisions related to obtaining informed consent from individuals for the use of their identity in AI-generated content.
- **Technological Evolution:** Design the legislation to be adaptable to the evolving nature of AI technology and its impact on content creation and distribution.
- **Privacy Considerations:** Address privacy concerns related to the collection and use of individuals' personal data in the creation of AI-generated content.
- **International Harmonization:** Ensure that the legislation is harmonized with international norms and agreements to address AI-generated content that crosses national boundaries.

Defining the contours of a federal right for AI-generated material would require extensive legal and policy analysis, as well as input from experts, stakeholders, and the public. Striking a balance between protecting individuals' rights, fostering creativity, and encouraging innovation would be a central challenge in crafting such legislation.

32. Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)? Who should be eligible for such protection? What form should it take?

*CHAT-GPT*

Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)?

The question of whether there should be protections against AI systems generating outputs that imitate the artistic style of a human creator, such as producing visual works "in the style of" a specific artist, is a complex and debated issue at the intersection of technology, art, and intellectual property. Here are some considerations:

### **Protections for Human Creators:**

- **Attribution:** If an AI system generates works that closely mimic the style of a human artist, there may be concerns about the potential for misattribution. Protections could ensure that viewers or consumers are informed about the AI-generated nature of the work to avoid misleading associations with a human creator.
- **Moral Rights:** Some legal systems recognize moral rights, which can include the right to attribution and the right to object to derogatory treatment of one's work. Artists may argue that their moral rights should extend to protecting the unique style they have developed from being replicated by AI.

### **Innovation and Creative Freedom:**

- **Fostering Innovation:** AI systems that can replicate artistic styles have the potential to foster innovation and

creativity by allowing artists to experiment with new styles or reinterpret existing ones.

- **Fair Use and Parody:** Protections should take into account the principles of fair use and parody, allowing for transformative uses of artistic styles for purposes such as commentary, criticism, or satire.

### **Balancing Interests:**

- **Public Domain:** Protections should consider whether an artistic style can enter the public domain, allowing for broader creative expression without restriction.
- **Impact on Artists:** Assess the potential impact on human artists and their ability to create distinctive styles and establish themselves in the art world.
- **Transparency:** Ensure transparency regarding the use of AI in the creation of artworks to maintain trust between creators, consumers, and the AI technology itself.
- **Cultural and Historical Context:** Recognize that styles and artistic movements often build upon or reinterpret existing traditions. Protections should account for the historical and cultural context of art.

The approach to protections against AI-generated imitations of artistic styles is a nuanced one, and it involves finding a balance between preserving the rights and creative expressions of human artists while also fostering technological innovation and creative freedom. Legal frameworks and industry best practices may need to evolve to address these challenges. Public discourse and collaboration among artists, technologists, policymakers, and legal experts can help shape the development of such protections.

Who should be eligible for such protection?

Determining who should be eligible for protection against AI-generated imitations of artistic styles involves defining the scope of protection and considering the rights of various stakeholders. Here are some key considerations regarding eligibility:

- **Human Artists:** Primary eligibility for protection is often extended to human artists who have developed unique and identifiable artistic styles. These artists may include painters, sculptors, musicians, writers, and creators in various artistic disciplines.
- **Living and Deceased Artists:** Protection can apply to both living and deceased artists. While living artists may seek protection during their lifetimes, estates or heirs of deceased artists may also seek to protect the legacy and artistic style of the artist.
- **Artistic Estates:** In cases where an artist's estate manages their artistic legacy, they may be eligible for protection to ensure the continued recognition and preservation of the artist's style.
- **Collectives and Collaborators:** Artists who work in collectives or collaborate with others may seek protection for their collective artistic style or the style that emerges from their collaborations.
- **AI Developers:** AI developers and creators of AI systems may also have a stake in this discussion. They may seek protection for the AI-generated style itself or for their contributions to the development of AI systems capable of replicating artistic styles.
- **AI-Generated Works:** There may be discussions about whether AI-generated works that closely imitate artistic styles should be eligible for protection, particularly if there is creative input from the AI system itself.

- **Public Domain Works:** Consideration should be given to works and artistic styles that have entered the public domain, as they may not be eligible for the same level of protection.

Eligibility for protection in the context of AI-generated imitations of artistic styles involves striking a balance between recognizing and safeguarding the creative contributions of human artists and considering the role of AI technology in art creation. The eligibility criteria would likely be shaped by legal frameworks, intellectual property laws, and policy decisions made by governments and cultural institutions. It's important for these decisions to be informed by public discourse, stakeholder input, and a clear understanding of the impact on the arts and creative industries.

What form should it take?

The form of protection against AI-generated imitations of artistic styles would depend on the legal and policy frameworks established by governments and cultural institutions. There are several possible forms of protection that could be considered:

- **Attribution and Transparency Requirements:** One approach is to require clear attribution and transparency in AI-generated artworks. This could involve labeling or marking AI-generated art as such to inform viewers or consumers that the work was created with the assistance of AI technology.
- **Moral Rights:** Expanding existing moral rights to include protection against the imitation of an artist's style by AI could be considered. Moral rights can include the right to attribution and the right to object to derogatory treatment of one's work.
- **Copyright and Related Rights:** Copyright law could be adapted or extended to cover the unique artistic style

developed by a human artist. This could include recognizing the artist's style as a protected aspect of their copyrighted works.

- **Contracts and Licensing:** Artists and AI developers could negotiate contracts and licensing agreements to define the terms under which AI-generated works that mimic an artist's style can be created, used, and attributed.
- **Best Practices and Guidelines:** Artistic and cultural institutions, as well as industry organizations, could develop best practices and guidelines for the responsible creation and use of AI-generated art, including considerations for artistic style.
- **Creative Commons Licenses:** Some artists and creators may choose to use Creative Commons licenses that specify the terms under which their works can be used, including AI-generated adaptations.
- **Public Domain Considerations:** Legal frameworks should also consider the treatment of artistic styles that have entered the public domain and whether such styles can be imitated by AI without restriction.

The specific form of protection chosen would depend on the legal and cultural context of each jurisdiction, as well as the policy goals of protecting artists' rights while fostering innovation and creativity. It's important for discussions around this issue to involve artists, legal experts, AI developers, and other stakeholders to find solutions that balance the interests of all parties involved. Public input and open dialogue can help shape the appropriate form of protection for AI-generated imitations of artistic styles.

33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws? <sup>[54]</sup> Does this issue require legislative attention in the context of generative AI?

*CHAT-GPT*

With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws?

Section 114(b) of the Copyright Act is primarily concerned with the rights and royalties associated with digital audio transmissions of sound recordings. It is a federal provision that addresses specific aspects of copyright law related to the digital transmission of sound recordings and webcasting, and it generally does not directly pertain to state right of publicity laws. However, there may be some indirect interactions or overlaps between federal copyright law and state right of publicity laws in certain situations:

- **Exclusive Rights:** Section 114(b) of the Copyright Act outlines the exclusive rights of sound recording copyright owners in the context of digital audio transmissions. These rights include the right to perform the copyrighted sound recording publicly by means of digital audio transmission. State right of publicity laws, on the other hand, primarily focus on protecting individuals' rights to control the commercial use of their names, likenesses, and voices.
- **Preemption:** Copyright law, including Section 114(b), is generally considered to be a federal matter and can preempt state laws in certain cases where there is a conflict or inconsistency between federal copyright law and state law. However, right of publicity laws typically govern different aspects of intellectual property, specifically the use of an individual's identity for commercial purposes.
- **Overlap:** There can be situations where the use of sound recordings in digital transmissions involves both copyright issues and right of publicity issues. For example, if



a digital transmission includes both a sound recording and the use of a celebrity's voice or likeness, there may be potential interactions between copyright law and right of publicity law.

- **Licensing and Permissions:** In practice, content creators, broadcasters, and digital streaming services often need to consider both copyright law and right of publicity laws when using sound recordings, especially if the recordings feature the voices or likenesses of individuals. Obtaining proper licenses and permissions is a common practice to address these legal considerations.

Overall, while Section 114(b) of the Copyright Act primarily deals with copyright issues related to sound recordings and digital transmissions, it is essential for those in the entertainment and media industries to be aware of and navigate the intersection of federal copyright law and state right of publicity laws when creating and distributing content that involves sound recordings and individuals' identities. Legal advice and compliance with relevant state laws are crucial in such cases.

Does this issue require legislative attention in the context of generative AI?

The issue of how generative AI interacts with copyright law and related rights, such as right of publicity, is a complex and evolving area of law. While there may not be a specific need for new legislation related solely to generative AI at the federal level, it's clear that this emerging technology poses unique challenges and opportunities for copyright and intellectual property law. Here are some reasons why legislative attention and updates may be necessary:

- **Clarification of Rights:** Legislation could help clarify how copyright and related rights apply to AI-generated



content, including issues related to authorship, ownership, and protection.

- **Fair Use and Transformative Use:** Legislative guidance may be needed to address the application of fair use and transformative use doctrines to AI-generated works, including whether AI-generated parodies or transformative works should be protected.
- **Rights of Attribution and Transparency:** Legislation could establish standards for attribution and transparency in AI-generated content, ensuring that consumers are informed about the nature of the content they are engaging with.
- **Liability:** Legislation might address questions of liability, such as who is responsible for AI-generated content that infringes on copyright or right of publicity.
- **International Harmonization:** As AI-generated content can cross international borders easily, legislative efforts may aim to harmonize copyright and AI-related laws globally.
- **Privacy Considerations:** Legislation may be necessary to address privacy concerns related to the use of personal data in AI models, especially if such data is used to create content that imitates individuals.
- **Access to Data:** Legislators may need to consider issues related to access to training data and datasets, especially in cases where public domain or publicly accessible data is used for AI training.
- **Public Input:** Legislative processes provide opportunities for public input and the involvement of stakeholders from the technology, entertainment, and legal sectors to shape the legal framework governing AI-generated content.

While existing copyright and intellectual property laws can apply to AI-generated content to some extent, the rapid advancement of AI technology and its impact on creative industries and rights-holders may warrant legislative updates and clarifications in the

future. Policymakers and legal experts are actively studying these issues to strike a balance between protecting rights and fostering innovation in the context of generative AI. Public discourse, stakeholder engagement, and interdisciplinary collaboration are essential components of this ongoing dialogue.

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

*CHAT-GPT*

In conducting its study on AI-generated content and copyright, the Copyright Office should consider a range of complex and evolving issues. Here are some additional areas and topics that warrant attention:

- **Orphan Works:** The Copyright Office may explore how AI can be used to identify and manage orphan works—works for which the copyright owner is unknown or cannot be located—and the potential role of AI-generated content in reviving orphaned works.
- **Machine Learning Fairness:** Consideration of issues related to bias and fairness in AI systems, including how AI models might perpetuate or mitigate biases present in training data and the legal implications of such biases in AI-generated content.
- **Data Privacy:** Exploration of data privacy and consent considerations when personal data is used to train AI models, especially when AI-generated content involves individuals' likenesses or personal information.
- **Ethical Use and Guidelines:** Discussion of ethical considerations and best practices for the responsible use of AI-generated content, including guidance on how AI should be used in creative, commercial, and public contexts.

- **User-Generated Content:** Examination of how AI-generated content interacts with user-generated content and platforms that host such content, including issues of moderation, intellectual property, and the role of platforms in content creation.
- **Creative Collaborations:** Analysis of the legal and creative implications of collaborations between human artists and AI systems, particularly in cases where artists use AI as a tool in their creative process.
- **International Perspectives:** A look at how other countries are addressing the intersection of AI and copyright, and consideration of international harmonization efforts.
- **Education and Outreach:** Consideration of the need for educational initiatives and outreach programs to inform creators, consumers, and the public about the legal and ethical aspects of AI-generated content.
- **Liability and Enforcement:** Examination of the roles and responsibilities of various parties, including AI developers, users, and online platforms, in preventing and addressing copyright infringement involving AI-generated content.
- **Long-Term Preservation:** Discussion of the challenges and opportunities for the long-term preservation of AI-generated works and the copyright implications of archiving and preserving such content.
- **Collaborative Research:** Encouragement of collaborative research initiatives among stakeholders, including legal experts, technologists, artists, and policymakers, to address emerging challenges and solutions.
- **AI in Education:** Exploration of how AI-generated content is used in educational settings and the copyright and educational policy implications.

- **AI in Journalism:** Consideration of the use of AI-generated content in journalism, including its role in automating news reporting and potential copyright issues in news content created by AI.

These topics represent some of the multifaceted issues that the Copyright Office should consider as it conducts its study on AI-generated content and copyright. The interplay between AI technology, intellectual property law, ethics, and societal impact is complex and requires a comprehensive examination to develop informed policies and guidelines for the future. Public input, collaboration, and interdisciplinary perspectives are essential components of this process.

End AI-generative responses

The questions presented and our responses constitute work in the filed both personal and academic. That in the process of earning degrees in Film and Studio Art and Founding our Organization, utilizing this technology has aided in the process of creating Fine Art. Utilizing this technology facilitates learning and positive outcomes for individuals with disabilities and impairments. Works created with AI-generative Models, tools and algorithms can and sometimes must, meet a threshold or criteria in the establishment of authenticity, Authorship or Registration in certain fields. It may not be Internationally necessary or require legislation, however. As an Organization working closely with this technology we believe it is *capable* of “expressing,” *beautiful*, or valuable “work,” outputs or, “expressions.” These are autonomous and meriting conservatory efforts. Such work, *collaboratively*, generated by a human user or an organization can be considered protected work as part of a larger US Copyright. Public Protections and Responsible Development and Innovation including the wider filed of the Arts and DEI accessibility, integration and interoperability, are what we believe to be the issues paramount to any potential legislation. As US Copyright holders with a growing Copyright Portfolio, we seek to work with other organizations, State Representatives, Federal Offices and Developers, Artists, Educators and Critical Thinkers around the topic of AI, including how it relates to emerging Blockchain Technology.

Tavarus Blackmon Art LLP

Tavarus (Blackmonster❤️) Blackmon, MA, MFA

<https://tavarusblackmonart.com>

Part-Time Faculty, CSU Sacramento

DEI Committee Chair, BOD, Root Division

Art and Art History Advisory Board of Directors, UC Davis

*Tavarus Blackmon*