FROM CANVAS TO CODE: ANALYSING THE GENERATIVE AI AND FAIR USE THROUGH THE LENS OF THE ANDY WARHOL VERDICT

VAISHNAVI CHANDRAKAR[*]

ABSTRACT

In an ever-evolving landscape where technology meets creativity, the clash between artificial intelligence, copyright law and artistic expression has prompted a revaluation of the concept of 'transformative use' which falls within the defence of fair use. The 2023 verdict of Andy Warhol Foundation v. Goldsmith redefined transformative use and has inspired new interpretations. This paper delves into the evolving concept of 'transformative use' and Generative AI within the context of fair use analysis. The paper aims to dissect the impact of the case's redefined definition of 'transformative use' on Generative AI. It analyses the contours of the verdict, scrutinizing prevailing jurisprudence on AI, Machine Learning and Fair Use and reframing them to suit the framework for contemporary technological advancements. The research paper sheds light on the complex interplay and balance between the Doctrine of fair use, artistic innovation and AI-generated content in the modern copyright landscape.

I. INTRODUCTION

"Success in creating effective AI could be the biggest event in the history of our civilization. Or the worst. We just don't know. So, we cannot know if we will be infinitely helped by AI or ignored by it and side-lined or conceivably destroyed by it."

– Stephen Hawking

An artist is someone who possesses the creativity and skills to create original works of art. In the realm of technological advancement, the advent of 'Generative AI' marks a pivotal moment for artists, as it introduces an array of approaches that redefine the creative process. Generative AI involves advanced models using deep learning to create diverse content like text and images, drawing from extensive training data to produce entirely new material based on learned patterns and information. The rise of AI art, showcased by technologies such as OpenAI's Dall-E 2, has both fascinated and unsettled observers. These advancements swiftly generate accurate and sometimes surreal images, like Kermit in the style of Edvard Munch or Gollum enjoying

---

[*] Fourth Year Undergraduate Student, (B.A., LL.B. (Hons.), Hidayatullah National Law University, Raipur.

watermelon.[1] As Stephen Hawking observed, the impact of successful AI creation remains uncertain, possessing the potential to shape the course of civilization in unprecedented ways. This uncertainty extends into the creative sphere as well, where artists navigate the intersection of traditional tools and innovative AI-generated images. In this evolving landscape, artists adopt various approaches—some harness the power of traditional graphic design tools to enhance AI-generated images, while others draw inspiration from these AI creations, birthing entirely new works through non-AI means. However, beneath these creative pursuits lies a complex web of legal and ethical considerations, the core question about fair use persists: *whether AI models trained on copyrighted material are encompassed within this legal doctrine?*

The **Andy Warhol Foundation for the Visual Arts v. Goldsmith Case**[2] has ignited a series of inquiries that extend beyond its immediate context. Fair use constitutes a vital aspect of copyright law, permitting individuals to utilize copyrighted material within certain limitations without seeking explicit permission. In the USA, fair use is determined by four key factors: the purpose and character of the use, the nature of the original work, and the amount and substantiality of the portion employed.[3] Transformative use, a subset of fair use, involves using the copyrighted material in a distinct manner and purpose from the original, adding new expression and meaning to copyrighted work. This paper aims to ascertain and analyse the bounds of 'transformative use,' reconciling the case and evaluating its alignment with earlier copyright decisions and balancing the innovation in AI and incentive to create for the copyright holders. The case will undoubtedly impact the application of the fair use doctrine to generative AI platforms.

II.   CONTEXTUALIZING WARHOL V. GOLDSMITH DISPUTE

In the recent *Andy Warhol Foundation for the Visual Arts v. Goldsmith Case*, the Supreme Court proposed a reasonable necessity standard for transformative fair-use analyses[4]. This copyright case involves two artists: Andy Warhol and Lynn Goldsmith. Andy Warhol, an American artist and filmmaker, is widely recognized for his transformative impact on contemporary art. His fame is anchored in iconic works featuring everyday items such as Campbell's soup cans and portraits of celebrities like Marilyn Monroe. He was a leading exponent of the 'Pop art' movement and elevated

---

[1] Lauri Clarke, *When AI can make art – what does it mean for creativity?*, THE GUARDIAN, (Dec. 13, 2023), https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney

[2] Andy Warhol Foundation for the Visual Arts Inc v. Goldsmith, 598 U. S. 508 (2023).

[3] *U.S. Copyright Fair Use Index*, USA COPYRIGHT OFFICE (Apr. 4, 2024), https://www.copyright.gov/fair-use/.

[4] Peter J Carol, *The transformative impact of Warhol v. Goldsmith - Artforum International*, ARTFORUM (Nov. 12, 2023), https://www.artforum.com/slant/the-transformative-impact-of-warhol-v-goldsmith-90667.

ordinary images to iconic status, prompting contemplation on the creation through art. Meanwhile, Goldsmith is a pioneer celebrity and rock and roll photographer. She documented stars from The Beatles and The Rolling Stones to the eclectic personas of David Bowie and Iggy Pop.[5]

In 1984, Vanity Fair, a magazine, secured a license from Lynn Goldsmith, which granted permission to use one of her photographs, the Portrait of Prince, in an article. In 1981, Lynn Goldsmith photographed a series of images of the emerging musical talent, Prince. Three years later, following the success of Prince's 'Purple Rain' in 1984, Vanity Fair, a Condé Nast publication, secured a license for one of Goldsmith's previously unpublished photographs. The specific black and white full-length portrait was intended for a planned feature, with the understanding that it would serve as an "artistic reference" for a unique illustration to be used only once. To bring this vision to life, the magazine hired the renowned pop-art artist Andy Warhol. Warhol transformed Prince's head from the photograph into a highly colourized silkscreen, creating an illustration for the feature titled "Purple Fame." Notably, Goldsmith shared credit for this collaborative endeavour.[6] However, Goldsmith's agreement explicitly stated that the usage was one time. Andy Warhol went on to create additional 15 works based on Goldsmith's photograph, which were incorporated into his estate following his passing in 1987. Meanwhile, Goldsmith was unaware of all of this until 2016, when Condé Nast, the parent company of Vanity Fair, opted to release a commemorative magazine in the aftermath of Prince's untimely demise. In an attempt to republish the 1984 image, Condé Nast reached out to the Andy Warhol Foundation ["**AWF**"]. During this process, they uncovered the existence of 15 other derivative works. Consequently, Condé Nast decided to license one of Warhol's other works, paying a significant license fee of $10,000. Subsequently, the 'Orange Prince' was featured on the cover of the commemorative issue, without any credit to Goldsmith.[7]

Upon discovering the extent of her photograph's transformation, Goldsmith expressed concerns regarding potential copyright infringement by the AWF. In a surprising turn of events, the foundation retaliated by suing Goldsmith. Initially, the District Court ruled in favour of AWF,

---

[5] Rosemary Feitelberg, *Supreme Court Victor Lynn Goldsmith Talks Warhol, Prince and Celebrity Facades,* WWD (Feb. 12, 2023, https://wwd.com/eye/people/lynn-goldsmith-prince-photograph-andy-warhol-supreme-court-interview-1235658276/.

[6] Hannibal Travis, *Warhol Foundation v. Goldsmith: Supreme Court rules for income streams over artistic freedom,* THE CONVERSATION (Feb. 12, 2023), https://theconversation.com/warhol-foundation-v-goldsmith-supreme-court-rules-for-income-streams-over-artistic-freedom-205986.

[7] Bruce E. Boyden, *The Stakes in Andy Warhol Foundation v. Goldsmith,* MARQUETTE UNIVERSITY LAW SCHOOL FACULTY BLOG (Nov. 12, 2023), https://law.marquette.edu/facultyblog/2022/10/the-stakes-in-andy-warhol-foundation-v-goldsmith

citing the principles of 'fair use' as outlined in Section 107 of the U.S. Copyright Act ["**the Act**"][8]. However, this decision was subsequently overturned by the Court of Appeals for the Second Circuit, setting the stage for the case's appeal to the Supreme Court.

The Supreme Court set forth a crucial turning point in the interpretation of 'transformative use' within the context of fair use analysis. At its core, the Court grappled with the question: whether the first factor of fair use analysis, "the purpose and character of the use, including whether the use is for commercial nature and nonprofit educational purposes (as outlined in §107(1) of the Act)", favoured Goldsmith.

The crux of the Warhol Case hinges upon the disagreement between the majority and dissenting opinions concerning the quote extracted from the landmark fair use case, ***Campbell v. Acuff-Rose Music, Inc***[9]. In *Campbell's Case*, the Court delineated 'transformative' use as that which "alter[s] the first work with new expression, meaning, or message." The quote implies that a comparison between the artistic expression, meaning, and message of the original and subsequent works must be made to gauge transformative intent. However, the majority in the *Warhol Case* adopted a more nuanced interpretation. The majority considered it indicative of a broader principle: the assessment of whether the secondary work's purpose differs from the original. This approach positions that the 'new meaning' attributed to a secondary work is relevant, albeit not determinative, in ascertaining its 'transformative purpose.' [10]To sum up the majority's holding: the transformation must pertain to the purpose of the secondary work i.e. the work must have a different purpose, rather than merely its expression, meaning, or message. For Instance, the act of critiquing or creating a parody of a creative piece, as exemplified in the Campbell case, does not supersede the purpose or object of the original work because it shares a different purpose altogether. The court formulated and applied this golden rule to the facts of the case. Goldsmith's and AWF's purposes were identical. Both encompassed portraits of Prince, employed and licensed by magazines to complement narratives about the musician. Consequently, the majority held that Warhol's subsequent use of Goldsmith's copyrighted works did not satisfy the first factor for it to be considered fair use.

---

[8] Copyright Act, 1976, 17 U.S.C. §§ 101-1332 (2012).
[9] Campbell v. Acuff-Rose Music Inc., (1994) 510 U.S. 569.
[10] Samuel Eichner, *Warhol's Ghost in the Machine: What Goldsmith v. Warhol Means for Generative AI*, IPWatchdog.com (June 8, 2023), https://ipwatchdog.com/2023/06/08/warhols-ghost-machine-goldsmith-v-warhol-means-generative-ai/id=162175.

III.   DELVING INTO THE CONTOURS OF FAIR USE ANALYSIS DELINEATED IN THE ANDY WARHOL JUDGEMENT

This section analyses the extensive deliberations of the court as it dissects the concept of fair use while comparing the works and delves into the nuanced interpretation of 'transformative use.'

A. The Broader Interpretation of the Campbell v. Acoff Ross Case

Central to the court's deliberation was primarily the fair use factor, encapsulated by "the purpose and character of the use, including whether such use is commercial or is for nonprofit educational purposes". The majority aptly elucidated this aspect by referencing the *Campbell Case,* emphasizing the distinction between works that merely "supersede the objects" of the original creation and those that introduce novelty with a distinct purpose or character. The court recognized that discerning whether a use mirrors an original's purpose or character, or encompasses new elements, hinges on a matter of degree.

While examining the *Campbell Case* in a more specific context, the court noted that the justification for certain uses on the ground of 'reasonable necessity' is to attain a new purpose. This rationale is evident in instances like parody, where the replication of the original was deemed necessary for the intended message. Likewise, forms of commentary or criticism targeting an original work were acknowledged to require borrowing elements to 'conjure up' the original. This rationale particularly gains prominence when assessing fair use, especially when both the original work and the derivative copy share identical or closely related purposes. This approach also holds relevance when the wide dissemination of a derivative work carries the potential to replace the original or its licensed derivatives.[11]

B. The Golden Rule for Analysing the First Fair Use Factor

The ruling underscores two pivotal notions. Firstly, the distinction between commercial and nonprofit usage emerges as an additional facet within the first, fair use factor. Secondly, this factor relates closely to the rationale behind the use. Broadly, a use possessing a distinct purpose is validated if it aligns with the copyright's goal of advancing arts and sciences while maintaining incentives for creativity.[12]

---

[11] Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985).
[12] Authors Guild v. Google Inc., 804 F. 3d 202, 214 (CA2 2015).

Conversely, a use sharing the same purpose as a copyrighted work amplifies the likelihood of substituting protected aspects of the original or its derivatives, thus undermining the copyright. Drawing from **Google LLC v. Oracle America, Inc. (2021)**[13], the court referenced the analysis of the first factor. Despite the intricacies posed by the functional nature of computer programs, the Court, in evaluating Google's use of Sun Microsystems' code, gauged the divergence of purpose from the original and assessed the strength of justifications for the use. Notably, while commercial, Google's use was grounded in repurposing Sun's code for its Android platform— shaping a distinct and novel computing environment for new products. The necessity for shared interfaces and reimplementation to facilitate programmer skills, lent further justification.

In essence, the court held that the first fair use factor's essence lies in the presence of a divergent purpose or character, a factor of degree. This degree is counterbalanced against the commercial nature of the use. Where an original work and secondary use share akin or highly similar purposes, coupled with the latter's commercial essence, fair use is inclined to be challenged, unless backed by other valid reasons for copying.

C. The Interpretation of Transformative Fair Use and Balancing its Impact on Creativity and Innovation

AWF further argued that their works constitute a commentary on celebrity, particularly Warhol's Prince Series, conveying the dehumanizing facets of fame. This new layer of meaning or message was posited by AWF to render the use 'transformative' in the fair use context. The Supreme Court stated that such an expansive interpretation of 'transformative use' could erode the copyright owner's exclusive right to create derivative works. As articulated by **Judge Leval,** a *"secondary author cannot freely extract substantial portions of the original author's expression solely due to how effectively it conveys the secondary author's different message."*[14]

The dissent extensively underscored a fundamental copyright tenet, the equilibrium between fostering pioneering endeavours and enabling subsequent innovation. The dissent warned that demanding payment from AWF for Goldsmith's creation might hinder creativity and suppress new ideas, casting a pall over innovation. However, the majority dismissed it holding that these concerns are unlikely to stand the test of time. Requiring AWF to compensate Goldsmith for her copyrighted work would not diminish the cultural fabric but reinforce incentives for artists to craft original pieces.

---

[13] Google LLC v. Oracle America Inc., 141 S. Ct. 1183 (2021).
[14] Campbell v. Acuff-Rose Music Inc., (1994) 510 U.S. 569.

IV.   GENERATIVE AI AND FAIR USE IN LIGHT OF THE ANDY WARHOL JUDGMENT

In November, a legal case was initiated by open-source developers against **OpenAI, GitHub Inc., and Microsoft Corp**., highlighting a significant concern regarding Copilot. AI Copilots are essentially conversational interfaces powered by advanced large language models ["**LLMs**"]. Their main goal is to aid users in tasks and decision-making within corporate settings. These copilots harness the impressive capabilities of LLMs to understand, analyse, and process vast datasets.[15] The plaintiffs argue that Copilot replicates their code without giving them proper credit, which wouldviolate their copyright rights. This disagreement emphasizes the pressing need for clear guidelines on whether the incorporation of copyrighted content in AI training aligns with the principle of 'fair use.'[16]Likewise, a collective legal action initiated in January 2023 is directed at **Stability AI, Midjourney, and DeviantArt**. It contends that generating derivative works using AI constitutes a breach of the principles established by the 'fair use doctrine' within copyright law.[17]

AI developers have long championed the argument that such training practices fall within the realm of 'fair use.' They assert that the transformative nature of AI, fundamentally creating something novel from existing data, aligns with interpretations such as the *Campbell and Google v. Oracle case.* The *Warhol ruling's* implication for generative AI lawsuits is clear: AI services may need to enhance their fair use justifications, as AI's inherent transformation might warrant closer examination within the context of established legal interpretations.

A.   The Fair Use Jurisprudence on Intermediate Copying, Reverse Engineering and Non-Expressive Fair Use in Machine Learning and AI

Non-expressive use is premised on the concept that copyright law protects expression, but not control over factual aspects of their work. For instance, when Google Books reveals the location and frequency of a keyword in a book, it imparts factual details about that book. Despite involving wholesale copying without permission, this type of use remains non-expressive.[18] Intermediate

---

[15] Sorab Ghaswalla, *Who or What is an AI Copilot?*, THE MEDIUM (Feb. 13, 2023), https://sorabg.medium.com/who-or-what-is-an-ai-copilot 845175f25ddb.

[16] James Vincent, The lawsuit that could rewrite the rules of AI copyright, The Verge (Nov. 8, 2022), https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data.

[17] Inyoung Cheong, *Generative AI and Fair Use: A Deep Dive into the Warhol Case*, MEDIUM (May 29, 2023), https://medium.com/@inyoungcheong/generative-ai-and-fair-use-a-deep-dive-into-the-warhol-case-2cecbb2bd3eb.

[18] Elise De Geyter, *The Dilemma Of Fair Use And Expressive Machine Learning: An Interview With Ben Sobel*, INTELLECTUAL PROPERTY WATCH (Aug. 23, 2017), https://www.ip-watch.org/2017/08/23/dilemma-fair-use-expressive-machine-learning-interview-ben-sobel/.

copying refers to reproducing copyrighted content as an intermediary step in crafting a fresh copyrighted work, often seen in cases involving computer programs. Notably, video game and software companies might 'reverse engineer' code from copyrighted programs while producing a non-infringing final product. For Instance, Phoenix, a U.S. software company, engaged in intermediate copying to create BIOS software compatible with IBM's version. Instead of directly copying IBM's code, Phoenix reverse-engineered the code, i.e. deconstructed the code to extract design information. This approach allowed them to reference the process without copying the proprietary code directly.[19]

The precedent set-in cases like **Sega Enters. Ltd. v. Accolade, Inc.**[20] recognises the 'Doctrine of Non-Expressive Fair Use' and offers valuable insights into how courts could potentially analyse the training of machine learning AI systems in the future. In this case, the court deliberated over liability for reverse engineering copyrighted computer programs to access their 'unprotected functional elements.' Accolade, Inc. sought to create video games compatible with Sega's 'Genesis' gaming console, opting against licensing Sega's code due to restrictive terms. Instead, they reverse-engineered the code by converting the system's object code into human-readable source code and using the interface specifications to develop their own Genesis-compatible games.

In this, a fair use evaluation led to a ruling in Accolade's favour. Key considerations emerged from the analysis of fair use factors. The first factor, addressing purpose and character of use, favoured Accolade due to its focus on studying non-protectable functional aspects for interoperability. The second factor, concerning the nature of copyrighted work, also supported fair use due to the mix of protectable and non-protectable elements in the code. While the third factor, evaluating the amount of copied content, did not heavily favour fair use due to full code replication, the fourth factor, assessing potential market impact, leaned towards fair use as limiting intermediate copying could impede market competition.

Accolade, Inc.'s ruling established the principle that reproducing copyrighted software to analyse non-protectable, functional elements is fair use if done for a legitimate purpose without alternative means. These cases also underscore that the same legal standard might yield divergent consequences. **Accolade, Inc. Case** permitted Accolade to compete with Sega, while the

---

[19] Ben Lutkevich, *reverse-engineering,* TECHTARGET (June 2021), https://www.techtarget.com/searchsoftwarequality/definition/reverse-engineering.

[20] Sega Enters. Ltd. v. Accolade, Inc., (9th Cir. 1992), 977 F.2d 1510, 1514 (concluding that reverse engineering of object code is a fair use of the copyrighted work if there is legitimate reason for, but no other means of, accessing the code's unprotectible elements).

Connectix Corp. Case[21] allowed a product that could diminish the need for purchasing a Sony PlayStation. The meticulous fair use analysis in the case, although theoretically robust, risks overlooking broader practical implications of excusing unlicensed use.[22]

B. Rethinking the Non-Expressive Fair Use Doctrine Considering the Developments in AI and Machine Learning

The non-expressive fair use doctrine is based on two principles that validate extensive, unauthorized copying conducted by machines. The initial principle asserts that machine-based consumption of copyrighted expression is not inherently infringing. If mechanical ingestion does not enhance human interaction with expressive works, it qualifies as non-expressive. The second premise suggests that these uses do not significantly impact copyright owners' markets, as their rights do not cover non-expressive components engaged in computerized analysis and value derivation.

However, the rise of machine learning challenges these assumptions. Firstly, machine learning enables computers to extract value from expressive aspects of works, not just factual elements. Consequently, these uses might cease to be non-expressive. Secondly, machine learning could introduce a novel market threat: instead of merely replacing individual works, expressive machine learning might entirely substitute human authors, disrupting market impact assessment.

Advanced machine learning reshapes fair use analysis in two key ways: first, intricate machine learning lacks non-expressiveness, and fails to meet the transformative criterion of fair use. Second, expressive machine learning introduces a new form of market substitution, influencing the evaluation of fair use's fourth factor, which is the 'effect of the use on the potential market of the copyrighted work.'[23] Echoing the factors scrutinized in the *Warhol case*, where original work and secondary use share akin purposes, commercial essence, and potential to act as a market substitute, fair use could be contested unless substantiated by valid justifications for copying.

---

[21] Sony Computer Entm't, Inc. v. Connectix Corp., (9th Cir. 2000), 203 F.3d 596, 599 (extending the holding in Accolade, Inc. to computer emulator programming where the software had to be copied for access to its non-copyrightable functional aspects).

[22] Eric Sunray, *Train in Vain: A Theoretical Assessment of Intermediate Copying and Fair Use in Machine AI Music Generator Training*, 13 AM. U. INTELL. PROP. BRIEF 1 (2021).

[23] Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis,* 41 COLUM. J. L. & ARTS 45 (2017).

V. AUTHORS GUILD V. GOOGLE INC. AND THE SHIFT IN TRANSFORMATIVE FAIR USE ANALYSIS

In 2015, the Second Circuit ruled on a case brought by literary authors (*Authors Guild*) against Google, Inc. Google had copied over twenty million books to create an online database for the Google Books project, enabling users to search for and view snippets of the content. The four fair use factors include assessing the purpose and character of the use, nature of the copyrighted work, the amount used, and the potential impact on the market or value of the original work. Courts consider these factors collectively to determine if the use of copyrighted material is fair. The court deemed Google's use transformative, favouring fair use on the first factor, as the functions provided information without substituting for the originals. The second factor also favoured fair use, given the transformative character of the use. The third factor favoured fair use due to limited snippet visibility, despite wholesale copying. The fourth factor aligned with fair use, as snippets were not a meaningful substitute for the original. The court thus held Google's copying and subsequent uses as non-infringing fair use.[24]

Though distinct from traditional reverse engineering, Google Books' massive copying for insightful search tools parallels the concept. Like reverse engineering cases, the *Authors Guild* Case involves intermediate copying, data mining, and developing a new program from the copied data. *Authors Guild* signals a potential shift in intermediate copying analysis, reflecting a trend towards transformative fair use. Prior to the *Authors Guild* decision, wholesale intermediate copying was permitted if it did not mirror the original work's expressive aspects. Now, post-*Authors Guild*, the satisfaction of the first fair use factor, if not the entire analysis, centres on the transformative purpose of intermediate copying.[25]

While the *Authors Guild Case* offers insights into transformative purpose, it diverges from intermediate copying in sectors like the AI music generator training. *Authors Guild's* transformative status rested on Google providing unavailable information to users. On the flip side, the objective of training generative AI systems is distinct from supplying information or commentary to end users. Instead, its primary aim is the generation of novel content using copyrighted compositions and sound recordings. In the absence of a distinct objective, such as an advanced musical search database, relying on the *Authors Guild Case* for transformative purposes in AI development is

---

[24] Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985).

[25] Kartik Chawla, *Authors' Guild v. Google- A Fair Use Victory and a Chance for Introspection,* SPICYIP (Nov. 8, 2015), https://spicyip.com/2015/11/authors-gould-v-google-a-fair-use-victory-and-a-chance-for-introspection.html.

problematic. Moreover, unless AI systems copy for social benefits like insight generation or education, the extensive use of copyrighted works is unjustified.

The legal precedents related to reverse engineering also offer valuable insights into the intent and nature of intermediate copying, particularly within the realm of the AI music generator training industry. In this context, intermediate copying plays a pivotal role in extracting essential elemental information from the works present in the training dataset. Unlike the instances observed in cases like *Accolade, Inc. and Connectix Corp.*, where intermediate copying was imperative for accessing non-protectable functional components of the copyrighted code. For instance, AI music generators employ intermediate copies for an extensive examination of every facet of the musical content. This comprehensive analysis encompasses a spectrum of attributes, spanning from non-copyrightable elements such as song structure and prevalent musical motifs (e.g., chord progressions and melodic intervals) to protected characteristics like timbre, vocal nuances, and distinctive performance qualities found within each sound recording. As a result, the depth and breadth of analysis and manipulation involved in AI music generator training exceed the boundaries of the targeted components in the context of reverse engineering scenarios.[26]

VI.   INSTANCES WHERE THE TRANSFORMATIVE WORKS GENERATED BY AI CAN BE HELD TO FALL UNDER THE FAIR USE RUBRIC

There are certain cases where the utilization or feeding of copyrighted material into AI systems may potentially align with the principles of fair use. In this section, we shall be analysing some of these instances where such cases might be held as transformative fair use, in the light of the principles established in the *Warhol Case*.

A. Use for Sophisticated Public Welfare Purposes

Machine learning often employs copyrighted data that could potentially be categorized as fair use under existing legal doctrine. A pertinent example is facial recognition technology, which necessitates a substantial collection of copyrighted individual photographs for model training. These photographs, however, are typically protected by copyright. The "**Labeled Faces in the Wild (LFW) dataset**", for instance, consists of images featuring 5,749 individuals sourced from news photos on Yahoo News. Replicating this dataset without proper authorization raises concerns about copyright infringement. Nevertheless, arguments can be made that such activities might not constitute infringement. Datasets like 'Faces in the Wild' do not replicate source images

---

[26] Cheong, *supra* note 17.

entirely. They only capture subjects' faces, closely cropped and compressed to low resolution, effectively eliminating most copyrightable elements.[27] Consequently, the risk of improper appropriation of copyrighted content is minimal, potentially averting copyright violations.

It is plausible that utilizing and disseminating the LFW database might not even require invoking a fair use defence. Yet, if such a defence were to be invoked, legal precedents set by the *Authors Guild* case indicate its plausible success. The training of facial recognition models on copyrighted images predominantly deals with factual information, such as the physical attributes of subjects' faces, rather than the artistic choices of photographers. Furthermore, the portions copied receive limited copyright protection and only minor segments are employed. Lastly, the application of these images for facial recognition would unlikely exert a substantial impact on the market for licensed photographs of notable individuals that accompany news stories.[28] Furthermore, this interpretation aligns with the principles established in the *Warhol Judgement*. In this context, the distinct purposes of both the original work and the secondary work remain evident and the secondary work does not function as a replacement for the copyrighted content within the market.

B. Output Based on AI Training and Content Generation

In the realm of AI and copyright, the interplay between training models on copyrighted data and generating content poses intriguing challenges. Training systems on copyrighted data are likely to fall under fair use while generating content using these models could raise copyright concerns. This analogy can be likened to creating fake money for a movie versus using it to make purchases. Training the model on a wide image collection to create novel pictures is less likely to infringe copyrights due to data transformation and limited impact on the original art market. However, fine-tuning the model on an artist's pictures to replicate their style may lead to stronger infringement claims. Vanderbilt law professor Daniel Gervais, explains, *"If you give an AI 10 Stephen King novels and say, produce a Stephen King novel, then you are directly competing with Stephen King. Would that be fair use? Probably not."*[29] Between fair and unfair use, countless scenarios exist, where input, purpose, and output influence the legal outcome.

---

[27] Machine Learning Datasets, PAPERS WITH CODE, https://paperswithcode.com/datasets?task=face-recognition&page=1 (last visited Apr. 9, 2024).
[28] Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985).
[29] James Vincent, *The scary truth about AI copyright is nobody knows what will happen next*, THE VERGE (Nov. 15, 2022), https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data.

C. AI Data Laundering: A Shield for Generative AI Liability

Data Laundering, a phenomenon of transforming pilfered data for legitimate applications, is taking on heightened significance as society's reliance on data continues to grow. Big Tech firms establish or sponsor a nonprofit entity responsible for crafting datasets. The entity's research-driven nature provides an avenue to utilize copyrighted material more flexibly. This dataset can then be harnessed by significant tech players for financial purposes.[30]

An influential factor in determining fair use involves assessing the origins of training data and models, particularly if they have been produced by academic researchers and nonprofit organizations. Stability AI, for example, instead of directly crafting the training data or training the models used in their software, facilitates the involvement of academic researchers. These researchers, working in collaboration with a German university, developed the Stable Diffusion model. This strategy allows Stability AI to market the model as DreamStudio, all while maintaining a legal distance from its creation.[31]

VII. CONCLUDING REMARKS: FAIR USE, PRODUCTIVITY, AND TRANSFORMATION IN AI-GENERATED CONTENT

In conclusion, the distinction between productive and consumptive uses, a pivotal aspect of transformative fair use, is deeply rooted in copyright jurisprudence. This distinction has found resonance in cases like ***Sony Corporation of America v. Universal City Studios, Inc***.[32], where the Supreme Court deliberated over the fair use implications of taping copyrighted television broadcasts. **Justice Blackmun's dissent** differentiated productive uses from ordinary, consumptive ones, emphasizing that productive uses result in added benefits to the public[33]. This concept of productivity has permeated the landscape of 'transformative use,' underscored by its centrality in determining the viability of a fair use defence.

In the evolving landscape of AI-generated content, this distinction becomes even more pronounced. The discussion around whether expressive machine learning constitutes a productive or consumptive endeavour holds considerable weight. Notably, AI's capacity to generate gigabytes of copies of copyrighted art for emulation and learning marks an impressive technological advancement. Yet, the notion of productivity has its limits. While machine learning may hold the

---

[30] Devansh, *Machine Learning Made Simple, Is Big Tech Using Data Laundering to Cheat Artists*?, Medium (Nov. 15, 2022), https://medium.com/discourse/is-big-tech-using-data-laundering-to-cheat-artists-ccf1a8c87b91.
[31] Authors Guild v. Google Inc., 804 F. 3d 202, 214 (CA2 2015).
[32] Sony Corporation of America v. Universal City Studios, 464 U.S. 417, 419-20, (1984).
[33] *Id.* at 478-79, (Blackmun, J., dissenting)

potential for future productivity, it cannot serve as a carte blanche for bypassing copyright protections.

In this context, *Warhol* Case's reinterpretation serves as a guiding precedent. The *Warhol* Judgment underscores the importance of balancing new creations against the potential market effects on the original work. Just as the court's analysis of *Warhol* emphasizes the delicate equilibrium between transformative intent and market impact, the question of fair use defence of AI-generated content pivots on whether the generated content plays a productive or consumptive role within the broader creative landscape. As this evolving terrain necessitates discernment, acknowledging the fine line between productive machine learning and mere consumptive use remains crucial.