GENERATIVE AI'S TWO INFORMATION GOODS

OREN BRACHA[*]

A large number of diverse copyright infringement claims are currently being brought against different actors involved in the production cycle of Generative Artificial Intelligence (GenAI). This Article argues that the key to analyzing such claims is to distinguish between two kinds of information goods implicated by such GenAI systems. A first kind of information good consists of discrete expressive works that form the training set and generated output of such systems. A second kind is aggregate "metainformation" about the discrete works in the dataset that form the model. These information goods have sharply different characteristics and raise distinct policy concerns. Copyright is an institutional tool designed to address a specific policy problem related to the production and use of discrete expressive works and is ill-suited for handling the distinct policy concerns posed by aggregate metainformation. Yet many GenAI copyright infringement arguments are attempts to use copyright for addressing claims related to the aggregate metainformation good, often dressed up as claims about the use of discrete expressive works. Such arguments should be rejected as lying outside copyright's domain. Other infringement arguments, involving generated output substantially similar to works in the training set, are proper claims about the use of discrete expressive works. These fall within the domain of copyright, but scope of liability principles should often prevent the extension of direct liability to GenAI producers. In such cases, upstream producers' liability for downstream infringement will be determined by secondary liability doctrines. This legal structure creates a lacuna: cases where producers do not fit one of the secondary liability categories but could have taken cost-effective precautions to reduce infringement risk. This lacuna should be filled by negligence-based liability that requires producers to take reasonable precautions to reduce the risk created by their technology.

Contents

I. INTRODUCTION

The Generative Artificial Intelligence (GenAI) revolution has descended on the field of cultural production.[1] In response copyright owners are suing. In a flood of copyright lawsuits plaintiffs target various actors involved in the production of GenAI systems with every infringement argument in their arsenal.[2] Copyright litigation, at least for now, appears to be our

---

[1] *See e.g.,* Kevin Roose, A.I.-Generated Art Is Already Transforming Creative Work, NY Times, October 21, 2022. https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html.

[2] At the time this draft was written there were 25 active lawsuits on the subject. Some of the more significant cases are: Andersen v. Stability AI LTD, Case 3:23-cv-00201 N.D.Cal. (hereinafter "Andersen v. Stability AI"); Getty Images (US), Inc. v. Stability AI, Inc. Case 1:99-mc-09999, D.Del. (hereinafter "Getty Images

central institutional tool for dealing with the disruptive effect of this technology. Does GenAI trained on copyrighted materials infringe copyright? Phrased on this level of generality, the question is meaningless. Production of expressive materials by GenAI involves a complex supply chain composed of different information processing phases and numerous actors engaged in diverse activities.[3] Copyright law, for its part, is an intricate set of rules and principles that create various rights and apply differently to different activities.

Where does one begin to disentangle this tangled web? The central thesis of the Article is twofold. Its first prong is that the way to make sense of the complex landscape of GenAI copyright infringement is to clearly identify and then keep firmly in sight the two distinct and different information goods relevant for this technology.[4] The second is that once such distinction is kept in sharp focus, the deep obscurity that mars the existing GenAI copyright debate gives way to conceptual clarity. Specifically, this conceptual clarity is crucial for guiding three dimensions of the analysis: identifying the very different policy issues at stake pertaining to the social effects of the production and use of different information goods; deciding which of these issues can be adequately addressed via the institutional tools of copyright; and with respect to the latter, properly developing the copyright policy analysis as it pertains to the relevant information good.

---

v. Stability AI"); Does v. GitHub, Inc., Case 3:22-cv-06823, N.D.Cal.; The New York Times Co. v. Microsoft Corp., case 1:23-cv-11195, S.D.N.Y. (hereinafter "New York Times v. Microsoft"). For media discussion *see e.g.*, Blake Brittain, Lawsuits accuse AI content creators of misusing copyrighted work, Reuters, January 17, 2023. https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17; Molly Enking, Is Popular A.I. Photo App Lensa Stealing From Artists?, Smithsonian Magazine, December 14, 2022. https://www.smithsonianmag.com/smart-news/is-popular-photo-app-lensas-ai-stealing-from-artists-180981281; Blake Brittain, Getty Images lawsuit says Stability AI misused photos to train AI, Feb. 6 2023. https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06; Christopher Mims, AI Tech Enables Industrial-Scale Intellectual-Property Theft, Say Critics, The Wall street Journal Feb. 4, 2023; Christopher Mims, Chatbots Are Digesting the Internet. The Internet Wants to Get Paid, The Wall Street Journal, April 29 ,2023; Jonathan Stempel, NY Times sues OpenAI, Microsoft for infringing copyrighted works, Reuters, December 27 2023. https://www.reuters.com/legal/transactional/ny-times-sues-openai-microsoft-infringing-copyrighted-work-2023-12-27/.

[3] See generally Katherine Lee, A. Feder Cooper, James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, Journal of the Copyright Society (forthcoming). Available at: https://ssrn.com/abstract=4523551.

[4] Credit for the method of analyzing public policy and specifically innovation policy problems by clearly identifying the relevant information goods involved goes to Talha Syed. *See* Talha Syed & Anna di Robilant, *Property's Building Blocks: Hohfeld in Europe and Beyond*, in SHYAMKRISHNA BALGANESH; TED SICHELMAN, HENRY E. SMITH EDS., THE LEGACY OF WESLEY HOHFELD: EDITED MAJOR WORKS, SELECT PERSONAL PAPERS, AND ORIGINAL COMMENTARIES (2022) (emphasizing the need for careful *resource-specific* analysis in the case of property theory and policy); and Talha Syed, *Pharma's Paten paradox,* YALE L. J. (forthcoming) (anchoring an analysis of pharmaceutical innovation policy on the foundational point that "innovation in pharmaceuticals consists of not one, but *two separate information goods*") (emphasis in original, draft manuscript on file with the author). I also build on the work of Salome Viljoen who pursues a similar method in her work on the analogous context of privacy and information surveillance. *See* Salome Viljoen, *A Relational Theory of Data Governance* 131 YALE L.J. 573 (2021).

There are always two kinds of information goods involved with GenAI systems. The first is discrete informational items that constitute the training set and the generated output. With respect to expression-producing GenAI these are expressive works. The second is aggregate metainformation, namely, information about the discrete works in the training set. This metainformation constitutes the model—the engine that powers the system. Consider the following analogy. Specific pairs of planal coordinates (X,Y) and (X', Y') are discrete information goods. Each embodies an information item about the location of a point in a plane. By contrast, $y=ax+b$ is aggregate metainformation. The formula represents information *about* a general relation held internally by multiple discrete information items [(X1, Y1); (X2, Y2)… (Xn, Yn)]. The metainformation formula "contains" the concrete information from which it was constructed—say four known pairs of coordinates on a line. If one knows how to use the formula, the preexisting discrete information items can be "extracted" from "it." But the formula can do so much more, because it allows one to generalize—to generate new discrete information items about countless previously unknown coordinates on the line. On a vastly more complex level, items in GenAIs datasets and output are like specific pairs of coordinates. The model is like the formula.

Once stated, the distinction may seem obvious, perhaps even trivial. But the conceptually clarifying effect on the GenAI copyright infringement debate is tremendous. And while conceptual clarity cannot, by itself, settle substantive questions of social policy, it goes a long way in setting the analysis on the right path and preventing it from falling into pitfalls of confusion or obscurity. The key insight is that the two kinds of information goods involved with GenAI are *conceptually* completely distinct from each other, have different traits, different use and market values, different dynamics of production and use, and different social effects and policy issues associated with them. Moreover, different institutional actors interact differently with the two information goods. The actor who assembles and reproduces the dataset is often not the one who produces the metainformation in the model, and both are frequently distinct from the actor who sets in motion the generation process.

Claims and policy concerns with respect to GenAI fall into three spheres, depending on their relation to the two distinct information goods. First, there are metainformation claims. These claims pertain to the social effects and dynamics set in motion by aggregate metainformation. Sometimes these claims are made directly, but often, espcially in the context of copyright litigation, they are dressed up as claims about the misuse of discrete works. Nevertheless, for all the ingenuity of these strategies, the fact remains that in these cases, what hides under thin rhetorical garments are metainformation claims. This Article argues that metainformation claims do and should fall outside the subject matter domain of copyright because the field was not designed and is institutionally unequipped to deal with them.

Second, there are discrete expressive works claims proper. These claims are about individual harms inflicted through the production and use dynamic of specific expressive works. In GenAI cases the hallmark of such cases is the existence of similar generated output. These claims fall squarely within the domain of copyright, but they raise in earnest the question of who among the many GenAI actors is liable for infringing acts. This Article argues that upstream GenAI producers are involved in the production of concrete infringing output on the level of metainformation and therefore they hold the position of gatekeepers of technology that has both

beneficial and harmful uses. As a result, they should not be deemed directly liable for specific acts of infringement by use of their systems.

Third, there are claims that are about the interface between metainformation and concrete expressive works. While actions in the domain of metainformation and those in the domain of discrete expressive works are distinct, they are not hermetically insulated from each other. Claims in this category assert that actors in the domain of metainformation bear some responsibility for the implications of their actions in enabling individually harmful acts in the domain of discrete works. This Article argues that upstream GenAI producers should have a duty to take reasonable precautions to minimize the social risk of infringement imposed by their systems. In legal terms: producers should be subjected to a negligence standard. GenAI, just as much as railroads or cars, is a beneficial technology that comes with social risk. What we should expect of its gatekeepers is to take reasonable precautions to reduce this risk. Yet it is exactly this liability standard that existing copyright law, with its all or nothing strict liability, is missing.

The Article proceeds as follows. Part II identifies the two information goods at work in the GenAI context, elaborates their distinct features, and explains the crucial importance of the distinction for legal and policy questions in the field of expressive GenAI. Part III discusses metainformation claims. It explains why many of the broad infringement arguments made in the GenAI copyright cases are metainformation claims dressed up as discrete expressive works claims, and why copyright is an inadequate institutional tool for dealing with claims of the former kind even when they raise genuine and important policy concerns. Part IV focuses on discrete expressive goods claims that belong in the domain of copyright. The Part identifies the generation of expressive output sufficiently similar to a copyrighted work as the litmus test for cases that fall within copyright's domain. It then argues that upstream producers are gatekeepers who should not bear direct liability. Part V turns to the appropriate liability standard by upstream producers. The Part argues that the answer to the question of whether, in technological terms, the metainformation embedded in GenAI models "contain" the informational content of works in the data set cannot, by itself, answer the legal liability question. It then explains the limitations of the main existing tool in copyright's arsenal for deciding gatekeepers' liability, namely, secondary liability doctrines. Part VI argues that secondary liability should be supplemented by a negligence duty imposed on GenAI gatekeepers to take reasonable precautions for reducing the risk of infringement. The Part discusses the theoretical justification for a negligence standard and suggests alternative doctrinal ways for its implementation. Part VII concludes.

II. TWO INFORMATION GOODS

A. Expressive Goods and Metainformation

It is Generally acknowledged that for purposes of legal and policy analysis there is hardly such a thing as any one unified GenAI system.[5] Rather, GenAI is a web, or perhaps a supply chain, composed of distinct information processing activities, unfolding in separate stages, executed by

---

[5] Lee et al *supra* note 3, at 4 (Observing that "'generative AI' is a catch-all name for a massive ecosystem of loosely related technologies").

different institutional actors, and organized in many different architectures[6]. By contrast, it is not generally acknowledged, but equally important, that the GenAI web of activities involves two distinct information goods. Abstracting away from many important details, the heart of a GenAI system is two processes of transforming information. The first one is extracting metainformation from concrete data. This is the training process in which information about patterns and relations within and among many individual informational items in the dataset is extracted.[7] This information about information (metainformation) is represented in the model as a complex set of parameters and functions, also known as "weights."[8] The second process is that of prediction or generation, where the metainformation in the model is used to generate new concrete informational items, typically in response to a specific trigger or prompt.[9] Thus, GenAI systems of all architectures and kinds involve two informational goods: concrete informational items that compose the training set as well as the generated output, and metainformation constituting the model.

Crucially, the two information goods have very different traits. The specific information goods (or "works") in the training set and output are discrete, and in the case of expressive GenAI—which is the focus here—they are expressive. The works are discrete in that each stands on its own two feet as having use value to people. Put differently, these works can be consumed individually for the specific expression embodied in them. Saying that the works are discrete should not be confused with an atomistic approach to expressive works. Unless one takes an extreme romantic authorship view, it is clear that discrete works are part of a broader and ultimately social network.[10] Neither creation nor consumption of works happens ex nihilo. Production of works needs other works as points of reference, building-blocks and sources of inspiration.[11] Similarly, the value of works in consumption comes from a meaning-making process that necessarily unfolds within a set of relations to other works.[12] Yet for all that, the works remain discrete in being individually consumable for the specific expression contained in them.

In the case of GenAI whose main purpose is generating material such as texts, images, or videos, the discrete works both in the training set and in the output are expressive. This means that

---

[6] *Id.*, at 5-6 (Describing the different phases of the "**generative-AI supply chain**").

[7] ETIENNE BERNARD, INTRODUCTION TO MACHINE LEARNING 11 (2021).

[8] *Id.*, at 272.

[9] *Id.*, at 11.

[10] See Michal Shur-Ofry, *Popularity as a Factor in Copyright Law*, 59 U. TORONTO . J. 525, 528-534 (2009) (using network theory to analyze the connection between expressive works).

[11] Jessica Litman, *The Public Domain*, 39 EMORY L.J. 965 (1990) (observing that "the very act of authorship in *any medium* is more akin to translation and recombination than it is to creating Aphrodite from the foam of the sea").

[12] Omri Rachum-Twaig, *A Genre Theory of Copyright*, 33 SANTA CLARA HIGH TECH. L.J. 34, 38 (2016) (observing that "for the audience" common expressive building blocks "function as raw materials for the creation of meaning and as a common language between the audience and authors").

these works have substantial use value related not to their informational content, but rather to the aesthetic forms they contain or in which they are organized.[13]

By contrast the metainformation that constitutes the model is aggregate social information and is primarily non-expressive. The metainformation is aggregate because it is information not about specific traits of specific works, but rather collectively on patterns and relations across a large group of works. It is social in the sense that these collective patterns and relations of works are ultimately traceable to actions and preferences, not of specific persons—whether creators or users—but to collective patterns of human behavior across society. The metainformation in GenAI models is sometimes metaphorically described as a vast pile of Lego building blocks.[14] The metaphor is useful insofar as it captures the fact that a model is not a database containing the works in the dataset, but rather is the result of extracting features of these works to enable assembling a large variety of new and different works in the generation stage. It falls short, however, in not clearly capturing the fact that the features extracted from the works in the dataset are not individual sub-elements of any particular work, but rather aggregate elements common to the set. Indeed, no individual specific work in the dataset or its elements are of much consequence to the model.[15] What matters is the aggregate or social aspect of the metainformation.

It is this collective or relational character of the information that gives the model its generative power, which accounts for most of its social value, quite distinct from and potentially far beyond the value of simply a database from which one can retrieve the individual works in the dataset. This power to "generalize"—to abstract the aggregate patterns and relations beyond the specific cases already contained in the dataset, and then use them to generate new concrete instances of works—stems from the collective character of the information.[16] This generalization power, rooted in the aggregate and relational character of the data, is also the engine that powers the vast value of the technology.[17] No one would invest the extensive resources required for developing, say, the Llama large language model, simply to create a textual database of existing materials. The name of the game is generalization that powers the production of new goods, and

---

[13] The relative importance of the expressive use-value of works may change with context. Many queries on Chat GPT are mainly motivated by the information content of the response provided rather than their aesthetic forms. Prompting Chat GPT for a poem or an image generator for a particular image is likely to be much more focused on the expressive value of the output. Regardless, as long, the relevant works have more than negligible expressive use value, they are expressive.

[14] See e.g. SUDIP DASGUPTA, THE MAGIC BEHIND GENERATIVE AI: A JOURNEY INTO GENERATIVE AI WONDERLAND 9 (2024).

[15] See Viljoen *supra* note 4, at 611 (observing in the analogous personal information context that "In a typical data flow, any one individual's data is essentially meaningless, and the marginal cost of any one individual defecting from collection is very low).

[16] See Bernard *supra* note 7, at 96 (explaining that "models must *generalize* to unseen data. Generalization is a fundamental property that every machine learning model should have").

[17] Viljoen *supra* note 4, at 610 (discussing the "horizontal" or relational character of machine learning data and observing that "Data's relationality is central to how data collection produces economic value").

the core element for achieving this purpose is extracting aggregate and relational social patterns from existing expressive works.

The metainformation in the model is non-expressive in the sense that it has no, or at most trivial, use value based on the direct consumption of the aesthetic forms, if any, embodied in it. Perhaps there are a few out there who would sit down in their TV Chair with a good cup of tea and consume chunks of the information in the GPT model for the sheer aesthetic pleasure of it. But it is beyond dispute that any direct aesthetic use value of the metainformation, is infinitesimally small in value relative to its non-expressive uses. The driver of the tremendous value of the metainformation is its informational-functional value, meaning the ways that its content can be leveraged into powering GenAI systems and unlocking the social value they can offer. To be sure, the metainformation itself is *about* expression—to enable the generation of expressive works the aggregate relations and patterns it represents must be patterns and relations of expressive forms.[18] And, no doubt, the metainformation's productive power is cashed out by being translated back into discrete works with expressive use value. In this way, the metainformation is a factor, or if you wish means, of production—a critical resource used to produce specific goods consumed for their expression.[19] But none of this changes the fact that the use value of the metainformation is in powering the apparatus that produces new expressive golden eggs, not in being such an egg itself.

Before proceeding a clarification is required to forestall some possible objections. The distinction between discrete expressive works and aggregate, functional metainformation is conceptual.[20] It is not ontological, because it is agnostic with respect to any claim about "what is really out there." The point of the distinction is to offer proper conceptual tools for analyzing the relevant subject matter in light of a purpose. The purpose is to understand the production/use dynamics and effects of the relevant information goods and devise institutional tools for regulating them.[21] The distinction also makes no claim about any physical object in which the information goods are embedded. Its point is that conceptually it is useful to distinguish between the two information goods, not denying the possibility that often both information goods may be embodied in a single physical object. Finally, the distinction makes no claim that actions in the domain of metainformation are always hermetically insulated from actions in the domain of discrete expressive works. Quite to the contrary, the point of the analysis is to properly understand the relation between the two domains. The claim is that a precondition to properly understanding the

---

[18] See Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines* (unpublished). GWU Legal Studies Research Paper No. 2024-52, GWU Law School Public Law Research Paper No. 2024-52, Available: https://ssrn.com/abstract=4909592.

[19] Viljoen *supra* note 4, at 607 (observing in the personal information context that the relational character of the system "is expressed technically through informational infrastructure that makes sense of data subjects via group classification and that operationalize classifications to act back on subjects.")

[20] See Talha Syed, *Legal Realism and CLS from an LPE Perspective* (unpublished), at 29, 44-45. Available at: https://ssrn.com/abstract=4601701 or http://dx.doi.org/10.2139/ssrn.4601701.

[21] *Id.*

interaction between the two kinds of information is to have a crisp conception of each and the distinction between them.

B. The Upshot

Why should we care? Because each of the information goods is subject to different dynamics of production and use, in which various institutional actors play different roles, giving rise to different social effects, implicating different human interests, and requiring very different legal-institutional responses. Consequently, conflating the two information goods inevitably leads to hopeless confusion in understanding the social processes at work, the normative stakes, and the adequate institutional responses to them.

The key point is that discrete expressive works and metainformation about these works, being two distinct resources, have separate production/use dynamics. And these separate dynamics give rise to distinct social effects, beneficial or harmful. To see this, consider an easy case. There is a rising concern of "AI Monopolies."[22] The concern is that the market for so called foundation models is likely to be dominated by a handful of firms, which in turn, through a system of vertical integrations and alliances might lead to very few Big Tech companies completely dominating GenAI markets.[23] This prospect raises dangers from the realm of competition policy, related to a host of anti-competitive practices and their deleterious economic effects.[24] But the concerns run deeper; they also pertain to the possibility that, as seen in previous episodes related to digital platforms, concentrated economic power could lead to other worrisome effects, ranging from economic inequity to political influence.[25] While this concern is hardly limited to expressive GenAI, it is not hard to imagine a particularly troubling version of it when the product of GenAI is cultural materials. In a nutshell, the fear is of concentrated control of "culture," and the manyfold of social evils that could follow.[26]

Laying aside the probability of such dark predictions being realized, for current purposes, the relevant question is: what are the economic drivers of the predicted trend of centralized control by a handful of giants? The answer is that the driving forces are rooted in the production/use dynamics of aggregate metainformation. While in the cultural sphere the effects of the economic processes may play out by shaping the character, prices, accessibility, and compensation for

---

[22] See e.g. Tejas N. Narechania, and Ganesh Sitaraman, *An Antimonopoly Approach to Governing Artificial Intelligence*, Yale L. & Pol. Rev. (forthcoming). available at: https://ssrn.com/abstract=4597080 or http://dx.doi.org/10.2139/ssrn.4597080; Jai Vipra and Anton Korinek, *Market concentration implications of foundation models: The Invisible Hand of ChatGPT*, available at https://arxiv.org/abs/2311.01550; Johan Moreno, FTC Concerned About Generative AI Monopolies, Forbes, June 30, 2023.

[23] Ganesh Sitaraman and Tejas N. Narechania, It's Time for the Government to Regulate AI. Here's How, Politico, 15 January 2024. Available at https://www.politico.com/news/magazine/2024/01/15/sitaraman-artificial-intelligence-regulation-00134873.

[24] Narechania, & Sitaraman *supra* note 22, at 23-29.

[25] *Id.*, at .29-33

[26] See e.g. Yuval Harari, Tristan Harris and Aza Raskin, You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills, NY Times, March 24, 2023.

creators of discrete expressive works, sitting firmly in the driver seat is the dynamics of producing and using aggregate metainformation. The horizontal relational and aggregate level of metainformation is what sets in motion the vertical power of GenAI giants with respect to specific relationships.[27] In a nutshell, the steep cost of producing a foundation model creates high barriers to entry.[28] The sources of the high cost are the need to obtain vast amounts of data, the extensive computational resources required to train models (i.e. to extract the metadata), and perhaps the scarcity of highly trained human talents to manage this process.[29] Moreover, much as in the context of digital search, the use side feeds backs into production: the more broadly a model is used the more information is generated that allows refining training and improving performance.[30] The result is positive network externalities that further raise the entry barriers: the few who could marshal the extensive resources to enter enjoy the advantage of the information generated by the large crowd of users of their network.[31] The material point is that the pattern of dominance by a handful of giants, although it may have significant effects on markets for discrete expressive goods, is driven by the economics of metainformation. It would be counterproductive, indeed in this case clearly strange, to address problems stemming from these dynamics of producing and using metainformation with institutional tools—copyright, for example—that were designed to address policy problems stemming from features of the production and use process of discrete expressive works.

A crisp conceptual distinction between the two information goods is needed as an inoculation against confusion in explaining the character and source of policy problems that is sure to result in mis-prescribing the remedies for them. The possible confusion is patently apparent in the example above, but it is writ large, albeit often in more subtle forms, over the entire policy and legal debate surrounding expressive GenAI. The phenomenon is a close cousin of the one identified by Salomé Viljoen with respect to a different branch of data capitalism, i.e.: personal information.[32] As Viljoen explains, existing debates tend to focus on harms associated with specific items of personally identifiable information and therefore prescribe remedies from the realm of "privacy rights" designed to protect individual interests.[33] However, the real driver of the personal information datafication trend and its social effects is a very different information good relevant

---

[27] See Talha Syed, *The Vertical and the Horizontal in Capitalism* (unpublished draft manuscript). Abstract available at: https://lpeproject.org/wp-content/uploads/2021/02/Syed_The-Vertical-and-Horizontal-in-Capitalism.pdf.

[28] Narechania, & Sitaraman *supra* note 22, at 18.

[29] Vipra & Korinek *supra* note 22, at 9-13; Tejas N. Narechania, *Machine Learning as a Natural Monopoly,* 107 Iowa L. Rev. 1543, 1578 (2022).

[30] Narechania *supra* note 29, at 1584 (observing that "Machine-learning-based applications that continue to internalize new data, including information drawn from their practical deployments, may gain an insurmountable lead over putative competitors in their initial competition for the market").

[31] *Id.*

[32] See Viljoen *supra* note 4.

[33] *Id.,* at 613-15.

for fueling AI technology: aggregate, social data, whose appropriate regulation requires different, more socialized institutional forms or governance.[34]

The same observations apply, mutatis mutandis, to expressive works. The driver of the power of GenAI in this area, including both its significant promise and looming threats, is aggregate metainformation. As in the personal information context, conflating metainformation with discrete expressive works is sure to result in a flawed diagnosis of the social dynamics leading to relevant policy problems. This misdiagnosis is sure to lead to mis-prescription with two matching aspects. One aspect is a mismatch between the problems and the institutional remedies. Usually, this mismatch would be in the form of turning to inadequate responses taken from the toolkit of institutional solutions to problems pertaining to discrete expressive works, when what is needed are different tools designed to address aggregate metainformation problems and therefore are more social in character. The second aspect is a blinkered institutional imagination—the result of the misconception of the problem restricting our ability to conceive of the full range of adequate solutions, whether existing or new.

In the area of expressive works, a clear understanding of the different policy problems associated with discrete works and aggregate metainformation and of the adequate institutional responses to them requires delineating how the conceptual distinction plays out in three spheres: claims or concerns that are about metainformation and its dynamics; claims or concerns that are about discrete expressive works; and claims or concerns that are about the interface between the two or about the way in which actions in one spheres relate to the other. The remainder of this Article takes up in turn each of these spheres.

III. METAINFORMATION CLAIMS

GenAI metainformation claims are policy or protected interests claims that are directly traceable to the production/use dynamics of the metainformation powering the technology. Such claims come in two formats: direct and sub-rosa. This Part describes the two types of metainformation claims and then explains why they should not be addressed via copyright.

A. Direct: Cultural Policy

Direct metainformation claims in the realm of cultural expression openly frame the policy concerns on which they are based as rooted in the dynamics of producing and using metainformation and the social effects these dynamics breed. Claims that are framed this way operate outright on a collective or social level. While the ultimate normative concerns are for individual human interests, the identified effects on these interests operate through aggregate metainformation, and therefore they unfold on a collective level and require addressal on that level. Thus, direct metainformation claims are openly claims about social cultural policy.[35]

---

[34] *Id.*, at 578-9.

[35] I bracket here particular set of concerns about the effect of GenAI within the cultural sphere that operate both on the individual and social level: those related to generating false, fictious, or biased information. See generally, Noam Kolt, *Algorithmic Black Swans*,101 WASH. U. L. REV.1177 (2014). I lay these concerns

One example of a groups of claims in this mold was already discussed above: concerns about concentration in GenAI markets and the various social ills that might follow in its wake.[36] Another host of concerns stems from the prediction that GenAI might significantly decrease the market share of human creators in certain markets for expressive works, or even displace such creators altogether.[37] Perhaps the most straightforward threat in this vein is the loss of livelihoods in creative industries.[38] Another relates to the loss, not of livelihoods, but opportunities for meaningful creative activity and its inherent value.[39] Yet a third branch of these concerns is about the possible negative effects of a dwindling human sphere of creativity on the sources and potential for paradigm-breaking innovation.[40] A fourth branch of concerns stemming from GenAI market dominance relates to corrosive effects on the plurality and variety of creative output.[41]

Three features of the concerns in the above, non-exhaustive, list should be highlighted. First, with respect to each of the concerns, the driving force is the productivity and cost-efficiency unleashed by the metainformation that powers GenAI. The ability to produce high quality expressive materials at low cost promises significant social benefits, including both satisfaction of demand and boosting of productivity.[42] At the same time, because the productive power unleashed often replaces that of humans, the result is their potential displacement from markets and the

---

aside despite their importance because their focus is on the content of generated information while the discussion here pertains to policies relevant for the creative and expressive side of cultural production.

[36] See *supra* text accompanying notes 22- 26.

[37] See Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis* 41 COLUM J. L. & ARTS 45, 77-79 (2017) Brauneis *supra* note 18; BJ Ard, *Copyright's Latent Space: Generative AI and the Limits of Fair Use,* 110 CORNELL L. REV. (forthcoming 2025), at 59-61.

[38] See e.g. Pamela Samuelson, *Generative AI meets Copyright*, 381 SCIENCE 158, 159 (July 2023) (observing that "Generative AI seems poised to have substantial impacts on the careers of professional writers and artists"); US experts warn AI likely to kill off jobs – and widen wealth inequality, The Guardian, Feb. 8, 2023. https://www.theguardian.com/technology/2023/feb/08/ai-chatgpt-jobs-economy-inequality.

[39] Oren Bracha, *The Work of Copyright in the Age of Machine Production*, HARV. J. L. & TECH. (forthcoming 2025), at 39.

[40] *Id.,* at 40-41.

[41] See Michal Shur-Ofry, *Multiplicity as an AI Governance Principle* (May 10, 2023). Available at: https://ssrn.com/abstract=4444354 (discussing the danger that large language models "could shift social perceptions toward uniformity and standardization, at the could shift social perceptions toward uniformity and standardization, at the expense of diversity and multiplicity."); Giancarlo Frosio, *Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity?*, in ENRICO BONADIO AND CATERINA SGANGA EDS., A RESEARCH AGENDA FOR EU COPYRIGHT LAW (forthcoming), Available at: https://ssrn.com/abstract=4527461, at 14-15 (discussing concerns of "Cultural Homogenisation" induced by GenAI); Octavio Kulesz, *Culture, platforms and machines: the impact of artificial intelligence on the diversity of cultural expressions* (UNESCO Intergovernmental committee for the protection and promotion of the diversity of cultural expressions, Paris 2018), at 12-13 (discussing GenAI effects on cultural diversity); Anil R. Doshi, & Oliver P. Hauser, Generative Artificial Intelligence Enhances Creativity but Reduces the Diversity of Novel Content (2023), https://arxiv.org/abs/2312.00506.

[42] See *infra* text accompanying notes 157-159.

attendant dangers claimed. Second, the economic interaction that generates these benefits and dangers is social and so are its effects. The metainformation that drives the process is social, being an aggregation of collective patterns of many existing works, and so is its cost-reducing effect.[43] Consequently, the resultant harms are diffused or systemic.[44] They are not visited directly through concrete harms to the interests of a particular individual, whether personal interests or in specific expressive works, but rather across society, with respect to the entire sphere involved.[45] Third, the various harms claimed are extra-market harms. The cost efficiencies of GenAI that drive the negative effects are a benefit in market efficiency terms. Moreover, the downsides identified can be easily incorporated into a market framework: what is lost in terms of jobs, creative opportunities, innovation, or variety, supposedly will only be lost if its market value is outweighed by the cost savings. Thus, to have any normative force each of the claims must depart from a strict market framework, either by pointing at some market "failure" disrupting the social calculus of costs and benefits and its realization through market prices, or more radically, rejecting the normative appeal of the maximization of market value criterion altogether.

B. Sub-Rosa: Cultural Policy Dressed Up as Copyright

While direct cultural policy claims certainly circulate in the GenAI context, there is also a widespread different format of metainformation claims. These are sub-rosa claims, where the social cultural policy concerns are dressed up as claims about individual interest pertaining to discrete expressive works. This strategy is common when the claims arise in copyright disputes. Or to put it somewhat differently: the strategy converts social cultural policy claims into ones of individual copyright harms.

Examples of cultural policy claims dressed up as ones of individual copyright harms are as varied as the ingenuity of lawyers devising them, but it would be useful to examine a few of them. Probably the most widespread and impactful claim of this kind is the argument that reproduction of copyrighted works strictly for purposes of the training process constitutes infringement, irrespective of the output of the system.[46] This argument takes central stage in several of the more

---

[43] See *supra* text accompanying notes 14-17.

[44] See Shur-Ofry *supra* note 41, at 40 (explaining that systemic harms "do not translate to immediate decisions affecting individuals" but rather consist "in the cumulative effect of LLMs' outputs across time that could yield the undesirable societal consequences"); Kolt *supra* note 35, at 1219-22 (2014) (discussing "systemic risk" associated by AI systems).

[45] Of course, at the end social harms do affect individuals including with respect to their interests in specific works. The point is that a particular action on the metainformation level, even with respect to a specific work, does not translate into a discrete harm to the individual or work involved, but accretes together with multiple other actions to affect large groups of individuals and works. See Viljoen *supra* note 4, at 608 (discussing the concept of "population level interests" with respect to social data that do not "reduce to the individual provenance of the data").

[46] See Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743 (2021) (explaining the argument that AI training copies infringe copyright).

important litigated cases and surrounding scholarly debate.[47] Its gist is as follows. Training an AI system—the process of extracting metainformation—requires creating digital copies of the works in the dataset. The creation of a digital copy falls squarely with the Copyright Act's definition of reproduction.[48] Hence, when the reproduced works are under copyright—as very often is the case—this reproduction is infringing. Note that that the argument applies to the training reproduction, irrespective of later stages, no matter whether any output of the AI system is similar to works in the dataset, or whether the system generates any output at all.

On the surface, the argument seems to be a bona fide copyright claim pertaining to harm done to discrete works. After all a copy (of a specific work) is a copy. However, slight pressure reveals that the argument has little to do with discrete expressive works and is, in fact, a metainformation claim. The argument aims directly at the metainformation and the entire GenAI system it powers. Its purpose is to strike at the core of the metainformation's production process, using the logic that any training on a dataset containing copyrighted works is tainted with infringement. The crucial feature that exposes the character of the claim is the non-expressive character of training reproduction.[49] While physically the work is reproduced, this act does not expand access to anyone to enjoy or consume the value of the work as an expressive work.[50] Reproduction in the belly of a machine for technical purposes does not further extend the use value of the expression. The upshot? A reproduction that is merely incidental to the technical process of extracting metainformation and does not extend access to the use value of an expressive work is all about the production of metainformation and its social effects. It has nothing to do with the production/use dynamic of the expressive work and individual harm to its owners.

The disguise of a metainformation claim as an expressive work one is effective due to two rhetorical strategies. First, the entire spotlight is placed on a reproduction of a physical object, leaving in darkness the crucial question of whether this physical fact has any bearing on the purpose or subject matter of the reproduction prohibition.[51] Second, there is a fast and loose deployment of the notion that discrete works are "used" in training. In a very loose sense, the

---

[47] *See e.g.*, Getty Images v. Stability AI, Complaint; Andersen v. Stability AI, Complaint; Christopher T. Zirpoli, *Generative Artificial Intelligence and Copyright Law*, US Congressional Research Service, May 11, 2023 (Observing that the "training process may involve making digital copies of existing works, carrying a risk of copyright infringement."); Enrico Bonadio, Palmen Dinev, and Luke McDonagh, *Can Artificial Intelligence Infringe Copyright? Some Reflections*, in RYAN ABBOTT ED., RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND ARTIFICIAL INTELLIGENCE 247 (2022) (Hereinafter "Handbook on IP & AI") (observing that training copies "may violate the right to reproduction").

[48] See 17 U.S.C. §106(1) (Giving copyright owners the right "to reproduce the copyrighted work in copies"). 17 U.S.C. §101 (Defining "copies" as "material objects… in which a work is fixed").

[49] Matthew Sag, *Copyright and Copy-reliant Technology*, 103 NW. UL REV. 1607, 1068 (2009) (explaining the concept of non-expressive uses); Mathew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295 (2023).

[50] See Bracha *supra* note 39 at 26 (arguing that "Non-expressive copies involve no enjoyment of any expression qua expression" and therefore do not infringe).

[51] *Id.*, at 23-24 (Criticizing the "Physicalist Fallacy" of this kind of reasoning).

notion is correct: to the extent, one employs the term "use" to denote extracting metainformation from a work. But the use is not one of the work as an expressive information good, which means allowing someone to consume the value of the expression qua expression. At the end, the rhetorical strategies cannot change the fact that the argument has nothing to do with the production/use of expressive works and everything to do with an attempt to control the aggregate metainformation production process.

Based on this analysis I argued elsewhere that the claim of infringement by training copies should be rejected on subject matter grounds: while such copies are reproduction under the statutory definition, they do not fall within the subject matter coverage of copyright.[52] By contrast, much of the scholarship on this topic advocates rejecting training copies claims on a different doctrinal ground. Applying a host of precedents from cases dealing with non-expressive uses, this approach concedes that training copies constitute prima facie infringement, but concludes that the reproduction, being non-expressive, is exempted as fair use.[53] The subject matter approach is preferable because it better captures the substantive grounds for the outcome—the fact that the relevant action is not within the domain of copyright at all. It is less complex to administer by both courts and litigants and is thus less prone to confusion and erroneous application.[54] For current purposes, however, this disagreement is of limited significance.

What is more interesting are some thoughtful responses from commentators who advocate finding training copies infringing by attempting to distinguish precedents that in other technological contexts exempted non-expressive uses as fair use. One argument in this vein is that training reproduction is different from other non-expressive reproduction because GenAI technology is "market encroaching."[55] The term means that, even in cases where there is no output similar to copyrighted works in the training set, the low-cost output of the system competes as substitutes with such works, thereby inflicting market harm on their owners.[56] A different argument is that, unlike prior non-expressive reproduction cases, the training reproduction actually is expressive in the relevant sense.[57] Such reproduction is expressive, the argument goes, at least in cases of training expressive GenAI, because the training does not simply extract any metainformation from the works (for example data that indexes references to certain content). The

---

[52] *Id.*, at 26 (arguing that non-expressive copying is not infringing because it "does not involve any copyrightable subject matter").

[53] See e.g. Lemley & Casey *supra* note 46, at 745; Samuelson *supra* note 38, at 159-161; Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, UCLA L. Rev. (Forthcoming); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887, 1914 (2024); Bonadio et al *supra* note 47, at 247; Jessica L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2680 (2020); Daryl Lim, *AP & IP Innovation: Creativity in An Age of Accelerated Change* 52 AKRON L. REV. 813, 847 (2018); James Grimmelmann, *Copyright for Literate Robots,* 101 IOWA L. REV. 657, 661-665 (2016).

[54] Bracha *supra* note 39, at 28-30.

[55] See Sobel *supra* note 37, at .81-82.

[56] *Id.*

[57] Brauneis *supra* note 18.

whole point of the training is to extract data about patterns of expression. This makes the use highly interested in the works as expression and therefore expressive.[58]

Ironically both objections, upon reflection, demonstrate exactly how the training copies infringement argument is a metainformation, rather than a discrete expressive work, claim. Saying that GenAI is "market encroaching" is simply restating the concern that the technology might displace human creators from markets for expressive goods. This concern is founded not on appropriating the expressive value of specific works, but rather on the cost-efficiency of GenAI in creating new, competing substitutive goods and the social effect that could follow in its wake. In other words, it is an argument rooted in the social production/use dynamics of the metainformation engine of the systems, rather than in the production/use dynamics of any discrete work. Similarly, saying that the metainformation extracted from works is "expressive" in the sense that it is about aggregate expressive patterns and relations simply identifies the kind of metainformation involved. Rather than denying that the relevant production/use dynamic is that of the metainformation and its social effects, the argument confirms it. One may identify the metainformation as expressive in that sense, but the fact remains that there is no extension of access to the expressive use value of protected discrete works.

A different argument that dresses metainformation claims in discrete works garments is that the GenAI model itself is a derivative work based on the works in the dataset, even in the absence of similar output. The reasoning here is that because the metainformation in the model is extracted from the discrete works, it constitutes a "derivative" of those works whose preparation is within copyright's rights to exclude.[59] Here the thin rhetorical packaging of the claim as a discrete works one is even easier to discern. So much so that one court dismissed this argument as "nonsensical."[60] While in some loose sense the model is based on the discrete works in the training set, it is not in any way a use of any concrete expression in those works *qua expression*, that is in a way that allows others to consume the expression, now transformed or recast in a "derivative" form.[61] Instead, the model, rather than being recast expression, is something very different, namely: information *about* aggregate patterns and relations within a group of expressive works.[62]

---

[58] *Id.*

[59] Kadrey et l v. Meta Platforms, Inc. 2023 WL 8039640, at 1 (N.D. Cal. 2023) (plaintiff arguing that "LLaMA language models are themselves infringing derivative works… because the models cannot function without the expressive information extracted."); New York Times v. Microsoft, Complaint (alleging that "the GPT LLMs themselves have 'memorized' copies" of the work in the training set which are "unauthorized copies or derivative works").

[60] Kadrey v. Meta Platforms Inc., 2023 WL 8039640, at 1.

[61] Oren Bracha, *Generating Derivatives: AI and Copyright's Most Troublesome Right,* 25 NORTH CAROLINA J. L. TECH. 345, 378 (2024).

[62] *Id.* The analysis assumes the model is not used to generate output substantially similar to a copyrighted work. For a discussion of whether a model should be considered a copy when it is used to generate similar output see *infra* Part V(A).

Another variant of the derivative argument switches the focus from the model to the output of the system by arguing that all such output, irrespective of any similarity, is a derivative of all works in the training set.[63] The heart of the claim is again that all output is, in principle, based on works in the dataset in the sense that it was generated using patterns and relations extracted from these works. Although to date this argument has enjoyed slightly more success than its model-as-derivative counterpart, it fails for similar reasons.[64] The argument trades strategically on a conceptual confusion. The output is based on works in the training set in the loose sense that it is generated by using aggregate patterns identified in those works during the training process, but it does not incorporate any specific recast expression from any specific work.[65] Once again, the argument turns out to be a metainformation claim lurking beneath a thin discrete expressive work cloak.

Why these rhetorical strategies? Why the insistence on disguising social metainformation claims as ones about individual harm to discrete works? One answer is in the question: quo bono? GenAI systems generate considerable social value and a successful copyright claim striking at the heart of the system provides a leverage to its holder for extracting a substantial chunk of that value. A different, more benign, answer is that some claimants are motivated by genuine concerns about the social dangers of expressive GenAI.[66] Copyright seems to be the most readily available vehicle for asserting these social policy claims—indeed, given the limited repertoire of existing cultural policy tools, perhaps the only one. Whatever mix of the two reasons is at work, resorting to copyright law requires packaging the claims as ones about the infringement of copyright in a specific work. Since copyright is about individual interests and harms with respect to discrete expressive works, to enter the copyright club, social metainformation claims must be dressed up in the proper attire.

C. Why Not

Why not? Even if many copyright arguments are social metainformation claims dressed up as discrete expressive works ones, what is wrong with that? Indeed, whether the claim is disguised or not, what is wrong with turning to copyright to provide solutions to genuine social problems created by GenAI's power, unleashed by its metainformation engine? The immediate legal answer

---

[63] See Tremblay v. Open AI, 2024 WL 557720, at 3 (N.D. Cal. 2024) (plaintiff arguing that "every output of the Open AI language models is an infringing derivative work"); Kadrey et l v. Meta Platforms, Inc. 2023 WL 8039640, at 1 (plaintiff arguing that "every output of the LLaMA language models is an infringing derivative work."); Andersen et al. v. Stability AI, Ltd. et al., 2023 WL 7132064, at 1 (plaintiff arguing that "every hybrid image is necessarily a derivative work" because "[e]very output image from the system is derived exclusively from the latent images").

[64] Anderson v. Stability AI, 2023 WL 7132064, at 8 (N.D.Cal. 2023) (expressing doubts that "copyright claims based on a derivative theory can survive absent 'substantial similarity type allegation'" but refusing to dismiss claims due to plaintiff's allegations that some output was substantially similar to plaintiffs' "style").

[65] Bracha *supra* note 61, at 383-84.

[66] Ard *supra* note 37, at 59.

is that social metainformation claims do not fall within copyright's subject matter domain.[67] Copyright applies to, and only to "works of authorship," a term which means specific expressive information goods.[68] Moreover, even with respect to expressive works copyright only applies to the expressive elements of that work.[69] Copyright does not extend to the content of information— designated in copyright doctrine as unprotectable "ideas."[70] And that is true whether the informational content is the one directly conveyed by a work or is metainformation about the work.[71] Metainformation about expressive works, including about aggregate patterns and relations in a group of works, is not expression. It is not part of the forms that humans can consume and derive aesthetic experience from, but simply information about such forms. Consequently, claims about use or extraction of metainformation are categorically outside copyright's domain.

This doctrinal reasoning is anything but dry formalistic legalism. To the contrary, it is grounded in deep substantive reasons that explain why copyright is an inadequate institutional tool for achieving the goals of proper regulation of metainformation and its social effects. BJ Ard has argued recently that "Copyright Cannot Solve Our AI Problems."[72] In the event that GenAI training on copyrighted works is found to be infringing wholesale, the most likely scenario is that producers will shift to train on public domain or licensed materials.[73] Far from removing the prospects of GenAI dominance in expressive markets and its perceived threats, the cost-efficiency of such systems in generating new materials will still drive their displacement of human creators. And far from adequately compensating creators for lost market share—before we ever mention non-market-value harms such as loss of the inherent value of creative opportunities—the most likely result, is some transfers to a few intermediaries who command the legal power to license large pools of works.[74] These include social media platforms who enjoy a broad latitude to use their users' materials by virtue of boilerplate terms of their user agreements, and licensing intermediaries who have already amassed large pools of materials.[75]

Worse still, rather than simply being ineffective, the copyright fix to GenAI's social problems is likely to backfire. The need to restrict training to licensed or public domain materials is likely

---

[67] Bracha *supra* note 39 at 22.

[68] 17 U.S.C. § 102(a).

[69] See e.g. Feist Publications, Inc. v. Rural Telephone Service Co., Inc., 499 U.S. 340, 359 (1991) (explaining that "copyright protects only the author's original contributions"); Ets-Hokin v. Skyy Spirits Inc., 323 F.3d 766, (9th Cir. 2003) (finding no infringement when "subtracting the unoriginal elements" left plaintiff "with only a 'thin' copyright").

[70] 17 U.S.C. § 102(b); Nichols v. Universal Pictures Corporation, 45 F.2d 119, 122 (observing that "too generalized an abstraction" is unprotectable because it is an "idea").

[71] Bracha *supra* note 39 at 17.

[72] Ard *supra* note 37, at 59.

[73] *Id.,* at. 65-66.

[74] *Id.*, at 69 (observing that "artists stand to gain" little from a right to exclude training on copyrighted works).

[75] *Id.*, at 68.

to exacerbate some of the very social policy problems that copyright is called upon to remedy. Contrary to the FTC's sanguine but little-grounded implied assumption that copyright and competition AI concerns point in the same direction, it seems far more likely that broad copyright liability for training on copyrighted materials will aggravate rather than ameliorate concentration concerns.[76] Copyright-induced dependency on a few intermediaries controlling large pools of materials and the cost of licensing, is sure to further raise buriers to entry for producing foundation models and exacerbate, rather than ameliorate, the problems of economic concentration and the ills that it brings in its wake.[77] Similarly, restricting the pool of training materials due to copyright and licensing constraints, is likely to aggravate concerns over homogenous GenAI output, diversity and pluralism.[78]

Ard's shrewd prediction of the inefficacy of copyright in remedying GenAI's social problems, raises a further question: is this merely a happenstance or are there deeper reasons that underwrite copyright's poor performance with such tasks? The answer is resoundingly the latter. Copyright's poor performance is rooted in its fundamental inadequacy as an institutional tool for addressing the relevant problems.[79] This inadequacy is rooted in the three characteristics of cultural social policy claims that emanate from features of social metainformation discussed above.[80]

First, there is a mismatch between the economic mechanism that drives GenAI metainformation cultural policy problems and the specific policy problem around which copyright is designed. Copyright is a solution to a concrete policy problem whose origin is the cost gap between producing and copying expressive works.[81] Generally, the cost of producing expressive works is significantly higher than copying existing ones. This gap works in tandem with another

---

[76] The FTC letter is a comment submitted in response to the U.S. Copyright Office's solicitation of comments on the subject of copyright and AI. *See* U.S. Federal Trade Commission, Comment Letter on U.S. Copyright Office Notice of Inquiry on AI and Copyright, 5-6, (Oct. 30, 2023), available at https://www.regulations.gov/comment/COLC-2023-0006-8630. The letter is far from a model of clarity, and so the FTCs assumption of consistency between broad copyright liability and competition concerns is implicit. The letter lists the FTCs a concern over the question of "[h]ow should liability principles apply to harm caused by AI tools trained on creative work that are used to generate new content?" side by side with a seemingly consistent concern over "entrenching the market power of… dominant firms."

[77] See also Ard *supra* note 37, at 68. For concerns over concentration in AI markets and related social effects see *supra* text accompanying notes 22-26.

[78] For these concerns see sources cited in *supra* note 41. For a similar argument see Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018) (arguing that broad copyright liability for AI systems could exacerbate bias problems with these systems by restricting the training set).

[79] See generally Blake Reid, *What Copyright Can't Do*, PEPPERDINE L. REV. (forthcoming) (analyzing the institutional inadequacies and limitations of copyright in addressing certain policy concerns in another context).

[80] See *supra* text accompanying notes 14-19.

[81] See Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 994-99 (1997); Oren Bracha & Talha Syed, *Beyond Efficiency: Consequence-Sensitive Theories of Copyright*, 29 BERK. TECH. L. J. 229, 237-38 (2014).

feature of expressive works: nonexcludability which means that, once published, it is very hard to exclude others from enjoying the work.[82] The upshot of these two features is that once a work is produced and made available, others will obtain access to its use value for a very low cost or price, thereby reducing the producer's ability to obtain profit, possibly to the point of not being able to recoup production cost. Any plausible normative basis of copyright—whether it is labor-desert, welfarist, or democratic—is rooted in this production/use dynamics and their effect on creators' fair compensation and incentive to create.[83] Property rights in expression—also known as copyright—is designed as a solution to this specific problem.[84]

GenAI metainformation policy problems have nothing to do with these dynamics. Their origin is not the gap between the costs of producing and copying any specific expressive work, but rather the cost efficiencies of producing new expressive works driven by the power of metainformation. This production cost efficiency is what enables the social benefits of GenAI. However, because these efficiencies outweigh the performance of human creators, often with comparable goods produced (or at least a superior combination of production cost and goods' quality), the result can be market displacement of creators and the attendant troubling implications.[85] The concerns related to concentration in GenAI markets, similarly, have nothing to with the production/copying cost gap of expressive works. Those are generated by steep production cost of the metainformation raising buriers to entry and positive network externalities in use.[86] In short, copyright is designed to address a specific cost gaps problem that is irrelevant for metainformation cultural policy concerns. Consequently, copyright has little to offer by way of solutions that target the economic dynamics that do drive those concerns.

Second, copyright is designed to address individual harms visited on discrete works, which makes it particularly ill-fitted for the deep social character of the problems associated with metainformation. Copyright is a property rights regime. As such, copyright—its entitlements and its remedies—are built to address individual market harms inflicted on individual owners of specific expressive works. This is not to say that the normative basis of copyright is necessarily limited to strict individual interests or claims. This may appear to be the case with labor desert, natural rights justification of copyright.[87] But many other normative justifications, ranging from

---

[82] See generally See Amy Kapczynski & Talha Syed, *The Continuum of Nonexcludability and the Limits of Patents,* 122 YALE L.J. 1900 (2013).

[83] See Bracha & Syed *supra* note 81, at 247 (observing that "the same incentive/access tradeoff at the heart of efficiency analysis is also highly relevant for the application of all the consequence-sensitive theories" including rights-based theories of copyright).

[84] To complete the standard account: copyright's property right ameliorates the cost gap problem by allowing creators to exclude others and charge a marked-up price. However, this right to exclude and the pricing power that comes with it creates a cost on access to expressive works that must be balanced against its beneficial effect. *Id.*, at 238-41.

[85] See *supra* text accompanying notes 42-45.

[86] See *supra* text accompanying notes 27-31.

[87] Upon scrutiny the most plausible version of labor-desert justifications is based on a claim for equitable reward for the intellectual labor of creators, which is a social justice claim.

welfarism to various democratic theories, set some social function or goal—anything from maximization of welfare to shaping a public sphere conducive of individual or collective self-governance–as copyright's lodestar.[88] However, even in the latter cases copyright is assumed to promote the coveted goal, which operates on a societal level, by protecting individualized claims against individual harms.

This structure of promoting social goals by protecting private claims has always been a source of tension within copyright.[89] At the end, however, the copyright regime is based on the assumption that the individual claims it protects can be calibrated in a way that tracks and serves the social goals.[90] With metainformation cultural policy claims this assumption disintegrates. These claims are manifestly and deeply social in nature. The harm they identify—whether flowing from the market displacement of creators or the concentration of GenAI markets—are systemic and distributed across society.[91] This social character of the claims is so strong that it is hard to see an institutional design that plausibly connects copyright's means of protecting individual claims in individual expressive works to the deeply social ends it is tasked to achieve, in an effective and tractable manner.

Third and finally, the extra-market character of GenAI cultural policy concerns is a square peg that does not fit withing the round hole of the market-oriented institution of copyright. Copyright, being a property right, is a market-based institution. It consists of rights to exclude that address the cost-gap problem by enabling owners to charge a price premium.[92] Copyright operates through market prices. Some valorize the market-orientation of copyright as making it superior to other institutional alternatives.[93] Others highlight the drawbacks of the market-price mechanism.[94] In

---

[88] See Bracha & Syed *supra* note 81, at 244-47 (discussing various consequence-sensitive theories of copyright).

[89] An early and unpersuasive attempt to reconcile the two is James Madison's argument in the Federalist 43 with respect to patents and copyright that "The public good fully coincides in both cases with the claims of individuals." THE FEDERALIST NO. 43 309 (Benjamin F. Wright ed. 1961).

[90] A prominent example of this assumption in copyright is its balance between providing incentive to create and limiting access cost. This balance is achieved through the various features of individual rights such as copyright's limited term or scope. However, in modern copyright the incentive/access balance always operate on a collective, social level. In a standardized general regime, the incentive/access tradeoff operates *across works*, not per work, with the aim of achieving an overall roughly attractive balance. See Bracha & Syed *supra* note 81, at 240-41 (observing that in copyright there is "an incentive/access tradeoff between different works").

[91] See *supra* text accompanying notes 43-45. See generally Nathalie A. Smuha, *Beyond the Individual: Governing AI's Societal Harm*, 10 Internet Pol'y Rev.1 (2021) (distinguishing individual, collective, and societal harm and exploring the relevance of the latter to AI).

[92] See Bracha & Syed *supra* note 81, at 238-39.

[93] See generally Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J. L. & ECON. 1 (1969).

[94] See generally Amy Kapczynski, *The Cost of Price: Why and How to Get Beyond Intellectual Property*

general, prices and markets work best within an efficiency normative framework, whose criterion is synonymous with market logic, i.e.: the maximization of value measured by willingness and ability to pay. Once the normative basis shifts away from efficiency—whether it is some version of individual or collective self-determination, human flourishing, or distributive equity—copyright's market price mechanism begins to falter. Thus, the market-price contours of copyright's institutional shape have always produced some friction within copyright law and policy.

When it comes to metainformation cultural policy concerns, The squeaky wheels of the discrepancy between extra-market normative goals and copyright's market orientation turn into a deafening roar. These concerns are strongly extra-market in the sense that to have any bite, they must reject a market efficiency criterion.[95] The reason: the driving force that generates the concerns is exactly the cost-efficiency in production of GenAI systems. Consider, for example, the concern that displacement of human creators from expressive markets might diminish opportunities for enjoying the inherent benefits of creative activity. An onlooker taking a market efficiency perspective might be puzzled by the concern. If the value of human creative activity for those engaged in it outweighs the cost savings of GenAI production, then human creators will absorb the cost difference and go on creating. If alternatively, GenAI cost savings are greater than the value of creative activity, then it is a good thing that the latter will be diminished to obtain the former. Any response to this objection necessitates departing from a normative criterion of market efficiency or wealth maximization, by rejecting efficiency's valued "good" (wealth or preferences satisfaction measured by willingness and ability to pay), its desired social "function" (overall social maximization), or both. Once this Rubicon is crossed, however, it is hard to see how copyright can serve as the tool for promoting desirable policy in this area. When tasked with such profoundly extra-market policy goals copyright—whose institutional form is deeply entangled with market prices and market value—is likely to misfire or even backfire every time.

In sum, there is a deep mismatch between the cultural policy problems traceable to the social dynamics of metainformation and the institutional form of copyright as a means for redressing these problems. The result is that copyright is an extremely inadequate tool for the task, one that is likely to deliver poor or even counterproductive results. Moreover, a turn to copyright will result in a vicious feedback loop. We turn to copyright because of the dearth of other institutional tools for addressing cultural policy problems. However, this only fuels our blinkered institutional imagination, our sense that there is little else out there and that copyright is the one-stop solution-center for all our cultural policy problems. What we need to do instead is look beyond our blinkered institutional purview and devise other solutions that better fit the problems—the economic dynamics from which they emanate, the social character of the harms and policy tradeoffs involved, and the extra-market orientation of the values at stake. These are considerable

---

Internalism, 59 UCLA L. REV. 970 (2012). With respect to competing institutional tools in innovation policy and their market orientation see Yochai Benkler, *Intellectual Property and the Organization of Information Production*, 22 INT'L REV. L. & ECON. 81 (2002).

[95] See Bracha *supra* note 39 at 43-44.

tasks that merit separate full treatment elsewhere. Accordingly, I turn now to analyzing GenAI claims that do properly fall within copyright's domain.

IV. EXPRESSIVE GOODS CLAIMS

GenAI expressive works claims are claims about individual harms related to the production/use dynamics of discrete expressive goods. Such claims fall within the proper subject matter domain of copyright. The discussion bellow explains the two main elements of analyzing these claims. First, the hallmark of expressive works claims, that belong in copyright, is generated output that expands the access to the use value of a concrete expressive work. Second, even when similar output exists and the claim is properly about expressive works, one cannot simply assume that liability applies directly to upstream GenAI producers. Rather, considerable further analysis is required to determine who is the actor on whom direct liability should be imposed, and which actors should be regulated by more indirect forms of liability.

This Part lays out the expressive-output criterion for triggering a valid copyright claim. It then explains how GenAI producers occupy the position of gatekeepers due to the locus of their activity in the interface between metainformation and expressive works. Turning to doctrine, the Part analyzes how copyright's main mechanism for deciding on whom to impose direct liability—known as "volition" doctrine—should be applied to GenAI producers, given their structural position as gatekeepers.

A. Substantially Similar Output

While many GenAI copyright infringement claims are in fact metainformation claims in disguise, others are proper expressive goods claim that belong in copyright. This is the case when the New York Times argues that the GPT model, whose training set included many of the Time's copyrighted articles, is sometimes used to generate output that is almost identical to these texts.[96] The same is true when a large language model trained on lyrics of many songs is used to generate as output texts that are highly similar to some of the lyrics.[97] Such cases fall squarely with the subject matter domain of copyright.

The litmus test for a proper expressive work claim is whether the relevant action expands in some way access to consumption of the expression in the work, qua expression. Typically, the practical meaning of this criterion in GenAI cases is that the system was used to generate output that is substantially similar to a copyrighted work. Why is this the criterion? Because potential human access to the expression is what implicates the use value of the information good that is copyright's subject matter, meaning, enjoyment of expressive forms. And this triggers the production/use dynamics of expressive works and the costs gap problem that copyright is designed

---

[96] New York Times v. Microsoft, Complaint (alleging that "the current GPT-4 LLM will output near-verbatim copies of significant portions of Times Works when prompted to do").

[97] Concord Music Group, Inc. v. Anthropic PBC, No. 3:23-cv-01092 , U.S. District Court for the Middle District of Tennessee, Complaint (alleging that defendant's "AI models respond to user prompts by generating identical or near-identical copies of Publishers' copyrighted lyrics").

to address.[98] Thus, the similar output criterion captures the cases that are within copyright's subject matter domain and which copyright's institutional tools are designed to handle.

Relevant cases are not limited to instances of reproduction. The extension of access to the work's expressive value can be done by any of the activities recognized by copyright's exclusive entitlements.[99] A GenAI system could be used to publicly display a copyrighted work,[100] publicly perform it,[101] or even to prepare a derivative work.[102] All of those actions, as long as human access to protected expression is extended, fall within copyright's ambit.[103] Finally, the criterion of expanding access to the work's expression only regulates the entry gate to copyright's domain. To win an infringement case a claimant still has clear all the usual hurdles of a copyright claim, including establishing infringement by showing that the generated output is substantially similar to the copyrighted work,[104] and surviving any defense including a claim that the fair use in the specific circumstances.[105]

B. Lumpy Information – Dual Use

When a GenAI case falls within copyright's domain, the next challenging question is who has direct legal responsibility for the infringement.[106] Some take it for granted that when infringing output is present, the producer who trained the model is directly liable.[107] This conclusion is too hasty. GenAI involves a supply chain of different actors running from upstream producers to downstream users.[108] For reasons of deep pockets and effective control, claimants impute direct liability for infringing output to the producer, as in cases where an internet host actively chooses

---

[98] See *supra* text accompanying notes 81-84.

[99] See 17 U.S.C. §106.

[100] Lee et al *supra* note 3, at 65-66.

[101] *Id.*

[102] Bracha *supra* note 61, at 396.

[103] Note that copyright does not place on copyright owners the heavy burden of establishing that someone has already taken advantage of the expanded access to the work's expression and consumed its expressive value. It is sufficient to establish reproduction, making derivative works, or public display or performance, even if no one perceived the expression yet because of any of those actions. However, copyright does require, under its subject matter principles, that access to expression was somehow extended in a way that creates the potential of actual consumption. This structure is second-order prophylactic design related to the division of evidentiary and practical burdens between opposing parties. Bracha *supra* note 39, at 27-28.

[104] *See* Arnstein v. Porter, 154 F.2d 464, 473 (2nd Cir. 1946).

[105] 17 U.S.C. §107.

[106] See Michael P. Goodyear, *Artificial Infringement*, (unpublished), at 45. Note that the discussion here is of liability for infringing output. The distinct claim that the model itself is an infringing copy whenever it generates similar output is discussed *infra* Part V.A.

[107] See Sag *supra* note 49 at 1914; Felix T. Wu, *The Structure of Secondary Copyright Liability*, 61 HOUS. L. REV. 385, 393 (2023) (observing GenAI are not ones in which the producer is "acting as a platform").

[108] See *supra* text accompanying notes 5-6.

which materials to upload to its systems.[109] Producers, by contrast, would argue that they are simply purveyors of neutral technology who are not subject to liability for infringing uses of their service, as in the case of a passive server provider whose system hosts materials selected by others, producers of VCRs, or business owners who sell access to their photocopiers.[110] In these latter cases, copyright law imposes liability on end-users for the specific use of the technology and subjects producers or service providers only to secondary forms of liability whose scope is limited in various ways.[111] Ironically, both sides to this disagreement are likely to invoke as support the partial-autonomy of GenAI systems-- the fact that the output is not a result of a preset set of rules and therefore is unpredictable.[112] One side may argue that this unpredictability prevents end-users from being the agents responsible for output generated in response to their prompts, placing the responsibility on producers. The other may respond that generated output is equally unpredictable to producers, placing the responsibility on users who initiate the specific generation through their prompts.[113]

How should the fundamental question of the responsible agent be resolved? The key is not the degree of predictability by producers or users, but their structural position with respect to producing the infringing output. The distinction between social metainformation and discrete expressive works is critical in this analysis because the way that actors interact with each of these distinct information goods shapes their structural position. Users participate in the generation process on the level of a discrete work, while producers shape it as gatekeepers through their activity on the level of metainformation. This distinction is of great import for the doctrinal analysis of direct liability and its underlying purpose to be discussed below.[114]

To understand the respective structural positions of producers and users assume a concrete example, say an image of Daffy Duck that is generated in response to a user's prompt "self-absorbed, whacky, greedy, ill-tempered cartoon duck."[115] Consider the producer first. What is its causal contribution to the generated Daffy? Training and making available for use a model whose informational content is such that the resultant response to the specific prompt is Daffy's image. The producer, by training the model, operates on the level of metainformation. However, the action results in effects in the domain of discrete works—enabling and shaping the generation of specific expression.

---

[109] See UMG Recording, Inc. v. Escape Media Grp., 2014 WL 5089743, at 22 (2014) (finding direct liability when plaintiff instructed employees to upload copyrighted music files to its website).

[110] See Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984); CoStar Grp. v. LoopNet, Inc. 373 F.3d 544, 550 (4th Cir. 2004).

[111] See *infra* Part V.B.

[112] See Goodyear *supra* note 106, at 44-45 (Highlighting the difficulty in identifying the direct infringer in GenAI infringement cases due to "the unique semi-autonomous nature" of the technology).

[113] *Id.*, at 46-52 (Discussing placing direct liability on producers or users).

[114] See *infra* See *infra* Part V.D.

[115] The example assumes that the generated image is substantially similar to an image of Daffy Duck that was in the system's training set.

This causal chain—an action on the metainformation level with effects on the level of discrete expression—is important because of the *aggregate* or *social* character of metainformation.[116] The specific generated output is causally traceable to design choices of the producer, but no design choice can be easily pinpointed as responsible discretely only for that output while not affecting other informational features and therefore functions of the system. When a GenAI system generates output similar to an item in its training set, we can say that the informational content of the work was conveyed by and therefore contained in the model.[117] Nevertheless, models are not giant databases, as in collections of discrete works. There is no specific part of the information in the model that represents discretely any particular work, just as there is no specific part of the formula y=ax+b that discretely represents a concrete pair of coordinates. Rather the information of a specific work is "conveyed" by the metainformation socially: by being bundled inseparably with information about other countless works. Daffy is, informationally speaking, "somewhere" in the model, but he is disintegrated into countless fragments that are fused with other information. Consequently, redesigning any feature of the model that allows extracting Daffy from it is likely to affect other functions and uses of it.[118]

Why does this social character of metainformation design choices matter? Because it makes the design of a metainformation resource *lumpy*. To borrow the words of Lee Fenell, models are "*lumpy* or *indivisible* because this is how the good delivers its value."[119] Redesign one part of a machine and you are likely to affect a wide range of its functions. Even more so, because of its social character, specific designs of models that affect concrete expressive outputs, are likely to affect others. The social character of models is what enables their utility in creating new information goods by generalizing, but it is also what makes their design choices lumpy.

This lumpiness in design places the model producer in the position of a gatekeeper of dual-sue technology. Unlike a server provider who chooses specific materials to host, producers do not engage in discrete activities of extending access to specific expressive works. Instead, they make lumpy design choices on the level of metainformation; choices that affect a wide swath of concrete uses, both legitimate and infringing. In this respect, producers are in the same structural position as manufacturers of VCRs and other information technology. They are gatekeepers of dual-use technology whose design choices operate on the social level to enable concrete beneficial uses and

---

[116] See *supra* text accompanying notes 14-17 .

[117] See A. Feder Cooper and James Grimmelmann, *The Files are in the Computer: On Copyright, Memorization, and Generative AI*, Chicago-Kent L. Rev. (forthcoming), at 6. Available at: https://ssrn.com/abstract=4803118.

[118] To be sure, a producer can remove the Daffy images from the training set. In isolation this design strategy is likely to prevent the infringing use while not seriously affecting any other uses of the system because no one item is of much consequence for training the model. The problem is that this strategy does not scale. The aggregate effect of removing from the dataset every item whose generation as output might infringe is likely to have a serious effect on the performance and non-infringing uses of the system.

[119] LEE ANNE FENELL, SLICES & LUMPS: DIVISION AND AGGREGATION IN LAW AND LIFE 9 (2019).

impose risk of infringing ones.[120] The parallel is incomplete. Unlike completely neutral technology such as VCRs, models impose risk of infringement on a specific and finite group of works: those included in their training set.[121] Nevertheless, the crucial fact remains: model producers, through their lumpy design choices, create aggregate social benefits and risks; they do not make decisions on the level of individual works.

The actor who engages with the technology on the discrete-work level is located closer to the specific generation process—typically, but not always it would be the end-user.[122] In the example above, the user who provided the prompt that resulted in the generation of the Daffy image engages in a discrete action that triggers and shapes the specific generation process. The point is not that the causal contribution of the user to the generated result is more important than that of the producer, but rather that its operation is limited to the level of generating the specific work. Nor is the point that the user intended or even predicted the generation of the specific image. For better or worse, copyright's strict liability makes questions of intent or knowledge of infringement irrelevant.[123] The point is that users, unlike producers, make their causal contribution to the generation of expressive output on a granular level that is limited to the specific individual work produced.

To understand how the very different structural positions of producers and users of GenAI shape the policy and legal analysis, we need first to explain the rather obstruse doctrine that governs the question of the responsible actor, known as "volition."

C. From Volition to Scope of Liability

When the question arises who among several actors, including producers of copying-facilitating technology, is responsible for an act of infringement copyrights go-to doctrine is "volition."[124] The term suggests that the substantive inquiry is whether an actor acted voluntarily or not, and many take it at face value.[125] However, the term is simply a misnomer. Laying aside exotic cases, locating liability in copyright cases has nothing to do with the voluntariness of the

---

[120] Mark A. Lemley & R. Anthony Reese, *Reducing Digital Copyright Infringement Without Restricting Innovation*. 56 STAN. L. REV. 1345, 1354 (2004).

[121] Cooper & Grimmelmann *supra* note 117, at 54.

[122] The discussion here brackets here the important question of the appropriate legal treatment of of the actor deploying the system, who is located in between model producers and end-users.

[123] See, e.g., Buck v. Jewell-LaSalle Realty Co., 283 U.S. 191, 198 (1931); De Acosta v. Brown, 146 F.2d 408, 411 (2d Cir. 1944); ABKCO Music, Inc. v. Harrisongs Music, Ltd., 722 F.2d 988, 180 (2d Cir. 1983).

[124] Religious Technology Center v. Netcom On-Line Communication Services, Inc., 907 F. Supp. 1361 (N.D. Cal. 1995).

[125] See e.g. Mala Chatterjee & Jeanne C. Fromer, *Minds, Machines, and the Law: The Case of Volition in Copyright Law*, 119 COLUM. L. REV. 1901 (2019) (describing the volition requirement as asking whether the relevant act was "genuinely willed or chosen by the so-called actor" and examining whether the AI system as such can have this mental state).

action.[126] The question is not who acted voluntarily, but rather, which of multiple actors—all of whom certainly did engage in "voluntary" (as opposed to involuntary, as in forced or unconscious)[127] "acts"[128] that were part of a causal chain leading to the infringement—should be deemed legally responsible for the infringing outcome.[129] The confusion probably stems from a paradigmatic scenario of a server provider who passively hosts infringing files that are uploaded and downloaded by others.[130] In this case, it is tempting to say that the server's owner is not the responsible actor because they are not the ones doing the uploading or downloading. In a loose analogy to classic non-volition cases in torts where one's body is (forcibly) used by others as a tool, one may wish to say that in this case the owner's server is being used by others with no voluntary involvement of the owner. But this analogy misses the point. The question is not whether the server's owner is responsible for the physical use of their servers by others. The question is whether the owner's own voluntary actions—namely, making the server available for use by others—which unquestionably are causally related to the infringing outcome, form a sufficient basis for imposing *legal* responsibility on them. The same applies to the cable provider who allows users to remotely record programing on its system, or the photocopier owner who allows others to use their machine.[131]

At this point, some might want to draw on a term familiar from tort and say that the question is really one of "proximate cause."[132] The question is which among multiple actions contributing to the harmful outcome is the "predominant" cause or "closer" to the harm. To spare the reader decades of intellectual debates and oceans of scholarly ink, it suffices to say that the post-Legal-Realism consensus is that so-called proximate cause questions have *nothing* to do with

---

[126] See Robert C. Denicola, *Volition and Copyright Infringement*, 37 CARDOZO L. REV. 1259, 1262 (2016) (Observing that the copyright requirement "Isn't Actually an Issue of "'Volition'")

[127] *See* RESTATEMENT OF THE LAW, SECOND, TORTS, § 14 (1964) ("Necessity of Act by Defendant").

[128] *Id.*

[129] Ironically, even views that insist on describing the doctrine as being about volition sometimes admit that all the relevant actions, including those not seen as giving rise to direct liability, are volitional. See e.g. Cartoon Network, LP v. CSC Holdings, Inc., 536 F.3d 121, 131 (2nd Cir., 2008) (explaining that "*Netcom* and its progeny direct our attention to the volitional conduct that causes the copy to be made" and then concluding that "[t]here are only two instances of volitional conduct in this case" but only one of which "supplies the necessary element of volition").

[130] See CoStar Grp. v. LoopNet, Inc., 373 F.3d 556.

[131] Cartoon Network, 536 F.3d 131.

[132] *Netcom* itself described the element that decide which actor has liability as requiring "volition or causation." Netcom, 907 F. Supp., at 1361. 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT §13.08[C][1] (2019) ("*Netcom* simply stands for the unremarkable proposition that proximate causation historically underlines copyright infringement liability no less than other torts"); David Nimmer, *Volition in Violation of Copyright*, 43 COLUM. J. L. & ARTS 1, 6 (2019) (the volition requirement is about cases where the "chain of causation is too remote to support legal responsibility"); Perfect 10, 847 F.3d, at 666.

causation.[133] In such cases all the relevant actions are causally related to the harm, in the basic but-for, sense by which questions of "factual causation" are typically settled.[134] And no amount of "physical" or "scientific" (or metaphysical!) reasoning regarding which cause is more "proximate," can resolve what is a social policy question, namely: what is the appropriate scope of legal responsibility of various actors for the harmful consequences of their actions.[135] The question, in other words, is about the "scope of the liability" that various actors owe to others, which is inescapably a question of policy or justice.[136]

Returning to the GenAI copyright context: how should the question, now properly understood to be about the scope of liability rather than either voluntariness or causation of the various actors involved, be best analyzed with respect to each of them? Which of these actors should be held responsible for the infringing, harmful outcome in which they are all causally implicated; namely, the generation of substantially similar output? In tort law, the scope of the liability inquiry—i.e., deciding which of the downstream causal effects of an actor's conduct properly lie within their legal responsibility—is often resolved on the basis of foreseeability.[137]

---

[133] See DAN B. DOBBS, PAUL D. HAYDEN & ELLEN D. BUBLICK, THE LAW OF TORTS (2nd ed. 2011), §198 (Observing with respect to "proximate cause that "that well-worn term has been justly criticized for years as inaccurate, misleading, and confusing, and has been rejected by the Third Restatement of Torts"); *Special Note on Proximate Cause,* RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL HARM (TENTATIVE DRAFTS), CHAPTER 6: SCOPE OF LIABILITY (PROXIMATE CAUSE) ("Although the term 'proximate cause' has been in widespread use in judicial opinions, treatises, casebooks, and scholarship, the term is not generally employed in this Chapter because it is an especially poor one to describe the idea to which it is connected. […] Hence, this Chapter is entitled, "Scope of Liability." That terminology more accurately describes the concerns of this Chapter: Tort law does not impose liability on an actor for all harm factually caused by the actor's tortious conduct. […] [T]his Restatement separates factual cause from scope-of-liability limitations.") (internal footnotes omitted). For the intellectual and social history of the rise and demise of the notion of "proximate cause" *see* MORTON J. HORWITZ: THE TRANSFORMATION OF AMERICAN LAW: THE CRISIS OF LEGAL ORTHODOXY 1870-1960 51-63 (1992).

[134] See Denicola *supra* note 126, at 1268 (Observing that the copyright requirement Isn't Actually an Issue of 'Causation' Either"). Less typical cases of factual causation that may be settled using alternatives to but-for analysis include cases of "multiple sufficient" or "probabilistic" causation. See MARK A. GEISTFELD, TORT LAW 252-55 (2008) (discussing cases of "multiple tortious causes"). The issues posed by these more complex cases and tests of factual causation remain distinct from those involved in determining scope of liability (so-called "proximate cause").

[135] Dobbs et al *supra* note 133 at §198 (observing that "The so-called proximate cause issue is not about causation at all but about the appropriate scope of legal responsibility").

[136] In modern tort law, there is a common analytical distinction—with procedural bite—between scope of liability and scope of duty issues. Historically, courts excluded from liability as "non-proximate cause" certain foreseeable harms, such as emotional distress. However, under modern tort doctrine, exclusion from liability due to foreseeability reasons is usually treated under scope of liability doctrine, while harms excluded for other reasons are treated under the scope of duty category (which operates as a threshold filter that is a question of law for the judge rather than of fact for the jury). Geistfeld *supra* note 134, at 152-53 ,270-71. Copyright has no such distinction. Volition is a catch all category used to exclude liability for certain harms by certain actors whether on foreseeability or, more commonly, other grounds.

[137] Dobbs et al *supra* note 133 at §198.

The principle is avoiding imposing responsibility on actors for harms that were not reasonably foreseeable as consequences of their actions. In copyright law, too, the issue of foreseeability may arise.[138] However, many copyright harms, especially in the context of infringement-facilitating technology, are perfectly foreseeable. To a reasonable server owner or application provider it is quite foreseeable that some others might use the technology that they make available to engage in infringing conduct. The same is true of the person offering access to the photocopying machine and the cable provider, who clearly should expect that some of their users would use the technology for infringing and harmful purposes.

This raises several questions. Should the scope of liability of the actors in the above examples be limited? If so, what is the ground for such limitation, other than foreseeability? Furthermore, does any such ground for limiting the scope of liability of certain technology providers apply to GenAI model producers?

D. GenAI Gatekeepers

As Robert Denicola shrewdly observed: "The volition requirement in copyright law defines the connection between the owner of a copying system and the copied work that is sufficient to justify attributing the copying of that work to the owner."[139] Specifically, limiting the scope of liability via "volition" doctrine functions as a policy lever that calibrates copyright's regulation of producers of dual-use technology. Dual-use technology, such as VCRs or internet servers is capable of legitimate and beneficial uses as well as infringing and harmful ones.[140] Two main considerations are interlaced in analyzing the question of what kind of copyright liability, if any, should be imposed on such actors. First, the legal standard must balance the social interest in reaping the technology's benefits against the interest of reducing its risk of harmful uses.[141] Second and relatedly, the interest of treating producers as centralized gatekeepers who are apt targets for cost-effective enforcement against harmful individual uses must be balanced with the concern of over-incentivizing such gatekeepers to make design decisions that frustrate beneficial uses.[142]

To balance these interests, copyright avoids imposing direct liability simply for producing or making available dual-use technology, even when users use it to infringe.[143] Underlying this rule

---

[138] For example, for a landlord who rents out general-use real estate it may be unforeseeable that a tenant would use the premises for a bootlegging operation. For the role of foreseeability in copyright law see generally Shyamkrishna Balganesh, *Foreseeability and Copyright Incentives*, 122 Harv. L. Rev. 1569 (2009).

[139] See Denicola *supra* note 126, at 1272. I wholeheartedly agree with Denicola's observations, but I develop a somewhat different account of what the sufficient connection and its purpose are.
[140] Lemley & Reese *supra* note 120, at 1354.

[141] *Id.*, at 1379; Douglas Lichtman & William Landes, *Indirect Liability for Copyright Infringement: An Economic Perspective,* 16 Harv. J.L. & Tech. 395, 397 (2003).

[142] Lemley & Reese *supra* note 120, at 1375-76. See generally Reinier H. Kraakman, *Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy*, 2 J.L. Econ. & Org. 53 (1986).

[143] Sony, 464 U.S. 442 (1984) (ruling that "[T]he sale of copying equipment" does not constitute infringement if the product is "capable of substantial noninfringing uses").

are two assumptions about sweeping gatekeeper liability for infringing uses: 1) that it would give too much power to those with individual harm claims over innovative technology, espcially emergent technology whose full social value has not yet developed,[144] and 2) that it risks over-impairing the legitimate uses of the technology by creating skewed incentives to producers.[145] The doctrinal device for reaching this result is the rule that gatekeepers who simply provide technology, rather than engage in individual decisions on how the technology is used, have no "volition."[146] Stripped of its misleading connotations, the rule simply means that the harm caused by individual infringing uses to copyright owners is outside the scope of liability of such technological gatekeepers. Direct liability is applied instead to end-users who put the technology to specific uses in individual cases. Gatekeepers are then subjected only to secondary liability that requires various additional elements and circumstances in which the gatekeeper is specifically involved in the individual infringing act.[147]

How should GenAI producers be treated with respect to direct liability for infringing uses of their systems? At first blush, it might seem that such producers are fundamentally different from traditional technology producers. GenAI models are not VCRs or servers.[148] A model producer does not simply make neutral or "empty" technology that is then applied to specific expressive works by users. Producers use a specific group of works in the dataset to produce the model.[149] In doing so producers impose risk of individual infringing uses, not on an unspecified and potentially infinite set of works determined by users, but on the concrete group of works included in the dataset.[150] A closer look reveals that despite this difference, there is a critical similarity between model producers and other technology makers: their structural position as gatekeepers of dual-use technology.[151] This role of the model producer, recall, is structured by the level of its involvement with infringing uses. Producers operate on the metainformation level which is lumpy in design.[152] Unlike a server provider who chooses to host specific works, they do not make discrete decisions on concrete infringing uses.[153] Instead, like passive server hosts or VCR producers, they make broad technological design choices that are likely to have varied effects across broad swaths of uses both harmful and beneficial.[154] In short, model producers are a special brand of gatekeepers

---

[144] Lemley & Reese *supra* note 120, at 1378-89.

[145] Assaf Hamdani, *Who's Liable for Cyberwrongs?,* 87 CORNELL L. REV. 901, 910 (2002).

[146] Netcom, 907 F. Supp., at 1361.

[147] See *infra* Part V.B.

[148] See Cooper & Grimmelmann *supra* note 117. For a discussion of whether models are themselves infringing copies of the works see *infra* Part V.A.

[149] Cooper & Grimmelmann *supra* note 117.

[150] *Id.*

[151] Frosio *supra* note 41, at 7-9 (analyzing GenAI as a dual-use technology).

[152] See *supra* text accompanying notes 119-121.

[153] *Id.*

[154] *Id.*

of dual-use technology. As such, deciding whether they should bear full direct liability for all infringing uses or be shielded by restricting the scope of liability, requires assessing the effects of these competing rules on producers' incentives and ultimately over the social costs and benefits of the individual uses of their systems.

What are the relevant risks and benefits of expressive GenAI? On the risks side, we have to lay aside both those risks relating to *social* cultural policies that are best handled by institutional means other than copyright,[155] and those pertaining to *individual* harms to interests distinct from those protected by copyright's right to exclude.[156] Unsurprisingly, what is left is the broad risk of the technology to facilitate individual acts of generating infringing materials, by providing powerful, low-cost, and widespread means for (re)producing expression.

The benefits side consists of two aspects: consumptive and productive benefits. The consumptive benefits are the more obvious ones—they pertain to GenAI's remarkable capacity for cost-effective satisfaction of demand in markets for expression.[157] Less apparent are the productive benefits of GenAI. Productive benefits consist of empowering human creative activity, rather than GenAI simply taking over production.[158] The source of these benefits is the use of GenAI in various hybrid models of creation, where the technology can supply raw materials, some of the expressive building blocks of larger projects, or simply creative tools integrated into more standard digital platforms.[159]

An important aspect of the production benefits of GenAI is their potential to deepen the democratization of cultural participation brought about by an earlier generation of digital

---

[155] See *supra* Part III.C.

[156] Consider, for example, concerns over divulgement of individual's personal information, or, even closer to copyright's domain but not to its right to exclude, attribution concerns about misrepresentation of a work's authorship.

[157] Bracha *supra* note 39 at 37 (referring to GenAI's "potential for high quality expressive production at low cost" and "more effective satisfaction of demand"); Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669, 1680 (2023) (observing that "AI threatens to lower the costs of creation itself, so that creative works… are cheap to produce in the first place").

[158] See Bertin Martens, Economic arguments in favour of reducing copyright protection for generative AI inputs and outputs, Working Paper 09/2024, at 3 (Arguing that "GenAI increases human productivity, not only for language, cultural and audiovisual media, the traditional domain of copyright protection, but across a wide range of industries and human tasks where GenAI models are applied"). See in general E. Brynjolfsson, D. Li & L. Raymond, *Generative AI at Work,* NBER Working Paper 31161, National Bureau of Economic Research (2023), at 1 (Finding, in a different context that "AI assistance increases the productivity of agents" and arguing that this happens because "generative AI systems work by capturing and disseminating the patterns of behavior that characterize the most productive agents, including knowledge that has eluded automation from earlier waves of computerization").

[159] Martens *supra* note 158, at 2 (observing that "human artists use GenAI to leverage their media productivity and explore innovative new media outputs"); Kulesz *supra* note 41, at 5 (observing that "the most effective formula is collaboration between human being and machine: so, far from doing away with artists, AI can enhance their capabilities"); Ard *supra* note 37, at 66-67 (discussing Adobe Firefly an integrated GenAI-based tool that allows powerful image manipulation); W Knight, When AI Makes Art, Humans Supply the Creative Spark, Wired, July 13, 2022.

technology.[160] GenAI tools, widely available at modest cost, can empower cohorts of non-professionals engaged in user innovation to produce expressive materials at an unprecedented quality.[161] As Katrina Geddes puts it, with GenAI "[o]rdinary individuals can now create sophisticated synthetic media by modifying, remixing, and transforming cultural works without any artistic training or skills" which "radically expands the range of individuals who can engage in aesthetic practice."[162] The point here is not simply that anyone, acting as a consumer, can through a simple prompt obtain a digital image of their liking, but rather that many, by enlisting the image-production power of GenAI and combining it with their own creative input, can generate a new comic book at a quality level that previously was beyond their reach.[163] The result of this democratization is substantial social benefits, not only, or perhaps at all, in welfarist terms, but also to values such as individual or cultural self-determination and human flourishing via socially equitable access to the benefits of creative activity.[164]

The question is how to shape legal incentives to producers, so their behavior results in an attractive mix of robust benefits for an acceptable level of risk. One answer is that producers should bear the entire social cost of the risk imposed by their technology. In simple terms, producers of GenAI should have direct copyright liability whenever their systems generate substantially similar materials. Such a rule will not eliminate the risk.[165] GenAI systems will always remain susceptible to some specific infringing uses. The hope is that by being forced to internalize the full infringement cost of their systems, producers will be adequately motivated to invest in its reduction. In other words, producers will be motivated to take reasonable precautions; namely: invest in such means and make design choices whose benefit in reducing harmful copyright infringement outweighs the cost, including any accompanying degradation of the system's potential for beneficial uses.[166] As a bonus, sweeping liability to the producers cuts on enforcement

---

[160] *See* e.g. Anupam Chander & Madhavi Sunder, *Everyone's a Superhero: A Cultural Theory of "Mary Sue" Fan Fiction as Fair Use*, 95 CALIF. L. REV. 597, 599–600 (2007); LAWRENCE LESSIG, REMIX (2009); William W. Fisher, *The Implications for Law of User Innovation*, 94 MINN. L. REV. 1417, 1418-22 (2010).

[161] Kulesz *supra* note 41, at 5 (observing that "AI lowers entry barriers and makes it possible for many more people to compose symphonies, make movies and write novels").

[162] Katrina Geddes, *How Art Became Posthuman: Copyright, AI, and Synthetic Media* (unpublished).Available at SSRN: https://ssrn.com/abstract=4865510, at 1.

[163] These hybrid techniques for creativity and their empowerment of low-resourced creators are revealed in following the details of some of the disputes involving attempts to register works created with GenAI involvement and the Copyright's Office to develop criteria for which aspects could be registered as made by human authors. *See* e.g. Letter from Robert J. Kasunic, Associate Register of Copyrights, Re: *Zarya of the Dawn*, to Van Lindberg, Taylor English Duma LLP (Feb. 21, 2023).

[164] See Bracha & Syed *supra* note 81, at 276-281; Fisher *supra* note 160, at 1458-72; Geddes *supra* note 162, at 33-35.

[165] Unless the cost of liability is so high that the system will not be produced at all.

[166] In other words, imposing direct copyright liability on producers for all copyright infringement by use of their system is tantamount to imposing strict liability on them for those harms with the outcome predicted by tort theory, when additional complications are ignored, being the taking of reasonable precautions. See *infra* text accompanying notes 268-270.

cost by allowing right owners to enforce against one centralized entity with effective control over the system, rather than attempting to chase countless end-users.[167]

Unfortunately, the hope for adequate precautions through full internalization of harm is too sanguine. Since we do not live in a frictionless world of perfect markets, sweeping direct liability for system producers is unlikely to result in precautions that adequately balance the technology's social costs and benefits. The reason: GenAI producers do not come close to internalizing the full social value of their systems. To the contrary, there are good reasons to assume that these systems are leaky and rife with spillovers of widely spread social benefits.[168] Nor would we want producers to fully internalize those benefits. On grounds of both equity and non-market values related to cultural creation, a world in which the tech giants suck the last drop of benefits created by GenAI from their legions of users is hardly an attractive one.[169] Together with the assumption of full internalization of benefits falls the postulation that full internalization of harms will lead to well-balanced precautions.[170] A producer calibrates its design decisions not to the social value of its technology but to its own private value. If producers are hit with the full infringement cost of their technology, but enjoy only a fraction of its social benefits, the result will be a systematic misalignment of incentives: over-valuing risk-reducing precautions relative to discounted social benefits.

The practical result with respect to expression-generating GenAI is a familiar chilling effect: a tendency to over-block or otherwise reduce the system's capacity with respect to the range of potentially generated materials due to a skewed over-incentive to steer away from infringement risk.[171] As in the case of previous generations of technologies and platforms—from internet servers to content hosting applications—full internalization of harm by an intermediary is sure to lead to

---

[167] Lemley & Reese *supra* note 120, at 1376.

[168] Martens *supra* note 158, at 1 (observing that GenAI creates substantial "spillover effects or externalities"); Frosio *supra* note 41, at 9 (observing in the context of GenAI that dual use technology has "significant positive externalities that benefit society at large").

[169] See e.g. Brett Frischmann, *Capabilities, Spillovers, and Intellectual Progress: Toward a Capabilities, Spillovers, and Intellectual Progress: Toward a Human Flourishing Theory for Intellectual Property* 14 REV. ECON. RESEARCH ON COPYRIGHT ISSUES 1 (2017) (examining justifying spillovers in intellectual property on the basis of the Capabilities Approach); C. Edwin Baker, *Giving the Audience What It Wants*, 58 Ohio St. L.J. 311, 316 (1997) ("the market's identification of preferences and its reliance on the existing distribution of wealth are often inappropriate for determining the production and distribution of media content.").

[170] Lemley & Reese *supra* note 120, at 1376 (explaining that the fact that "intermediaries do not bear the full social cost of taking down challenged content means that enforcing copyright law by requiring them to do so creates negative externalities, tilting the law too far in favor of copyright owners"); Dotan Oliar, *The Copyright Innovation Tradeoff: Property Rules, Liability Rules, and Intentional Infliction of Harm*, 64 STAN. L. REV. 951, 1014 (2012) (acknowledging that the "copyright-innovation intersection" analysis must be adjusted in cases of significant externalities).

[171] Hamdani, *supra* note 145, at 916 (2002) (arguing that "given the different incentive structure" of intermediaries "imposing strict liability" on them "would result in overdeterrence in the form of excessive monitoring… and overzealous censorship policies."); Wu *supra* note 107, at 398.

over-blocking.[172] The chilling effect applies not only to consumptive uses, but also to productive ones. Too zealous infringement-averting measures at the system design level entail not only fewer satisfied consumers, but too deep impairment of GenAI's potential for empowering human creativity both professional and user-oriented. And it is the latter category of distributed user innovation that would suffer the most, for two reasons. First, distributed user creativity tends to rely heavily on strategies of "glomming on" to, adapting, and imbuing with new meaning dominant cultural materials.[173] As a result it is strongly dependent on users ability to incorporate exactly the copyrighted materials that will be most affected by over-blocking.[174] Second, the most significant normative value of distributive user creativity is not in its market, or even welfare-enhancing value. The normative benefit is primarily for values such as individual and cultural self-determination, as well as central dimensions of human flourishing such as broad opportunities for meaningful creative activity and communal interaction.[175] These benefits would be discounted twice by producers: first because their market value is far smaller than their normative social value, and second because even with respect to market value the producer only internalizes a fraction.

The result of over-blocking induced by full direct liability is a new "Snoopy problem." The "Snoopy Problem" has become a term of art in GenAI copyright policy debates.[176] It refers to the fact that the generation of GenAI output substantially similar to copyrighted works is most likely with respect to dominant materials whose patterns gain salience in the training set, such as popular characters (hence "Snoopy").[177] The new Snoopy problem—now being a normative rather than technological—is that Snoopy & co. might be placed out of reach, including for legitimate uses by those who need them the most. The reason is the lumpiness of model-design: the difficulty of implementing system designs that block infringing uses without also blocking or frustrating legitimate ones.[178] The likely result of over-incentivizing producers' precautions is that dominant cultural materials (including, for example, popular characters) would simply be unavailable for non-infringing, socially beneficial uses. Imagine yourself trying to make a Barbie parody or a

---

[172] Sharon Bar-Ziv and Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown*, 50 CONNECTICUT L. REV. 339 (2018).

[173] *See* Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 10-12; Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 535 YALE L.J. 535, 567 (2004) ("Copying can serve as self-expression . . . and it can work as affirmation, a way of connecting to a larger group.").

[174] See Bracha & Syed *supra* note 81, at 272.

[175] See *supra* note 164 and accompanying text.

[176] See Sag *supra* note 49 at 334.

[177] *Id.*

[178] See Ard *supra* note 37, at 33 (arguing under the label "The Spider-Man Problem" that "it is difficult to stop… [AI] systems from reproducing famous characters without also trenching on legitimate expression").

subversive version of Gone with the Wind when the GenAI tools on which you rely simply would not produce anything resembling Barbie, Ken, Scarlet, or Rhett.[179]

Full internalization of infringement harm through sweeping direct liability of GenAI producers, is likely to lead to over-zealous preventive measures with unfortunate results in the cultural sphere. What is the alternative? Under existing law, the main alternative is to set the scope of liability so that the producer is not seen as the direct infringer simply because a model was used to generate infringing output. This leaves as the direct infringer actors with more discrete involvement in initiating the *specific* generation process that produces infringing output. Generally, these would be actors closer to the downstream side of the GenAI supply chain. Typically, but not always, these would be end-users who feed the system with the prompt that triggers a specific output.[180] When (and only when) liability for output is appropriate these actors should bear it; not because they act more voluntarily, more culpably, or with greater foreseeability of the harmful result, but rather because their causal contribution to this result is on the individual level of a discrete act responsible for a discrete harm, rather than the systemic design level that sweeps together harmful and beneficial effects.

E. Objections

Locating direct liability downstream, often with end-users, is likely to attract vigorous objections. Consider two cases of a GenAI system generating a Daffy Duck in response to a user's prompt. In the first case the prompt is "Daffy Duck" In the second it is "whacky duck cartoon figure." Many would not object to direct user liability in the first case, but would find it hard to accept in the second.[181] It seems that the in the second case the user acts innocently and is less responsible for the infringement.[182] And on the flipside, the model producer seems to be more responsible in the second case.[183] After all, since it does not seem to be the user who was bent on having their cartoon duck being Daffy, it must have been the model's designer who made that decision.[184] Despite their surface allure, these instinctive objections should not change the analysis. The attraction of the objections is in the residual power of assumptions about voluntariness, state of mind, or proximity of causation.[185] But scope of liability doctrine is not about any of these elements, but about the better locus of liability in light of the structural position of the actors.

---

[179] See Geddes *supra* note 162, at 41 (observing that "user seeking to challenge the heteronormativity of Marvel comics will struggle to accomplish this using generative AI if prompts containing the names of Marvel characters are automatically blocked or rephrased").

[180] Goodyear *supra* note 106, at 48.

[181] The two examples are based on Lee et al *supra* note 3, at 84.

[182] *Id.*

[183] *Id.*

[184] *Id.*

[185] For the reason why these assumptions are irrelevant see *supra* text accompanying notes 124-136.

As for the user, under copyright's existing strict liability standard, their intent or knowledge with respect to the generated output is irrelevant.[186] State of mind is relevant neither for general liability nor for deciding which actor has "volition."[187] One may believe that it is improper to impose liability in circumstances of lack of knowledge or intention with respect to an infringing result, but this would be a criticism of copyright's extreme strict liability standard, not a reason that goes to specifying the identity of the actor who is subjected to this standard.[188] Recall that copyright's "volition" doctrine has nothing to do with state of mind or voluntary actions.[189] Similarly, despite the label of "proximate cause" the doctrine is also not about the relative importance of one's causal contribution to the infringement.[190] What matters is the structural role of the actors. In both Daffy examples the user is the actor whose actions are on the individual level of a discrete act resulting in the output, rather than the general system design level.

As for the GenAI model producer, even in cases where the user's contribution is less foreseeably connected to infringing output, the producer is not in the same position as a server host who picks specific hosted materials, notwithstanding the surface appearance of similarity. The crucial feature that distinguishes a passive server host from one who is subject to direct liability is crossing the line between providing general purpose technology and making discrete decisions on individual infringing uses. The model producer remains firmly, even in the second case, on the side of lumpy design of technology that is susceptible to both infringing and legitimate uses, rather than making individual use decisions.

At the end, what matters in cases of dual-use technology is neither state of mind nor the relative importance of causal contribution, but the structural position of the actors. Recall that in complex technology cases "volition" is a way of distinguishing the gatekeeper from an actor who uses the technology concretely in a specific case, to assure that a different liability standard applies to each.[191] The relevant question is whether an actor is a gatekeeper who designs technology capable of diverse uses, or someone who uses the technology in an individual instance to produce infringing output.[192] In both examples above the structural positions of the parties remain constant with the producer being the gatekeeper; and so, the scope of their liability should remain constant as well, with the user being the direct infringer.

---

[186] See *supra* note 123.

[187] See Perfect 10, 847 F.3d, at 666 (observing that the volition requirement has nothing to do with an "act of willing or choosing").

[188] See generally Patrick R. Goold, *Moral Reflections on Strict Liability in Copyright*, 44 COLUM. J.L. & ARTS 123, (2021).

[189] See *supra* text accompanying notes 124-131.

[190] See *supra* text accompanying notes 132-136.

[191] See *supra* Part V.D.

[192] For the reasons why the person who directs the specific instance of use, rather than the gatekeeper should be deemed the direct infringer see *supra* text accompanying notes 122-123.

A different objection to locating direct liability with downstream actors is a sense that it is unfair to place the liability burden on these actors while letting the upstream "big fish" escape unscathed.[193] Downstream actors would often be smaller and not deep-pocketed, espcially by comparison to upstream producers. Moreover, the specific uses of generated copyrighted materials might be of small impact. And finally, what of the interest in reducing enforcement cost by centralizing it via gatekeeper liability?

These objections too are mostly misplaced. Some of them go to copyright's general scope. If one thinks that certain personal, non-commercial, or other uses should be allowed beyond what is currently privileged by fair use or other exemptions, either generally or specifically in the GenAI context, then that should be the argument.[194] Moreover, focusing on the individual traits of the specific actions of downstream actors, rather than those of the upstream producer, is exactly the right frame for analyzing such questions. As for upstream producers, they are not allowed to escape completely unscathed. The analysis above suggested they should be free from sweeping direct liability for all infringing activity simply by virtue of producing or training a system, not that they should be free from any liability including one that befits their role as gatekeepers who create social risk and are well placed to reduce it.[195]

In sum, to track the underlying purpose of "volition" doctrine in dual-use technology cases, direct liability should be assigned by the structural position occupied by the relevant parties: as a gatekeeper who engages in lumpy design decisions, or an actor who puts that technology into specific uses in individual cases. Generally, the parties occupying the latter position will be on the downstream side of the GenAI supply chain.

V. Metainformation Cum Expressive Goods

Perhaps the most challenging group of GenAI claims in the field of expression are those that relate to the interface between the spheres of discrete works claims and metainformation claims. Rather than trying to disguise metainformation claims or impute to producers direct responsibility for all similar output, this kind of claims takes a different stance. The claims here assert that actors in the sphere of metainformation should bear responsibility for the effects of their action or inaction within that sphere on harmful actions by others within the distinct sphere of expressive works. The crucial question is how law and policy should treat this interface and what kind of liability should metainformation actors have for the effects of their action with respect to discrete expressive works transgressions. In short, these claims confront the role of metainformation actors as gatekeepers more directly, rather than simply assuming that they are the agents directly responsible for specific infringing output. This Part first criticizes one absolutist answer to the question of upstream liability: that whenever infringing output is produced the GenAI model itself should be considered a reproduction and its maker a direct infringer. The Part then discusses copyright's standard toolkit

---

[193] Ard *supra* note 37, at 63; Benjamin W. Sobel, *Copyright Accelerationism* (unpublished).

[194] See e.g. Geddes *supra* note 162 at 60-71 (advocating the enactment of a privilege "for non-commercial user-generated content").

[195] See *infra* Part VI.B.

for handling gatekeepers, namely: secondary liability doctrines. It argues that these doctrines are rooted in a paradigm of liability for individual, discrete harm and therefore are insufficient to govern the activities of producers whose activity on the level of social metainformation creates broadly distributed social risks.

A. Copy Fundamentalism

One possible approach to GenAI gatekeepers' liability circles back to direct liability for all infringing generated output. Ostensibly, this answer is grounded in technological truism and doctrinal necessity. The argument is that infringing output renders the GenAI model itself a copy of the infringed work because the informational content of that work is conveyed by the model.[196] In such circumstances, the argument goes, the model fits to the letter the statutory definition of "copies" as "material objects… in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[197] The digital files containing the model are physical objects. The specific copyrighted work is "fixed" in these objects. Under the statutory definition a work is "fixed" in a tangible medium of expression when "its embodiment in a copy… is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."[198] Because the work can be perceived, reproduced or communicated from the information in the model, as the generated output indisputably establishes, it is "fixed" in the "copy" of the model.[199] Similarly, the existence of the output also establishes the definitional element of a "copy" that requires the physical object to be one "from which the work can be perceived, reproduced, or otherwise communicated," in this case indirectly "with the aid of a machine."[200] If the physical embodiment of the model is a copy of the work then it follows that the producer of the model directly infringes copyright by reproduction, not because it is responsible for the generated output, but rather because it is responsible for making the model.[201]

Note that the argument has a quantum mechanics quality. Under its logic, although the information is embedded in the model at the time of training, it is only possible to know that a model is a copy of a work once it is successfully used to generate similar output. Until that moment the model hovers, Schrödinger-cat-like, in indeterminate superposition, potentially being both a copy and a non-copy of a particular work. Only the generation of similar output, if and when it

---

[196] See Cooper & Grimmelmann *supra* note 117 ("If a generative-AI model memorizes its training data, the training data is *in the model*"); New York Times v. Microsoft, Complaint (alleging that when plaintiff's works were "memorized" by the model, as evidenced by similar output, their "expressive contents have been substantially encoded within the parameters of the GPT series of LLMs. Each of those LLMs thus embodies many unauthorized copies or derivatives" of plaintiff's works).

[197] 17 U.S.C. §101 (defining "copies").

[198] 17 U.S.C. § 101 (defining "fixed").

[199] Cooper & Grimmelmann *supra* note 117, at 21.

[200] 17 U.S.C. §101.

[201] 17 U.S.C. §106(1) (giving the owner the right to reproduce the work in "copies").

happens, collapses the copy function into a determinate state.[202] Fortunately, further consideration of this kind of (meta)physical speculations can be laid aside. The failure of the argument is rooted in the fact that applying the legal concept of a copy cannot hinge completely on either technological assertions or metaphysical speculations about physical facts, both pertaining to the question of whether the physical embodiment of the model "really" "contains" the specific copyrighted work.[203] Copyright law defines the term "copy" for its own purposes, and therefore the *legal* concept can only be applied in light of the purposes underlying it. It is not dictated by any physical or metaphysical necessities, detached from considering this purpose.[204]

What is copyright's purpose in defining "copies" and then applying direct liability to the person who makes them? The answer is that, given technological and social circumstances, making a copy is a standard and widespread way of expanding access to the use value of the work—an element that goes directly to the production/use dynamics of expressive works on which copyright is focused.[205] And more specifically relevant to our context: whether someone makes a copy is a distinction that separates actors who engage in a discrete act that expands access to the use value of an expressive work, and others who engage in other legitimate activities that may carry with them the social risk of facilitating such specific access-expanding acts. This is the relevant difference between someone who uses a VCR to record a copyrighted movie (a maker of a copy) and someone who produces VCRs (not a maker of a copy), not simply the technological fact that a video cassette "really" contains the movie.[206] Copyright purposes require different legal treatment of these actors and consequently copyright's concept of "copies" must be informed by this purpose.

How then should we read and apply the term "copies" in light of its purpose? Recall that for a physical object to be a copy of a work it is not sufficient that the work "can be perceived, reproduced, or otherwise communicated" from it.[207] It is also necessary for the physical object to

---

[202] The less colorful way of describing this situation is to say that "the model has memorized these data, but we do not currently have the means to know that it has done so." Cooper & Grimmelmann *supra* note 196, at 29. Should this initial indeterminacy be troubling? Perhaps not, but the question is relevant only if one thinks that the appropriate meaning of the term "copies" flows directly from facts about the physical world. The latter is not my view.

[203] Or in Cooper's & Grimmelmann's words whether the individual work is really "in the computer." Cooper & Grimmelmann *supra* note 117. Thus, it is true that "Copyright law does not determine technical facts." *Id.*, at 8. But it is equally true that technical facts do not determine copyright law. To rephrase: for purposes of copyright law "copying" is *not* a "physical fact." *Id.,* at 55. It is a legal concept. And there is an ocean of difference between the two.

[204] See generally Felix Cohen. *Transcendental Nonsense and the Functional Approach*. 35 COLUM. L. REV. 809 (1935).

[205] See *supra* text accompanying notes 98-103.

[206] This is not to say, of course, that there aren't differences between VCRs and models. See Cooper & Grimmelmann *supra* note 117, at 54-57 (explaining why "models are not VCRs"). It is only to say, that the mentioned common feature is one that is highly significant for copyright's purpose and therefore should inform the application of legal concepts.

[207] 17 U.S.C. §101 (defining "copies").

be one in which the work "is fixed."[208] Is the work "fixed" in the model? Generally, courts have identified two elements to the statutory definition of fixation.[209] First, and somewhat overlapping with the definition of "copies," to be a fixation an object must be "sufficiently permanent or stable to permit" that work "to be perceived, reproduced, or otherwise communicated."[210] Second, the first element must obtain "for a period of more than transitory duration."[211]

What is broadly overlooked in this breakdown of the statutory definition of "fixed" is that the text also says that these two conditions apply to "its embodiment," namely, to the physical embodiment of the work. It is not surprising that this term is broadly overlooked because in traditional fixation cases, even those involving complex technological settings, it is safe to assume that when a work can be "perceived, reproduced, or otherwise communicated" from a physical object that object is an embodiment of that work. But with respect to a GenAI model this assumption cannot be taken for granted. Indeed, it is the very question on the table. While the presence of similar output makes it clear that the work can be perceived or communicated from the model, it does not magically, without more reasoning, resolve the question of whether the metainformation in the model from which the work can be constructed is an "embodiment" of the work.[212] And here, unlike other cases, this question requires an answer before we can conclude that the work is "fixed" in the model.

Note that no amount of "plain meaning" reasoning can read this question out of the statutory definition of "fixed." The reference to the "embodiment" being one of the work is in the statutory text. And any argument that this part of the "plain" text is meaningless, superfluous, or reducible to the "permit to be perceived" prong requires support. The support for the opposite position—maintaining that "embodiment" of the work in an object is not completely reducible to any object that permits the work "to be perceived, reproduced, or otherwise communicated"—comes from the underlying purpose. Copyright has very good reasons for treating differently those who make objects whose primary purpose is expanding access to the work's use value (an "embodiment") and others who create objects that serve many different useful purposes but also run the risk of sometimes empowering actions that expand such access (an object from which the work can be made perceivable that is not an "embodiment" of it).[213] The possibility of an object that permits the work to be "be perceived, reproduced, or otherwise communicated" without being an embodiment of it opens up exactly this space that is required by the underlying purpose.

The challenging question then, that is fairly unique to the GenAI context, is whether the model is an "embodiment" of the work whose generation it enables. And while one answer, per above, is

---

[208] *Id.*

[209] Cartoon Network, 536 F.3d at 131.

[210] 17 U.S.C. §101 (defining "fixed").

[211] *Id.*

[212] Unless one tautologically defines embodiment as a physical object from which the work can be perceived, reproduced or otherwise communicated.

[213] See *supra* Part IV.D.

affirmative,[214] another, which is better supported by the purpose of the doctrine, is that it is not. The model is not an embodiment of any particular work. It is, rather, metainformation about aggregate patterns of the works in the dataset, that is capable of many uses. The fact that sometimes the metainformation can be used to recreate a work in the dataset, shows that the information was conveyed by the model, but it does not transform the model from aggregate information about a group of works to an embodiment of a specific work.[215]

Some may strongly object to this reasoning, on a combination of technological and legalistic grounds. When a model generates output closely similar to data in its training set, technologists often say that the data was "memorized" by the model.[216] What they mean is that the data can be extracted (or be "perceived, reproduced, or otherwise communicated") from the model and therefore the model contains or at least conveys that data. One may be tempted to draw the conclusion that this also means that the model is an "embodiment" of the work in the copyright statutory sense. This conclusion seems to be further bolstered by the way that copyright law treats certain kinds of metainformation about expressive works. A digital file in mp3 or any other compression format does not directly contain the information which is the musical or other expressive work "in" the file.[217] What this file contains is metainformation about the work from which the work itself can be reconstructed, given a device that can decode the compression standard. In fact, the same is true on a more basic level: any digital file, even not compressed or encrypted, is metainformation about the expressive work it "contains."[218] We regard these files as containing the works because we use them with standardized devices that seamlessly convert the metainformation into directly perceivable representations of the expressive works. And yet there is no question that copyright law treats such digital files as copies that embody the expressive works.[219] This is the modern approach of technological neutrality in copyright that stands in stark contrast to premodern attitudes.[220] It is facilitated by the capaciousness of the statutory definition that encompasses "any method now known or later developed" for fixing works, whether these

---

[214] See *supra* text accompanying notes 196-201.

[215] To better grasp the point, consider a low-tech analogy. Assume that Jill, a scholar of computational literary theory, produces, as part of her research, a complex set of tables that represent the statistical probability of certain terms following others in Jack's poems. Jill's table are useful for various literary research purposes about Jack's poetic corpus and the work of other poets in his poetic movement. Jillian uses Jill's charts to construct a text that is very close to one of Jack's poems, having never seen the poem. In this case the basic purpose of copyright's doctrine informs the conclusion that Jill's tables should not be seen as a "copy" of Jack's poem, although Jillian made the poem perceivable from the tables.

[216] Cooper & Grimmelmann *supra* note 117, at 15-30.

[217] For a short technical explanation of the mp3 file format see "Technology" at https://www.mp3-history.com/en/technology.html.

[218] Cooper & Grimmelmann *supra* note 117, at 32.

[219] UMG Recordings, Inc. v. MP3.com, Inc., 92 F.Supp.2d 349, 350 (S.D.N.Y. 2000) (finding reproduction of music in MP3 format to be infringing).

[220] For the premodern attitude see White-Smith v. Apollo, 209 U.S. 1 (1908).

methods make the work perceivable directly or indirectly "with the aid of a machine."[221] If copyright treats these forms of metainformation as copies why wouldn't it do the same with GenAI models?

The above reasoning proceeds too quickly. While it may be appropriate for technologists to regard for their purposes models as copies of extractable information, copyright law has its own purposes, and the same conclusion may not follow within it. What matters for copyright law is not whether a technology formally satisfies a technological definition of copying but how the technology functions in practice with respect to relevant human interests.[222] Copyright law treats the making of copies as infringement because in our society this is a widespread and standardized way of expanding the availability of the expressive use value of works. In circumstances where standardized decoding technology is ubiquitous, digital files, including in compression formats, perform this function of expanding access to specific expressive works, and there is good reason to regard them as an "embodiment" of the work and treat them like other copies. But most GenAI models do not perform the same function.[223] A model is different from an mp3 file because of the aggregate nature of the metainformation in it and its lumpy design.[224] Models are not libraries of information about specific works, but aggregate information about collective patterns of large swaths of works, geared toward the generalization necessary for generating new materials.[225] As such, models are not standardized means for expanding access to the expressive use value of discrete works.[226] They are primarily productive engines for generating new works that necessarily carry with them the risky side effect of being used in specific instances to expand access to existing ones.[227] That the risk materialized in a specific case does not change the general functional character of the model.

And this different functional character is a good reason to distinguish a model from digital files, compressed or otherwise. Under existing technological circumstances in our society, the latter are simply means for expanding access to discrete copyrighted works. Accordingly, they are seen by copyright as an "embodiment" of those works and are designated as "copies" in which the

---

[221] 17 U.S.C. §101 (defining "copies").

[222] Thus, Cooper and Grimmelmann are right to say that "The Copyright Act's definitions of fixation and copies are functional not formal." However, they are wrong in inaccurately identifying the relevant function. Cooper & Grimmelmann *supra* note 117, at 21.

[223] In principle, a model could be trained for the sole purpose of reconstructing a single work in which case it should be treated as a copy of this work. See Sobel *supra* note 37, at 63-64.

[224] See *supra* text accompanying notes 119-121.

[225] See *supra* text accompanying 116- 118.

[226] Contrary to Cooper & Grimmelmann' observation that it is only "relative novelty" or "immense scale" of Gen AI models that makes them seem very different than other digital storage technologies. Cooper & Grimmelmann *supra* note 117, at 33.

[227] This, of course, would not be true in the unlikely case that the considerable resources needed to train large foundation models would be invested to create a model with such a high level of fit and so little capacity to generalize that its main use is as a glorified form of a database that expands access to the works in its dataset. For the principled possibility of such a model see Sobel *supra* note 37, at 63-64.

work is "fixed." Models, by contrast, are primarily means for generating new works that impose risks of some unauthorized expanded access to existing works. As such, their producers should be treated not as the makers of copies who bear direct liability, but as technological gatekeepers whose beneficial activity should be balanced by the law against the social risk they create, even in specific cases when that risk materializes. None of this denies that system producers can limit infringing outputs including by model design decisions.[228] It is only to say that they should be treated as gatekeepers who create system-wide risks and benefits through design choices, rather than parties responsible for discrete transgressions. The doctrinal way of achieving this result is conceptualizing GenAI models as aggregate metainformation about discrete works rather than an "embodiment" of them, and hence not seeing them as "copies" in which specific works are "fixed."

B. Secondary Liability

I have argued that the best way to conceptualize model producers is neither as responsible directly for infringing output nor as reproducers of works in the model itself. Instead, their treatment should reflect their position as making design choices on the metainformation level; namely, as gatekeepers who produce dual-use technology with many beneficial uses but also attendant social risk of infringement. Copyright's primary doctrinal instrument for regulating such technological gatekeepers is secondary liability doctrines. These doctrines originated in tort law principles but acquired a life of their own in copyright.[229] And while they sometimes applied to discrete or individualized activities, their main function, espcially in cases involving technology, has become to govern the activity of gatekeepers.[230] In this capacity secondary liability doctrines function in a generalized or even socialized manner: they regulates entities who create systemic social risk of infringement across a broad swath of works or are well positioned to mitigate such risk.[231] Despite this systemic character of the activities they often regulate, secondary liability doctrines remained moored in different ways to the individual harm paradigm from which they emerged.[232] The clearest expression of this is the fact that all secondary liability doctrines require establishing specific, direct infringement by someone, only by reference to which, gatekeepers can be found secondarily liable.[233] Once direct liability is established, two groups of doctrines apply in a non-mutually-exclusive manner. Vicarious liability is based on the supervisory power of the

---

[228] Cooper & Grimmelmann *supra* note 117, at 6, 61-64.

[229] Metro-Goldwyn-Mayer Studios, Inc. v. Grokster Ltd., 545 U.S. 913, 930 (2005).

[230] See Wu *supra* note 107, at 392-93.

[231] *Id.*

[232] *Id.*, at 393-96.

[233] 2 PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT §8 (3d ed. 2005). Note, however, that in technological gatekeepers cases, the application of this individual direct infringement element is often stretched in various ways to encompass systemic risk scenarios. In these cases, the direct infringement analyzed would often be an undifferentiated mass of widespread infringing activity. See e.g. A&M Records, Inc. v. Napster, Inc, 239 F.3d 1004, 1014 (9th Cir. 2001) (framing the analysis with the proposition that defendant's "users are engaged in the wholesale reproduction and distribution of copyrighted works,").

gatekeeper over the directly infringing activity, while the different variants of contributory liability are based on a more "horizontal" element of facilitating the infringing activity.

1. Vicarious Liability

Vicarious liability is based on a party's supervisory relationship with the direct infringer. It applies when a party has a right and ability to control the infringing activity as well as a direct financial interest in it.[234] In the context of technological gatekeepers, the most salient feature of vicarious liability is the non-systemic and status-quo oriented way in which the right and ability to control element is applied. Typically, courts analyze this element by reference to some combination of the gatekeeper's formal right to supervise and regulate, its practical capacity to do so, and the extent to which it actually exercised this capacity.[235] Crucially, this analysis is done with respect to a specific relationship between the defendant and the direct infringer.[236] Moreover, this relationship is analyzed under the status quo of the actual system design.[237] Courts require an existing "monitoring and supervisory relationship" and generally reject arguments that sufficient control could be achieved by redesigning the system, introducing precautions, or restructuring the relationship between the parties.[238] Instead, they insist that the right and ability to police "is cabined by the system's current architecture."[239]

Under this standard, GenAI producers may be subject to vicarious liability in some circumstances. For example: if a producer has a specific corporate relationship with the direct infringer or is involved in the deployment of a system in a way that allows for monitoring and control of users' activity, these circumstances may give rise to the kind of specific supervisory power that courts require for imposing vicarious liability.[240] Nevertheless, this form of liability is limited by its focus on the individual relationship under the existing technological and institutional

---

[234] Shapiro, Berenstein & Co. v. H.L. Green Co., 316 F.2d 304, 307 (2nd Cir. 1963).

[235] Goldstein *supra* note 233, at §8.2.2.

[236] Metro-Goldwyn-Mayer Studios, Inc. v. Grokster Ltd., 380 F.3d 1154, 1164 (9th Cir. 2004) (observing that "right and ability to supervise" element "describes a relationship between the defendant and the direct infringer").

[237] *Id.*, at 1166.

[238] *Id.* (rejecting an argument that right and ability to control exists because "the software itself could be altered to prevent" infringement); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1175 (9th Cir.2007) (ruling that defendant's failure to change the mode of operation of its search engine "is not the same as declining to exercise a right and ability to make third-party websites stop their direct infringement"); Luvdarts, LLC v. AT&T Mobility, LLC, 710 F.3d 1068, 1071 (9th Cir. 2013) (rejecting an argument that right and ability to supervise exists because defendant could establish a system "that would give them the right and ability to supervise the infringing activity").

[239] Napster, 239 F.3d 1024.

[240] An additional, factor in shaping the scope of the cases where vicarious liability is imposed goes to the question of how broadly courts interpret the element of direct financial interest in the infringing activity. See Alfred C. Yen, *Third-Party Copyright Liability after Grokster.* 91 MINN. L. REV. 184, 200, 205 (2006); Lemley & Reese *supra* note 120, at 123-24.

architecture. Under the auspices of this doctrine courts largely refuse to look at the more systemic question of what this architecture could or should be.

2. Contributory Liability

Contributory liability is based on a more horizontal, facilitative or cooperative relation to direct infringement. It applies when a party has knowledge of the direct infringement and is materially inducing, causing, or contributing to it.[241] A key element of the doctrine as it regulates technological gatekeepers is the way courts apply the knowledge requirement. Courts insist on "actual knowledge of specific acts of infringement" rather than general knowledge that the "system allows" some infringement.[242] While there is some ambiguity on whether "reason to know" might be sufficient, even such lesser degree of knowledge must be with respect to specific, individualized act of infringement.[243] And similarly, while some case law suggests that willful blindness may suffice to satisfy the knowledge requirement, courts tend to insist that this too must apply to a subjective suspicion of specific infringing activity.[244] In addition courts insist that the knowledge of concrete infringement must be simultaneous with an ongoing contribution to the infringement.[245] The net result of courts' insistence on knowledge of specific acts of infringement simultaneous with ongoing contribution is a highly individualized character of contributory liability. The doctrine does not impose liability for awareness of a general infringement risk imposed by the system and failure to mitigate it, but only for knowing acts of facilitating specific infringing actions.[246]

There are two more variants of contributory liability. The first involves cases where a party has no actual knowledge of a specific infringing activity but only "constructive knowledge," meaning

---

[241] See Gershwin Pub. Co. v. Columbia Artists Mgmt., Inc., 443 F.2d 1159,1162 (2nd Cir. 1971).

[242] Napster Inc., 239 F.3d at 1021. Aee also Luvdarts, LLC v. AT&T Mobility, LLC, 710 F.3d at 1073 (contributory liability requires "specific knowledge of infringement").

[243] See Casella v. Morris, 820 F.2d 362, 365 (11th Cir. 1987) ("The standard of knowledge is objective: 'Know, or have reason to know'"); Napster, 239 F.3d at 1021 (defendant "knew or had reason to know"); Viacom Intern., Inc. v. YouTube, Inc., 676 F.3d 19, 30 (2nd Cir. 2012) (Ruling in an analogous statutory context that the phrase "facts or circumstances from which infringing activity is apparent" in 17 U.S.C. § 512(c)(1)(A) refers to "'knowledge of specific and identifiable infringements.'").

[244] Compare In re Aimster Copyright Litigation, 334 F.3d 643, 650 (7th Cir. 2003) ("Willful blindness is knowledge, in copyright law") with Luvdarts, LLC v. AT&T Mobility, LLC, 710 F.3d at 1073 (ruling that to show willful blindness plaintiff must show that defendant "took deliberate actions to avoid learning about" specific infringing action on its systems).

[245] Grokster, 380 F.3d 1162 (explaining that plaintiff must show that plaintiff had knowledge of a specific infringing activity at the same time that it contributed to it).

[246] BMG Rights Management (US) LLC v. Cox Communications, Inc., 881 F.3d 293, 310 (4th Cir. 2018) (ruling that "proving contributory infringement requires proof of at least willful blindness; negligence is insufficient").

general statistical awareness that the system is susceptible to infringing uses.[247] Under *Sony Corp. of America v. Universal City Studios, Inc*, to establish liability on this basis the technology must not be one that is "capable of substantial non-infringing uses."[248] Thus, hiding under the misleading term "constructive knowledge," is the one branch of secondary liability that directly looks at system-wide risks and benefits. However, the reach of the doctrine is deliberately limited. Liability is imposed only in the limited group of cases where the technology has very limited beneficial potential. Thus, Sony's constructive knowledge liability, despite an unresolved Supreme Court split on the exact standard for its safe-haven, applies only in extreme and rare cases where the technology at issue is hardly a dual-use one.[249]

The second variant of contributory liability is inducement.[250] It applies when a party "distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement."[251] Like constructive knowledge, the doctrine operates on the product or system level, but its focus on a culpable state of mind shifts it away from considerations of systemic risks and benefits.[252] The Supreme Court muddied the water on this issue by enumerating system design and failure to take precautions as possible objective indicia from which intention can be inferred.[253] But realizing the friction with *Sony*, it restricted such considerations to the status of cumulative factors that cannot in themselves establish inducement.[254] Accordingly, inducement is somewhat of a chimera: some considerations of system design can be smuggled in under cover of objective indicia of intention, but at the end, the doctrine's overall grounding in culpability prevents it from being a vehicle for methodically evaluating the systemic risk and benefits of technology.

As with vicarious liability, the various branches of contributory liability could apply to GenAI producers, in specific technological and institutional circumstances.[255] A firm may train a model with such a level of fit to its training set or deploy it in such a manner that makes the system

---

[247] See Sony, 464 U.S. 417 (1984); Napster, 239 F.3d 1020 (distinguishing between knowledge based on the "architecture of the… system" and based on specific "conduct in relation to the operational capacity of the system").

[248] Sony, 464 U.S. 442.

[249] See Grokster, 545 U.S. 948 (J. Ginsburg concurring) (Sony's safe haven requires showing "a reasonable prospect that substantial or commercially significant noninfringing uses" are "likely to develop over time") and Grokster, 545 U.S. 057 (J. Breyer concurring) (Sony's safe have applies "unless the product in question will be used almost exclusively to infringe copyrights").

[250] There is some disagreement on whether inducement is a variant of contributory liability or an independent form of secondary liability. See WILLIAM F. PATRY, PATRY ON COPYRIGHT 21:79 (2018).

[251] Grokster, 545 U.S. 919.

[252] Wu *supra* note 107, at 394-95; Yen *supra* note 240, at 227 ("Grokster sent a strong signal that the Supreme Court considers fault the primary theory of third-party copyright liability").

[253] Grokster, 545 U.S. 939.

[254] *Id*., fn 12.

[255] See Goodyear *supra* note 106, at 64-69.

incapable of substantial non-infringing uses. It may have a business relationship with a further downstream user of a model it trained that makes it aware of specific infringing uses. And it may operate in a way that betrays a clear intention to encourage infringement from users. How broad this set of cases is depends on how courts interpret key elements of the doctrine such as the strictness of the concrete and simultaneous knowledge requirement, the exact meaning of *Sony*'s safe-haven formula, or the extent to which system-design factors can establish intention. However, it is clear that these doctrines operate in a—narrower or broader—penumbra of individual cases, rather than applying to the core of GenAI's systemic social risk.

3. Taking Stock

In sum, doctrines imposing secondary liability for infringing output may apply in specific circumstances to upstream GenAI producers. How wide this range of circumstances is depends on the extent to which courts are willing to loosen the key limiting elements of these doctrines.[256] At the end, the range of cases where secondary liability is imposed upstream is limited. The reason is a mismatch between the individualist underpinnings of the doctrines—their focus on specific harms, individual supervisory or facilitative relations, or individual culpability—and the systemic character of the risk of GenAI with respect to expressive works.[257] The more typical and significant risks imposed by these systems are not about knowing facilitation of specific infringing activities. Instead, the main risks are of a more general or distributed nature: enabling infringement on an aggregate social level as a side effect of the system. Similarly, the more significant and typical GenAI risks involve, not discrete failures to exercise existing supervisory power, but rather questions of system-wide design, namely: questions about how systems should be designed in the first place to reduce the systemic risk of infringement.

Even when it comes to those variants of secondary liability that begin to grope toward a more systemic framework, the risk imposed by GenAI systems tends to be so incidental to the purpose of these doctrines that it would often fall outside their ambit. The significant value, both social and private, of these systems is overwhelmingly in independent satisfaction of demand and empowering production rather than as tools for generating infringing output. Thus, the typical case would be a system whose primary use is new production rather than primarily as an engine of infringing output wrapped in token non-infringing uses. The result: cases of constructive knowledge that fail the *Sony* safe harbor, despite the split on its exact content, are unlikely to be very common.[258] Similarly, while not impossible, cases where the business logic behind a GenAI system is profiting from the value of the system as a vehicle for infringing output are not likely to be central. As a result: inducement cases, despite some room for play with deducing intention from design features, are not likely to be central in this area.

Secondary liability doctrines have a proper role to play in regulating the activities of GenAI technological gatekeepers. However, because of the mismatch between the persisting individual

---

[256] Yen *supra* note 240, at 193-212.

[257] See Wu *supra* note 107, at 393.

[258] *Id.*, at 394 (observing that "platforms are always capable of 'substantial noninfringing uses'").

harm and relations focus of these doctrines and the deep systemic or social character of GenAI's primary risks in the expressive realm, this role is limited.

VI.   NEGLIGENCE

Is there anything missing in the existing framework of copyright law as applied to GenAI claims that properly fall within its domain? The analysis below suggests that copyright has a liability gap. Caught between the overkill of direct liability and inaptness of secondary liability, it has insufficient tools to adequately deal with GenAI's systemic social risk. The analysis further argues that the way to address this gap is by subjecting GenAI gatekeepers to a negligence standard, a standard that could be applied in several doctrinal ways.

A. The Liability Gap

When it comes to GenAI gatekeepers copyright suffers from a liability gap. The source of the gap is the polarity of   existing legal alternatives. GenAI gatekeepers can be treated as direct infringers with respect to all infringing output. The likely result of bringing to bear this heavy artillery is over-blocking whose deleterious effects are visited mainly on users.[259] The alternative is subjecting GenAI gatekeepers only to secondary liability. While secondary liability doctrines can deal with specific scenarios of improper behavior, their clinging to an individualized framework makes them inapt for properly addressing the most significant issue, namely: incentivizing gatekeepers to adequately reduce the systemic risk associated with their systems.[260] Neither of these alternatives is adequate in achieving what should be the main concern in this area, which is moving GenAI gatekeepers to design and operate their systems in ways that reduce the overall social risk of infringement associated with them, while not unduly compromising their demand-satisfaction and productive social benefits.[261] Put differently, the major concern in this area is in the paradigm of the law of *accidents*: an actor who is engaged in a legitimate and useful activity that comes with the incidental risk of harmful side-effects.[262] It is socially useful and desirable for the railroad (or a GenAI system) to be built and operated. The pertinent question is how to deal with the attendant systemic risk of the activity. We would not necessarily like to deter the actor from engaging in the socially useful activity altogether or force it to obtain permission from every single person who incurs risk of being adversely affected. But we may want to induce it to reduce the risk associated with its activity by appropriately designing and conducting its operation. Copyright law currently has no mechanism for handling this accident framework.[263]

---

[259] See *supra* text accompanying notes 171-179.

[260] See *supra* text accompanying notes 256-258.

[261] See generally Margot Kaminski, *Regulating the Risks of AI,* 103 B. U. L. REV. 1347, 1351 (2023) ("lawmakers have turned to the tools of risk regulation to govern AI systems").

[262] Oren Bracha & Patrick R. Goold, *Copyright Accidents*, 96 B. U. L. REV. 1025, 1032-34 (2016) (discussing the accidents paradigm as applied to copyright).

[263] One may argue that secondary liability doctrines already impose a duty to take reasonable precautions. Thus Alfred Yen has argued that "Contributory liability is a form of fault-based liability because it mimics tort law's inquiry into a defendant's potential negligence" by inquiring "whether the defendant behaved

As long as this liability gap persists, three alternative unhappy results might follow. The first, is leaving the GenAI systemic risk problem mostly unaddressed. The second is that courts, moved by the need to mitigate the risk and the perceived unfairness of letting gatekeepers completely off the hook, might turn to the overkill of sweeping direct liability, resulting in over-blocking. The third is that courts might try to twist and bend existing law, in particular secondary liability doctrines, to regulate the systemic risk created by GenAI gatekeepers.[264] The problem here is that the basic framework of those doctrines was not designed to deal with the problem of systemic risks and benefits on the social level. Thus, manipulating various elements of the doctrines haphazardly to achieve desired results in specific cases is likely to result in various distortions. The rules are likely to be costly to administer, confusing or obscure, and therefore prone to generate much error cost. Additionally, manipulating elements of the doctrines to deal with systemic risk could bleed into other areas of their application, creating undesirable results.

How should the problem be addressed, instead? The answer is already given in the clear identification of the problem as one from the realm of accidents. What is lacking in existing law is a doctrinal mechanism for regulating the activities of GenAI gatekeepers, so an adequate balance is reached between a system design that reduces the infringement risk and the need to avoid over-blocking and excessive frustration of the system's social benefits. In other words, GenAI gatekeepers must be moved to design their systems and activities in ways that incorporates reasonable precautions, where reasonable here means taking those precautions whose social risk-reducing benefits outweigh their social functionality-impairing cost.[265] Cashing out this goal requires analyzing two questions. The first pertains to the preferable institutional standard for achieving the goal of reasonable precautions. The second is how to embed the preferred institutional standard in copyright doctrine.

B. Risk-Creating Technology: Negligence or Strict Liability?

How should an actor, engaged in a beneficial social activity, be made to take reasonable precautions that adequately reduce the social risks associated with the activity? There are (at least) three familiar institutional alternatives. The first, to be mentioned here briefly and then laid aside,

---

reasonably" in light of the knowledge he had. Yen *supra* note 240, at 215-16. There are two difficulties with this claim. First, courts have rejected the claim that contributory liability is based on a duty to take reasonable precautions. See Cox Communications, 881 F.3d 310 (ruling that proving "negligence is insufficient" for contributory liability). Second and more fundamentally, such an individualistic framing of fault as based on behavior in light of specific knowledge in particular cases leaves entirely unaddressed the crucial issue of system-wide precautions to reduce systemic risk. See also Peter S. Mennell and David Nimmer, *Legal Realism in Action: Indirect Copyright Liability's Continuing Tort Framework and Sony's De Facto Demise,* 55 UCLA L. REV. 143, (2007) (arguing that the way that the *Sony* safe haven for dual use technology has been applied in practice is consistent with general tort principles and specifically with the criterion of the unavailability of reasonable alternative design that could reduce the risk of infringement)

[264] See Wu *supra* note 107, at 408-409.

[265] See Ian Ayers & Jack M. Balkin, *The Law of AI is the Law of Risky Agents without Intentions* (unpublished), at 1 ("The law should hold [AI] risky agents to an objective standard of behavior" which means "holding the people and organizations that implement these technologies to standards of reasonable care and requirements of reasonable reduction of risk").

is regulation. The regulatory option is based on trying to anticipate and then delineate in detail—in some mix of a statutory text and administrative agency action—the specific precautions required of GenAI gatekeepers.[266] This alternative deserves, perhaps, a closer examination. It will be bracketed here, however, because of the current early and highly dynamic state of GenAI systems development. Generally, regulatory schemes, that are based on standardized specific criteria or centralized edicts rooted in expertise, tend to work well when the relevant socio-economic circumstances are relatively stable or settled. By contrast, in a dynamic, fast-changing and heterogenic environment, decentralized frameworks with multiple sources of authority and case-sensitive decisions (such as the common law) tend to have informational advantages in reaching overall better results.[267] Thus, given the dynamic and volatile stage of the technology, an open-ended standard to be applied by common law courts seems preferable at present.

The remaining two alternatives are the general legal standards of negligence and strict liability. Tort lore teaches that in a frictionless world free from transaction-costs the deterrence effect of the two competing standards will be identical.[268] Under negligence, liability is imposed only when a party fails to meet a reasonable precautions standard set externally by courts. Wanting to avoid liability the party would adhere to the standard.[269] Under strict liability, a party is liable for any harm it caused. Nevertheless a "rational" party would only take cost-effective precautions, meaning precautions whose benefit in reducing expected harm outweighs their cost.[270] If the reasonable precautions standard of negligence only requires taking cost-effective precautions, the deterrence effects of the two standards converge.

So much for the theoretical frictionless universe. By contrast, in the real world, which is ridden with transaction-costs, each of the liability standards has relative advantages and drawbacks. The main advantage of strict liability that is relevant for our context is its lower administration cost.[271] Disputes are much cheaper to resolve under a clear-cut rule that uniformly makes GenAI gatekeepers liable for any harm caused by infringing uses of their systems than under a negligence standard that necessitates the additional inquiry of failure to take reasonable precautions.[272] This ex-post simplicity also translates into ex-ante self-executing clarity under which gatekeepers, knowing they will bear the entire harm, draw on their superior information in setting their own

---

[266] See Kaminski *supra* note 261, at 1404-1406 (describing four models of administrative risk regulation).

[267] See generally Friedrich Hayek, *The Use of Knowledge in Society,* 35 Am. Econ Rev. 519 (1945) (arguing that decentralized institutions for information production are superior when applied to dynamic and complex social phenomena); Friedrich Hayek, Law, Legislation, and Liberty, vol. 1: Rules and Order (1960) chs. 3-4 (arguing for the superiority of the decentralized and adaptive common law). See also A. I. Ogus, *Law and Spontaneous Order: Hayek's Contribution to Legal Theory*, 16 J. L. & Soc'y. 393, 396-98 (1989) (discussing Hayek's views of the common law).

[268] Steven Shavell, Economic Analysis of Accident Law 181-82 (1987),

[269] *Id*., at180.

[270] *Id.*, at 179-180.

[271] Robert Cooter & Thomas Ulen, Law and Economics 359-361 (5th ed. 2008).

[272] *Id.*

standard for what precautions would be cost-effective.[273] This consideration is shored up by the fact that the behavior and system design of GenAI gatekeepers is often non-transparent to external observers. When it is costly to observe the relevant behavior of a party to adequately determine whether reasonable precautions had been taken, the higher administering and error cost of negligence in both monitoring and enforcement is exacerbated.[274]

However, negligence has its own major advantage over strict liability in implementing the relevant substantive social policies. Imposing strict liability on GenAI gatekeepers for all materialized risk generated by their systems is completely equivalent to regarding them as the direct infringers with respect to all copyright infringement on their system. As a result, the major drawback of sweeping direct liability for gatekeepers, discussed above, equally applies to a strict liability for risk standard.[275] That drawback is the mismatch between the socially desirable level of precautions and the level that will be produced by the private cost/benefit calculus of the gatekeeper.

Two factors drive this mismatch. First, recall that GenAI systems involve substantial positive externalities.[276] The activities generate substantial social benefits, both consumptive and productive, that are not internalized by the gatekeeper.[277] Nor is such complete internalization desirable. Consequently, a GenAI gatekeeper, motivated by its private calculus of the risk-reducing benefits of precautions relative to their functionality-impairing cost imposed on legitimate uses, will systematically assign lower private value to the latter relative to its full social value. The predictable result: over-blocking of legitimate uses resulting in excessive frustration of the consumptive and productive benefits of the systems.[278]

The second factor that further widens the mismatch between the private assessment of precautions and their social cost/benefit effect is non-market-value benefits. Not only do GenAI gatekeepers consider only the fraction of legitimate uses benefits they internalize, but they also measure such benefits in the currency of market value. The normative social value, however, espcially for wide-spread productive benefits, cannot be measured in market value currency. These benefits serve extra-market normative interests such as fostering social conditions for individual and collective self-determination and ensuring broad and equitable opportunities for cultural activities central to human flourishing.[279] Measured from such normative vantagepoints the social benefits cannot be reduced to maximizing subjective preferences, much less market value executed through price.[280] Thus, strict liability sets in motion a private calculus for reasonable precautions

---

[273] *Id.*

[274] Shavell *supra* note 268, at 31-32.

[275] See *supra* text accompanying notes 168-179.

[276] See *supra* text accompanying note 168.

[277] See *supra* text accompanying notes 157-164.

[278] See *supra* text accompanying notes 171-179.

[279] See *supra* text accompanying note 175.

[280] Bracha & Syed *supra* note 81, at 249.

that systematically undervalues the social benefits of GenAI systems twice: first by considering only gatekeepers' private benefits, and second by measuring these benefits in the currency of market value.

Negligence, exactly because it is an externally set standard, rather than one which is self-executing by private market decisions, is much less susceptible to these miscalculations. Under this rule, courts can better align private gatekeeper incentives to social value by correctly setting the reasonable precautions standard. A GenAI gatekeeper should only be required to take precautions whose effectivity is such that the expected social value of reducing infringement harm outweighs the social cost of impairing the system's functionality. In considering the social cost courts should be particularly mindful of any expected broad negative effects on potential legitimate uses of the system and weigh particularly heavily substantial negative effects on users' productive uses relevant for democratic interests whose value cannot be reduced to market value.

To be sure, there is no exact optimization here, but merely a roughly good enough standard. Specifically, courts should be wary of requiring as "reasonable," precautions whose implementation cost is prohibitively high, that would deeply impair the system's functionality, or that are likely to lead to a substantial amount of over-blocking of legitimate uses of copyrighted works. Courts could further adjust the standard by imposing gatekeeper liability only in cases of high degrees of literal similarity between copyrighted works and generated output, thereby pushing gatekeepers to focus on precautions that are likely to result in less over-blocking.[281] In this endeavor courts can be assisted by expertise based on developing research on "safety" measures in designing AI systems and their effects.[282] Finally, the lack of observability of the gatekeeper's behavior problem could be mitigated somewhat by requiring a measure of transparency from GenAI producers, either in a separate regulation or simply as a brand of precautions under the negligence standard.[283] In combination, these features of a well-implemented negligence standard will make it considerably superior to strict liability in capturing the social, as opposed to private, costs and benefits of GenAI systems and in setting the deterrence bar for system design choices of gatekeepers accordingly.

At the end, strict liability's failures relative to negligence outweigh its comparative advantages. The systematic miscalculation of social costs and benefits is a first-order concern that, unlike the second-order advantage of lower administration costs, goes to the heart of the underlying normative interests. Negligence is likely to do much better on this crucial substantive front, while its second-order disadvantages with respect to enforcement cost and availability of information can be mitigated.

C. Implementation

There are at least two major alternatives for implementing a GenAI negligence standard in specific legal doctrine. The first, preferable alternative, is creating a new direct negligence duty to

---

[281] Wu *supra* note 107, at 403.

[282] See e.g. See Sag, *Copyright Safety supra* note 49 at 338-343; Cooper & Grimmelmann *supra* note 117.

[283] See Sag *supra* note 49 at 340.

Gatekeepers requiring them to implement reasonable precautions. The second, considerably inferior, but perhaps somewhat more realistic, alternative is implementing the negligence standard through the existing fair use doctrine.

The straightforward way for implementing a negligence standard is simply to create a new rule that imposes such a duty on GenAI gatekeepers. There are two variants to this option. The first is for courts to create and apply a duty to actors who produce and deploy GenAI systems to take reasonable precautions to reduce the risk of copyright infringement by users of their systems. The contours of this negligence-based duty should be as described above, and the sanction for failure to meet the duty should be liability for any infringement harm causally connected to the failure.[284] Courts have both a source on which to draw in developing such a rule and the power to do so. As for a substantive source, courts could borrow and adapt from tort law the principles that apply to negligence liability imposed on gatekeepers.[285] The structural parallel of this branch of tort law— requiring gatekeeper actors to take reasonable precautions within the sphere of their control to reduce the risk of harmful acts by others—to the case of GenAI gatekeepers is apparent.[286] As evidenced by the development of secondary liability doctrines, judicial power to develop doctrine in this way exists in spades. In the past courts did not hesitate to judicially develop the norms of copyright law by incorporating and adapting familiar tort principles.[287]

An alternative implementation of a direct gatekeepers' negligence regime is via a statutory safe-haven scheme. The parallel here is section 512 of the Copyright Act that creates various immunities to actors involved in Internet communication.[288] The difference between the two regimes is that while the crux of section 512's safe-harbor immunity for hosts of internet content is a notice and take-down norm with respect to material posted by users, a GenAI safe-harbor should condition the immunity on taking reasonable precautions to prevent infringement by users.[289] While the basic norm of this mechanism is one of immunity, it could indirectly introduce negligence liability as its flipside. In other words, this construct assumes as a baseline a norm under which GenAI gatekeepers are seen as directly liable for all infringing output generated by their systems, but then carves out an immunity when reasonable precautions are taken. The net result: gatekeeper liability only in cases of failure to take reasonable precautions. Much as the 512

---

[284] See *supra* text accompanying notes 281-283. Additionally, it should be considered whether copyright's remedies, espcially statutory damages, that can often be excessive, should be limited in this context to avoid overdeterrence. Wu *supra* note 107, at 405-406.

[285] See Mark McKena, *Probabilistic Knowledge of Third Party Trademark Infringement*, 2011 STAN. TECH L. REV. 10, 11 (2011) (suggesting in a parallel context the possibility of treating cases of probabilistic knowledge of trademark infringement by borrowing from tort law and imposing liability "when the defendant failed to take precautions in the fact of a known risk of infringement"). Wu *supra* note 107, at 396-98 (arguing that trademark case law has started to shift in this direction).

[286] 17 U.S.C. §512.

[287] Peter S. Menell & David Nimmer, *Unwinding Sony*, 95 CAL. L. REV. 941, 994 (2007) ("For nearly two centuries, courts have looked to tort principles in determining the contours of copyright liability").

[288] 17 U.S.C. § 512.

[289] See 17 U.S.C. § 512(c)(C).

regime, such a statutory scheme could preserve the existing niches of gatekeeper secondary liability to address the specific scenarios of individual improper behavior covered by them.[290] Note that although the main intervention under this scheme is legislative, courts would be left to pour specific content into the reasonable precaution condition for the immunity, thereby preserving the decentralized informational advantage of a common-law standard in a volatile socio-technological context.[291]

What if Congress fails to legislate in this area and courts are hesitant to innovate by borrowing from tort law? The remaining, last-ditch, alternative for introducing negligence-based gatekeeper liability is by incorporating it into the existing fair use doctrine. The problem is that, as Ian Ayers and Jack Balkin put it, "legal doctrine generally thinks of fair use in terms of individual determinations that compare a single derivative way used in a particular way with ana asserted original."[292] To implement a negligence standard that adequately regulates the systemic risk of GenAI, fair use analysis will have to be pried out of this individualistic mold and recast into a system-wide one.

Embryonic traces of such analysis exist in two rulings of the Second Circuit with respect to the Google Books project.[293] When faced in those cases with the argument that a non-expressive background reproduction of scanned books created the risk of a security leak that could expose the full text of the copyrighted books to massive copying by others, the court refused to change its ruling that the use was fair. It relied on "the extensive security measures" the defendants "have undertaken to safeguard against the risk of a data breach."[294] Incorporating a negligence standard into fair use would involve shifting this analysis of system-wide, risk-reducing safeguards, that played a secondary role in the Google Books cases, to the center of the stage.

The analysis would involve a few elements. First, the base-line rule would be that GenAI gatekeepers are liable for all infringing output of their systems unless they manage to show that their activity is privileged under the four-factor analysis of fair use.[295] Second, the focus of such fair use analysis would be the gatekeepers' systemic actions and in particular whether it has undertaken reasonable system-design precautions, not the character of a specific infringing activity. As a result of this broad systemic risk and precautions focus, factors (2) and (3)— pertaining to the nature of the copyrighted work and amount and substantiality of the portion of

---

[290] 17 U.S.C. § 512(A)-(B).

[291] See *supra* text accompanying note 267.

[292] Ayers & Balkin *supra* note 265, at 8.

[293] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 100 (2nd Cir. 2014); Authors Guild v. Google, Inc.804 F.3d 202, 227 (2nd Cir. 2015).

[294] HathiTrust, 755 F.3d at 100. See also Authors Guild v. Google, 804 F.3d at 227 (ruling that defendant's "demonstration of the effective measures it takes to guard against piratical hacking" was "a sufficient showing of protection of its digitized copies of Plaintiffs' works" to "shift to Plaintiffs the burden of rebutting" the fair use conclusion).

[295] 17 U.S.C. §107.

the work used—would have little weight.[296] They would come into play mainly if they reflect on the adequacy of the precautions taken.[297]

The heavy-duty work will be assigned to factor (1) of the purpose and character of the use.[298] Specifically, courts would make two inquiries under this factor: a. whether the main purpose of defendant's use is a legitimate activity with substantial public benefits in the expressive realm that only carries with it an incidental risk of infringement; and b) whether defendant has implemented reasonable precautions to reduce this incidental risk in designing its system and pursuing its activity. It is, of course, into the second inquiry under this factor that the negligence analysis is incorporated, including a standard of reasonable precaution that considers the proper measure of the full social value of the system's consumptive and productive benefits.[299] Finally with respect to factor (4)—pertaining to the market effect of the use on the copyrighted work—courts should emulate the treatment of highly transformative uses.[300] Under this treatment the analysis of factor (4) is colored by that of factor (1), specifically: when a use is considered highly transformative, courts do not consider as cognizable market harm any lost licensing fees that the copyright owner could have earned for such use.[301] Similarly, courts should take into account as cognizable market harm only loss caused by risk that could be prevented by implementing reasonable precautions.

The net effect of these features would be conducting a gatekeeper negligence analysis under the mantle of the fair use doctrine. However, this doctrinal implementation is far from ideal. Much as in the context of non-expressive uses, fair use is an inapt doctrinal tool for this task.[302] Fair use is designed to be a back-end exemption doctrine sensitive to the circumstances of individual cases, not a mechanism for conducting systemic risk and reasonable precaution analysis. Using it for the latter purpose is likely to result in conceptual confusions, and prolonged skirmishes over case-specific circumstances, resulting in high cost of administering the doctrine, and high error cost. In short, gatekeeper's negligence incorporated into fair use should be an inferior last resort, to be relied on only if courts and legislature fail to develop a more adequate doctrinal vehicle.

---

[296] 17 U.S.C. §107(2)-(3). Arguably, these two factors already play a minor role in the fair uses analysis of cases involving non-expressive uses of copyrighted works.

[297] For example, if on a system-wide basis, in many cases memorized output only incorporates quantitatively and qualitatively small amounts of copyrighted works, this may affect assessment of whether the precautions taken are reasonable.

[298] 17 U.S.C. §107(1).

[299] See *supra* text accompanying notes 281-283.

[300] 17 U.S.C. §107(4).

[301] See Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 592 (1994) (refusing to consider the licensing market for parodies of the work under the fourth fair use factor); Bill Graham Archives v. Dorling Kindersley, Ltd., 448 F.3d 605, 614 (2nd Cir. 2006) (refusing to count against fair use potential lost revenue from licensing of copyrighted images for uses "transformatively different from their original expressive purpose.").

[302] See *supra* text accompanying notes 52-54.

VII. CONCLUSION

GenAI tends to excite strong reactions, as much in the cultural sphere as elsewhere. The technology's socially disruptive character fixates many observers' gaze on its considerable potential dangers or benefits. As a result, many tend to grasp at any legal or institutional means within reach to try to combat the dangers or make sure that the innovation race is not lost, and the benefits are not forgone. When GenAI operates in the cultural sphere, the means at which people tend to instinctively grasp are whatever familiar tools that happen to lie within reach in the rather disorderly shed of copyright law. The order of the day is crush GenAI dangers with sweeping copyright liability or ensure its benefits by shielding it under the umbrella of fair use.

This Article has offered a different approach. I have argued that the key for facing the challenges of GenAI in the cultural sphere is a firm conceptual understanding of how this technology operates in society. The foundation for such an understanding is having crisp concepts of the two distinct information goods relevant for GenAI: discrete expressive works and aggregate metainformation. Armed with such conceptual understanding, we can correctly analyze the dynamics of production and use of each kind of information good as well as their interface, with the aim of producing an explanatory account of the social effects of the technology. The next step is matching adequate institutional tools for dealing with the different social effects with respect to the two distinct information goods. For some of the challenges there may already be adequate tools within copyright or outside it. Other challenges may require adjustment of existing institutional tools. Yet others will require developing new institutional tools that are currently not in our arsenal.

Following this line of analysis reveals two key points about the relative adequacy of our existing institutional tools and the deficiencies in our arsenal. The first point is about copyright's sphere of institutional adequacy. Copyright—a legal regime designed as a solution to a specific cost-gaps problem in the production of discrete expressive goods—is ill-suited to deal with the policy problems that are rooted in the metainformation dynamics of GenAI. The latter problems are about cost efficiencies in production, systemic social effects rather than individualized harms, and extra-market normative concerns—all of which are out of synch with copyright's basic design. Trying to face these problems using copyright is likely to lead to poor results or even backfire. We need to turn to other institutional tools to address metainformation-related social policy problems. And herein lies the first deficiency. Many of these alternative tools do not yet exist. One benefit of the foregoing analysis is being able to see these deficiencies. The deficiencies become more visible once we remove the blinkers on our institutional imagination imposed by an instinctive turn to copyright as the remedy for all maladies in the cultural sphere. The task, once we are free of copyright-centrism, will be to develop the missing institutional tools to adequately face the unique challenges of GenAI's social effects powered by its metainformation engine.

The second point is internal to copyright. Copyright, while not the only alternative, is an appropriate institutional tool for addressing GenAI issues that relate to the production/use dynamics of discrete expressive goods. In simple terms, copyright should apply to cases where the output of GenAI systems expands access to the use value of the expression in discrete copyrighted works. Trouble starts, however, when copyright is called upon to deal with the interface between metainformation and upstream actors who produce it and individual harms inflicted on discrete

expressive works. The cause of the trouble is the contrast between copyright's atomistic character—its focus on isolated harms to specific works—and the more social or systemic effects of metainformation even within the sphere of expressive works. Existing copyright doctrine ever seeks for specific harms to specific works, whereas the risks and benefits of GenAI are systemic. Hence the second deficiency: copyright's doctrine oscillates between all or very little. GenAI gatekeepers can be subjected to sweeping liability for all infringing activity on their systems or to very limited secondary liability in cases of specific misbehavior related to individual acts of infringement. What is missing is the concept of GenAI systemic or social risks and benefits. In other words, what is missing in copyright, is an understanding of GenAI gatekeepers within an accidents paradigm, as actors engaged in a beneficial social activity that incidentally generates social risk. Developing the legal doctrine for addressing this lacuna by applying familiar institutional tools is within close reach, given initiative and will by either courts or legislatures.