

Course Name – Artificial Intelligence, Law and Justice
Professor Name – Dr. Krishna Ravi Srinivas
Department Name – Center of Excellence in Artificial Intelligence and Law
Institute Name – NALSAR University of Law
Week – 03
Lecture – 14



Artificial Intelligence, Law and Justice

Session 14
Artificial Intelligence and Copyright-Part-I

Dr. Krishna Ravi Srinivas
Adjunct Professor of Law &
Director, Center of Excellence in Artificial Intelligence and Law
NALSAR University of Law



Artificial Intelligence, Law, and Justice. In this class, we are going to see AI and Copyright. As I said in the previous class, this is session 14, the first of the four sessions on AI and Copyright. This is a very hot topic.



A Quick Overview

- Debates on AI and Copyright are raising fundamental questions because AI is also a 'creative' technology
- Traditional roles and understandings of 'Author', 'Authorship' are being questioned
- So is the applicability of copyright for training purposes in AI
- Old issues like 'fair use' are being revisited and challenged
- In this 4 sessions we quickly cover many important issues and Identify trends, dilemmas and proposed solution
- As some themes are to be discussed in different contexts for different purposes some repetition is inevitable



As I said earlier, we will look into some of the emerging issues, although we may not go into greater detail in terms of case law, the new cases, but we will look at some of the

fundamental issues that come up again and again in such debates and why the AI and copyright problem is not something that is going to be solved soon. See, fundamentally, they raise a lot of very important questions, not because AI is merely a technology. It's a creative technology that could mimic, act as, and that can simply imitate humans. When traditional copyright law or practice is based on the notion of author and authorship, of course, the whole idea of author and authorship has been questioned in the sense that people have claimed whether the idea of author can be claimed as individual and original. All those things have been questioned in different contexts. So, the point here is that copyright law envisages the author as the creator, but the author may not be the owner or may not be the one who holds the copyright, and then identifies with something the author has with the authorship.

So, authorship is something that cannot be wished away, that cannot be ignored. If there is a work of art, there is an author, and the author's contribution is also classified as authorship. But can we say in another way: copyrighted works can be used for fair use for training and educational purposes? In different contexts, there are limitations. In the famous case in India, as you all know, Xeroxing textbooks was permitted as it was considered fair use. But in the USA and other countries, that is a criminal offense, or it is an offense. You would be fined heavily for it. So, the applicability of copyright for training purposes in AI—is it applicable or not? —is a big question, more like a million-dollar question. Here the whole issue of fair use is being revisited and challenged. The fair use debate has come up again and again when digital technologies, or even Xerox, were brought in, or when people started digitizing images and then sharing them. The fundamental thing that we need to understand is that technology and copyright issues have been present for quite some time, right from the age of copyright being used to publish books.

It has evolved with each technology, and each time there was a crisis or issues that looked insurmountable or unresolved technical solutions, irrespective of them, legal solutions have been brought, acts have been amended, rules have been changed, and more importantly, when digital technology became more prominent, countries adopted something like the Digital Millennium Copyright Act so that no copyright act covers the impact of digital technology; it also recognizes the potential of digital technology for the production, consumption, and distribution of copyrighted works. So, the technical issues or the technology that poses issues have been sorted out; amicably or not, they have been sorted out, but AI poses a whole lot of new problems. So, we cannot simply extract, or you cannot simply say that there is an analogy; we can extend the same logic here. In the four sessions, there are a lot of things that we will cover, but this cannot be comprehensive on account of the complexity of the issues. Two, some of the things that we will talk about are the ones that come up again and again, so there could be some repetition here.



Four Factors Dominating the Discussion



- **Control**
 - Importance of managing and directing processes
- **Compensation**
 - Fair and adequate remuneration
- **Transparency**
 - Clarity and openness in operations
- **Legal Certainty**
 - Ensuring legal clarity and stability



Let us look at the factors that are dominating or that are the key things in the discussion. Control is important in managing and directing the processes of holding copyright works, creating copyright works, and determining who has ownership and the right to control the production, distribution, and sharing of those works. Then, compensation - if the fair use principle is given for AI training; authors will not get any compensation because it's fair use. So, if authors have to be given compensation and fair and adequate remuneration for the works being used, what sort of new principles would we need, or should we extend some of the older ideas and say that based upon the earnings, some sort of royalty should be given to users by the AI companies? We don't have clarity on it. Then the transparency and openness - see the fundamental problem: no AI system developer or AI system company is going to fully tell you that this is what they use for training, nor are they going to disclose everything about what they use for training. Some bits and pieces may be available. Even for open-source models, the information that might be available would be adequate for someone to understand, but that may not be 100%. The idea of knowing exactly what happened when the LLM was developed or when the AI tool was developed using pre-existing material that is already copyrighted is very difficult to determine. And then, the lack of transparency is a huge problem because the authors themselves won't know which of their work has been used. Then the legal clarity and legal certainty, as we said, are totally missing because there are some fundamental issues that need to be tackled.



Legislators Trying to Recalibrate the Balance of Interests



- **Strengthening the Position of Rights Holders**
 - Efforts to protect the rights of individuals and entities
- **Strengthening the Position of AI Companies**
 - Singapore's approach to support AI companies
- **Indecision**
 - Lack of clear direction in some regions
- **Conditional Control**
 - European Union's regulatory measures
- **Country-Specific Approaches**
 - China's stance on AI regulation
 - United Kingdom's policies
 - United States' regulatory framework



So how do the legislators and policy makers try to wriggle out of? One way to look at it is that it strengthens the powers or the rights of the individual holders, particularly authors and then entities. When we say entities, it could also refer to copyright societies, which are organizations that manage the copyrighted works on behalf of the authors and other copyright owners. Then one approach is to strengthen the position of AI companies by telling them that we support innovation and will ensure that they are not unduly burdened by copyright limitations and restrictions. So, expand the scope of fair use or come up with a solution so that both the creators and those who use the AI systems trained through copyrighted works are happy. The recent UK consultation on the use of copyrighted materials for the AI industry became very controversial because it tried to come up with some solutions that many authors and creators were totally unhappy about.

But not all countries have taken a clear-cut stand, and in many countries, there is an absolute lack of clarity on it. And then the European Union, as usual, is trying to find some solutions, incorporating certain things, and given the European Union's long history of trying to quickly address solutions as and when technology raises some problems, this is also evident there. So, their regulatory measures are broadly compatible and in tandem with the European Union's larger approach towards digital services, the digital economy, and the creative economy. So, their approach is very different: comprehensive in one sense, but too complex in another. The country-specific approaches to copyright matters are also very important. But precisely because each country or many countries are trying to come up with their own solutions or a set of solutions in different contexts, we are not seeing any harmonization of that globally.




Overview of AI in Artistic Creation

- **AI as a New Artistic Tool**
 - Enables creation of sounds, pictures, and texts
 - Protected by artistic freedom under Art. 13 EUCFR and Art. 10(1) ECHR
- **Risk of Copyright Claims**
 - AI use often seen as a threat to human creativity
 - Technological progress in reproduction and distribution challenges culture industry's value chain
- **AI Leading to New Copyright War**
 - Concerted action by rights holders
 - Artists rarely featured in the conflict
- **Crucial Question for Artists**
 - Whether AI's input taints its output
 - 21st century's paintbrush seen as unlawful by many lawyers and content industry




Although the binding agreement for copyright is TRIPS. So, in artistic creation, AI is really something acting like a bull in a China shop. It is a new artistic tool. It can create a whole lot of things that you never thought of. So, is AI an artistic tool protected by artistic freedom under Article 10.1 of the ECHR? That has to be really thought about. Can AI claim that, as an artist, I have some rights and my artistic freedom cannot be questioned? No, we don't know about it. We really have to look at how the Europeans try to see these things. Often, if AI can imitate human creativity, what humans have done, AI can come up with thousands and thousands of outputs that are identical and could be indistinguishable from the original output, it's a threat not only to copyright but also to creativity, and we will see this idea of technological process in reproduction and distribution challenges in the cultural industry's value chain.

See the value chain here is: I write a book, then someone publishes it for me, then someone buys it from him or her, and then someone sells it on e-commerce platforms or sells it in different versions like an e-book, digital version, bound book, print copy, or hard copy. Then there is someone who can do certain things, in the sense that someone can use the book I wrote to develop a comic strip, someone can make a film out of it, and someone else can make a drama out of it. So copyrighted works are also conceptualized as part of a larger value chain, so the problem with AI is that the technical progress in reproduction and distribution is always a problematic one, and then AI makes these things much worse, disrupting the cultural industry's value chain. It could potentially disrupt the industry because AI has huge potential to reproduce identical copies and can also mimic styles, so there are a lot of new copyright wars, and of course, there are crucial questions for the authors as well.




Artistic Use of AI

- **Artists' Use of AI**
 - Generate images and sounds
 - Manually modify or place in self-designed contexts
- **Not Limited to Prompts**
 - Fine-tune using specific material
 - Part of artistic practice protected under Art. 13 EUCFR and Art. 10(1) ECHR
- **AI as a Tool, Not an Alternative**
 - New tool for artists
 - Same acceptance problems as art created using new technical means
- **Historical Comparison**
 - Photography faced similar acceptance issues




But when authors themselves use AI, they can generate images and sounds. They can manually modify or place things in a self-designed context. So, authors can use AI. I can use AI to create a new painting and then identify the use of AI in a specific way in the painting itself, or make a claim and disclaimer that I used Canva for this, or that I used this specific Midjourney for this, and then I used my own work and AI to develop a new painting. So, the question here is that AI, as a tool, is someone's primary tool, but for some, it is not an alternative in the sense that if AI as a tool is being created, where does the question of originality come in? Where does the question of duplication and copying come in? So, it cannot be an alternative; it can be a tool, depending upon what the authors or artists want to use it for. Historically, photography also faced similar acceptance issues earlier. So, the point I am trying to raise here is that the artistic use of AI is a controversial one because not all artists are comfortable with AI. That could be a personal choice, but that choice also stems from the fundamental fact that some artists could see AI as a threat to their creativity and creative freedom. But then this has less to do with that than it is with a simple copyright issue.



Copyright Issues with AI



- **AI's Dependency on Copyrighted Material**
 - AI needs to be fed material to create new content
 - Often uses copyrighted material for training
- **Reproduction of Copyrighted Images**
 - AI must 'see' artworks to produce similar styles
 - Involves reproduction of copyrighted images
- **Potential Copyright Infringement**
 - Generated output may infringe on copyrights
 - Liability under copyright law is complex
- **Distinction Between Input and Output Phase**
 - Different legal considerations for input and output



AI's dependence on copyrighted material is well known. So, it often uses copyrighted materials for training. AI must see artwork. So copyrighted images are something that AI really has to see. And for training and other purposes that have to be reproduced with the images, photographs, paintings, and crafts, everything has to be reproduced. So, question number one is whether involving them for training purposes is a violation, or whether the act of an AI system seeing the copyrighted works to produce something is a violation, whether it is being totally reproduced or not. Question number two: the generated output might include copyrighted images as part of it; then where does the question of liability come in? If the generated output explicitly infringes on copyrighted material, liability has to be established, but the current laws are not able to handle it. Therefore, the liability here is very complex because an AI tool may use an image for training purposes and then derive something from it or can use it in such a way that its usage may not be explicitly visible or disclosed, but it could still have derived something from it. So, the question here is where the input is, where the output is, and what is the way AI deals with these things, which is a black box to many artists.



Input and Output Phases



- **Input Phase of Generative AI Models**
 - Digitised image and text material used
 - Material collected by web crawlers or selected by artists
 - Most input material protected by copyright
- **Training Process and Copyright Infringement**
 - Reproductions made during training constitute copyright infringement
 - Exception for text and data mining under CDSM Directive
 - Right holders can reserve such uses
- **Liability for Copyright Infringement**
 - US company liable for copyrighted content in learning process
- **Output Phase and Further Infringement**
- **Exceptions and Limitations**




When an input phase of AI models occurs, digitized images are used, material is collected by web crawlers, and circulated by artists. The most important input material is protected by copyright. Here the problem is: if some images are collected by web crawlers and bots that roam the web and then used, is it a copyright violation or not? Then the training process, is there a question of copyright infringement? It was held that a U.S. company was liable for copyrighted content in the learning process. But then, that is a very debatable point. So, what are the exemptions and limitations for such uses?



Legal Implications and Pastiche



- **AI Learning and Copyright Concerns**
 - AI learning is considered copyright infringement if text and data mining exceptions do not apply
 - AI output is only safe from copyright claims if input phase is covered by exceptions
- **Artistic Use of AI**
 - Artists using AI as a creative tool must have input phase protected
 - Selection of materials for fine-tuning AI is part of artistic practice
 - Technical character of fine-tuning does not change its artistic nature
- **Legal Precedents and Artistic Freedom**
 - Sampling in music is a known practice covered by artistic freedom
 - Reproductions in this context are protected by CJEU and German Federal Constitutional Court
 - Artistic freedom protects both the process and the final artwork




And then this is an artistic use of legal precedents. These are some of the things. But the more important question is, "What is 'pastiche' when you do something, when you derive something, or when you combine certain things to come up with something new? How do copyright exemptions or principles adopt or are to be used, or are they valid in the

context of AI?" This is an important question. Because artistic freedom actually protects both the process and the final artwork. So, in the recent controversy where a lot of people used AI to imitate the style of a particular artist and then came up with their own images, this became a huge issue. Are we such a process, are they violating artistic freedom, or are they imitating the artistic process, and by infringing on the rights of the author? We are not sure.



AI Copyright War Dynamics

- **Copyright as Fundamental Right**
 - Motivates new creations by rewarding old ones
- **Silicon Valley's Attitude**
 - Eric Schmidt's controversial statement at Stanford
 - Blatant disregard for IP rights
- **Rights Holders vs. Creators**
 - Economic interests often disguised as defense of the arts
 - Creators' fundamental rights overshadowed by corporate interests
 - Rights holders focus on existing works, artists create new ones
 - In a recent case a federal judge in *Thomson Reuters v. Ross Intelligence* rejected a fair use defense in an AI training context. But there is no clarity as there are many cases pending in many courts on this or similar matters

Then we need to look into the whole AI copyright war dynamics. The standard argument is that copyright is a fundamental right. It creates incentives for people to write books, compose poems, make films, create music, and do thousands of things because copyright is there to protect their interests, and it could be a good source for them in case of misappropriation or using their work in a way that the author would not approve of. So, it's a fundamental right that motivates creation, that incentivizes creation. And it is also true that some people, particularly in the industry, have a total blatant disregard for IP rights. Let us be very clear about this. The point here is that there is a whole question of right holders versus creators. Often, the right holders and the creators have totally different interests. So economic interests are often designed or disguised as a defence of the arts by the rights holders who have economic stakes, and then the real creators' fundamental rights are overshadowed by corporate interests. The rights holders focus on copyright protected existing works, whereas authors are often focused on the new ones that they can create, which they can derive, and for which they can earn money. And in a recent case, a federal judge in Thomson Reuters versus Ross Intelligence said the fair use defence in the AI training context is not available. So, the use of AI as an excuse and then saying that I have used it only in the context of fair use is not acceptable, nor is that defence available for those who develop LLMs and other AI products.



AI Copyright War and Political Economy



- The Political Economy of AI cannot be ignored in any discussion on AI and copy right. According to a recent report "In Web 1.0, we worried about lower direct rewards for artists dulling incentives. In Web 2.0, we found out that if there are indirect rewards (such as YouTube fame),this can compensate. Now we have a world where indirect rewards are also challenged because human creators will ultimately become more anonymous, at least insofar as they contribute to the synthetic output of generative AI." Lutes, Brent A. ed., Identifying the Economic Implications of Artificial Intelligence for Copyright Policy: Context and Direction for Economic Research, U.S. Copyright Office, 2025. P61
- According to **Joshua Gans**", it is demonstrated that an ex-post 'fair use' type mechanism can lead to higher expected social welfare than traditional copyright regimes" Copyright Policy Options for Generative Artificial Intelligence <https://www.nber.org/papers/w32106>
- Irrespective of criticisms of copyright as an incentive and an exclusive right, it is obvious that AI is going to create new issues for creative artists including writers. On the other hand many of them who have assigned their copyrights to the publishers or to others, may not gain much when the copyright holder licenses copyrighted materials for training and other purposes



The political economy of this cannot be ignored because, in a recent study commissioned by the US Copyright Office, it is a very interesting study, they made an interesting observation. In Web 1.0, we were worried about the loyal direct rewards for the authors. And then with Web 2.0, there were some new channels like YouTube available; you would get indirect rewards, you would get revenue, and you would get money based on the eyeballs, based on how many people see you. We have a world where indirect rewards also pose challenges because human creators ultimately will become more anonymous; AI doesn't give you the due credit, it simply absorbs your work without even telling the persons whose works it has absorbed and then used. More importantly, AI uses a lot of work for creating synthetic data and then synthetic output. Unfortunately, the authors of works that are used by AI to create synthetic output may not even be aware of that. So, what are they ending up with? They are ending up as more anonymous suppliers to synthetic output generated by generative AI. Hence, there is a need to look at the economic implications of AI for copyright policy.

So, the copyright policy, which are the traditional economic incentives argued, does not seem to be something that can really work well. So, according to Joshua Gans, he conducted a study and then said, "ex-post fair use mechanisms may be better than the traditional copyright regimes." Ex-post fair use is when my work has been used for AI training or for some other purpose by AI systems. After it has been used, I should have a say over whether it could be considered fair use. Some compensation can be given for it, or some mechanism can be created to identify the fair uses ex post, or after the use has occurred. And then try to compensate the authors or evolve a system where companies automatically give a specified amount to an author based upon the usage of their materials. For example, if a company develops an LLM in law and has taken substantial chapters from various legal professors' books, it can say that this is fair use and compensate the professor by looking at the quantity or the number of pages used, and then determine the amount to pay for fair use ex-post. So, it is also possible that some sort of licensing can be made available, but AI is going to create new issues for creative

artists and others. Many publishers already take the copyright for themselves, or the copyright is assigned to them when the publisher gets the work published or the article submitted and then approves it for publication. So in that case, what happens is that the author almost surrenders every possible right except the moral right of the author, and then the publishers become the de facto owners who can negotiate because technically, although the author is the author, the copyright has been transferred to the publisher.



AI Copyright War and Political Economy



- To give a simplified picture we have the Big Five (Face Book, Meta, Microsoft, Amazon and Google) and others in AI, and, big publishers, music labels, media empires engaged in copyright wars but the creators, artists and others who produced them, often are not in a position to engage in a protracted litigation even when they know that their creative works have been used for training or (mis) appropriated otherwise.
- On the other hand much of the battle is also on the lines of ' fair use', 'incentives for creators' , 'need for a flexible approach towards copyright to incentivise AI' , and, ' creating new schemes for rewarding and redefining limits of copyright protection'
- But if go beyond the rhetorical claims and assertions we can easily understand that the economic stakes are high.
- Given the insatiable appetite of AI for data and the potential for endless generation of synthetic data, the copyright war is also about who will control and direct the development of AI?
- Political Economy aspects and gains from valorisation through copyright have always been the elephants in the room whether we like it or not



To give a simplified picture, the big five, that is Facebook, Meta, Microsoft, Amazon, Google, and others, and big publishers, music labels, engage in a copyright war, but creators are those who are really marginalized because they have neither the resources to engage in protracted legal litigation, nor might they even know which works have been used. So much of the debate that happens under the label of fair use incentives for creators is why we value copyright as a legitimate right, and then we try to create a flexible approach so that we can incentivize AI, creating new schemes for rewarding and redefining the limits of copyright protection. The rhetoric is very much there, but then the underlying economic rationale is the cost-benefit analysis of each and every option for those who develop models versus those who have the copyright on many of these data, images, or works that are going to be used for developing and training purposes, so we need to pierce the veil of the rhetoric to find the real piece. Given the insatiable appetite, at some point AI will run mostly on synthetic data, and AI can generate its own synthetic data. So, what exactly will happen, and for how long this copyright issue will prolong, is a matter of debate.

Then, who is going to really control the subsequent direct development of AI if more and more synthetic data is going to be created by AI companies and then used? With synthetic data derived from copyrighted works, what sort of things should we be able to monitor or who will account for such use? Or will that become a broad canvas or broad umbrella of fair use, where anything you use for synthetic data creation is perfectly valid as fair use? So, the political, economic aspects and gains from valorisation to copyright has always been the elephant in the room, whether we like it or not. Some of these

questions have been raised even earlier, while others are being raised now. But they are the elephants in the room. Like the political and economic aspects of royalty, how much money should be paid to the publishers who hold the copyright, by the people who develop AI when systems are being trained using copyrighted materials? And who benefits from the valorisation of the copyrighted materials? Valorisation will happen when copyrighted materials are digitized and then sold as a digital product, or further value addition occurs through different means by artists or by the publishers themselves in various ways. So, the question here is that the political economy in the copyright war is the core of the war.



- From another vantage point it is argued “Law constructs and reinforces power in the GVCs of AI by providing the foundation for a power-enabling form of contractual governmentality and by cementing IP fences that perpetuate the uneven distribution of value across the chain. Meanwhile, law amplifies the geopolitical framing that wants transnational AI development to be shaped based on national interests.”
- Petros Terzis, Law and the political economy of AI production, *International Journal of Law and Information Technology*, Vol31, Issue 4, Winter 2023, Pages 302-330, <https://doi.org/10.1093/ijlit/eaac001>
- There are national variations of ‘fair use’ and national policies on AI that will play a key role in deciding some of the contentious issues.
- Striking a balance between innovation promotion and upholding interests of copyright holders/ authors/creators may be an ideal but how that will be negotiated and agreed upon is not clear now.

Although we can always put it in terms of rights, liabilities, and responsibilities of different stakeholders. Another vantage point, it has been well argued, is that law construction reinforces power in global value chains of AI, providing the foundation for a power-enabling form of contractual governmentality and then submitting IP fences, fences that perpetuate the uneven distribution of value across the chain. Or in other words, the value across the chain depends on who holds the IP, who has been able to license it, and who has been able to make the best of it. For authors, the value chain offers practically nothing to them because the IP fence created by the publishers and other copyright owners is what makes the difference. So, the law amplifies the geopolitical framing that seeks to shape transnational AI development based on national interest. So ultimately, the global development and deployment of AI are good and fantastic. We all talk about it.

But the ground realities in each country make a huge difference. So, the variations in the idea of fair use are too huge and vary across countries. And then national policies on AI also have some strong objectives, and some countries even talk of sovereignty in AI. And then when copyright becomes a contentious issue, will we see copyright law being tweaked to fulfil national objectives in the sense that doctrines like fair use, doctrines like fair and adequate compensation, and doctrines like reuse rights and other associated

rights, or many rights which authors enjoy on account of the TRIPS agreement and other things, will they be tweaked enough so that incentivization for innovation is much more important than paying some amount to the copyright owners as well as to authors? We really do not know because the battle is just heating up now. So how do we strike a balance, or are there any golden rules, or is there a golden means to address this? This is an important one, and how they will be negotiated upon and agreed upon is not clear, but in the three subsequent sessions, we will try to address them in some other context as well, because part of the issue here is that the copyright issue in AI is not just copyright alone. It is also a question of machine language; machine learning-related issues. So, we need to differentiate between the issues. We need to look at the issues in a broader sense of copyright wars, but also in a narrower sense of the technological use of copyrighted materials for different purposes in AI.



Next Session

- In the next session we will discuss the legal aspects in Machine Learning and Copyright



In the next session, we will discuss the legal aspects of machine learning and copyright.