THE CHANGING DIFFICULTIES POSED BY GENERATIVE AND ARTIFICIAL INTELLIGENCE (AI), AS WELL AS THE LEGAL NUANCES OF DIGITAL AND ONLINE CONTENT, PARTICULARLY WITH REGARD TO COPYRIGHT IN THE DIGITAL AGE

Dr. Debabrata Roy, Associate Professor, Faculty of Law, The ICFAI University, Tripura, Agartala-Simna Road, Kamalghat, West Tripura [1]

ABSTRACT

Copyright and intellectual property rights (IPR) structure has been reevaluated in light of the swift integration of the generative AI (Gen AI) sector into society. There are considerable ethical and legal issues concerning Gen AI's capacity to generate original content derived from data obtained from human-created materials. These issues cannot be adequately addressed by the current copyright and IPR frameworks, which are based on the concept of human authorship. This paper adopts a multi perspective methodology to investigate how Gen AI might infringe upon existing IPR regulations by duplicating or altering copyrighted works. The results highlight the lack of transparency in Gen AI platforms and the presence of legislative gaps. This study suggests creating a flexible ethical system that could work alongside a worldwide fair use policy to address these problems and guide how we develop and use Generative AI responsibly. The researchers brought together insights from various experts to understand how these new challenges and potential solutions fit into the bigger picture of how society and technology are evolving together. The study emphasizes that we need countries to work together and do more research to make sure that intellectual property laws and policies stay fair and useful as generative AI continues to shape our future.

Keywords: AI, Generative AI, Ethical Framework, Legislative loopholes, Copyright, IPR, ChatGPT.

[1] Associate Professor, Faculty of Law, The ICFAI University, Tripura, Agartala-Simna Road, Kamalghat, West Tripura

Introduction:

Generative AI, or GenAI for short, uses clever algorithms like GANs and transformer models along with machine learning and deep learning to create brand new content when you ask it to. You've probably heard of ChatGPT from OpenAI, which launched in November 2022 and became a huge hit incredibly fast, reaching 100 million users quicker than any other consumer app ever has. OpenAI trains their GPT models on enormous amounts of data, and these models can create stuff that looks and sounds so much like what humans would make that it's pretty mind-blowing. This technology has the potential to shake up how businesses work across tons of different industries. Economists think GenAI is going to have a massive impact on the world economy, predicting it could add somewhere between 2.6 and 4.4 trillion dollars every year globally. That would bump up AI's total economic impact by about 40 percent. Looking ahead, experts believe AI will take over about half of all jobs sometime between 2040 and 2060, and GenAI is actually speeding this up by about ten years compared to what people thought before. What's really amazing about generative AI is how it can create fresh content in so many different ways - whether that's writing text, making images, coding software, or even creating videos like you can do with OpenAI's Sora. Popular tools like ChatGPT, Gemini, and Bing AI all run on Large Language Models that have learned from incredible amounts of information - we're talking billions of sources including websites, research papers, books, and news articles. But here's the thing - there's quite a bit of uncertainty about how exactly these models get trained and whether it's all above board legally speaking. The companies behind these AI systems aren't being very open about their processes, which leaves people scratching their heads about transparency and who's accountable for what. This whole situation brings up some serious questions about protecting human creativity, who gets credit for original work, and who actually owns the content that gets created. As these large language models become more common and generative AI spreads everywhere, we're seeing all sorts of tricky legal issues pop up, especially when it comes to copyright laws and figuring out who can claim authorship of AI generated work.

Copyright is a special type of intellectual property that keeps original creative work safe - things like books, art, music, and other creative stuff. It gives creators the sole right to use, share, and change their work, usually for a set amount of time. Intellectual property rights cover all the legal protections for things people create, like artwork, inventions, designs, and pictures. While copyright laws deal with exact copies of pixels, text, and software, the content that AI

creates is basically built from language models that learned by soaking up and using original copyrighted material, which kind of questions whether AI-made content is really "new." Having AI as a creator that isn't human creates a big problem for our current legal system that was set up with human creators in mind. This situation brings up a bunch of tricky issues. Take, for instance, when an artist uses a Gen AI tool to make digital art - the AI creates something based on styles and features it learned from training data that includes copyrighted stuff. The person who made the original art might say they own the copyright to what the AI produced because it used their work. But the digital artist and the AI Company could argue back that what the AI made counts as a new, derivative piece of work. These kinds of problems have pushed some artists and creators to take legal action against companies like Stable Diffusion and mid journey there's this Getty case about using 12 million licensed photos without permission. On top of that, people have accused Open AI and Microsoft of software piracy when they developed Copilot, which you can now find built into MS Office.

During March 2024, interview with The Wall Street Journal, Mira Murat, who serves as Open AI's Chief Technology Officer, brought up worries about how copyrighted content might be getting misused when training AI systems. She also pointed out that Open AI hasn't been very open about how they handle their data. After her remarks came out, a lot of people started wondering if Open AI is really doing enough to protect the rights of the people who create content. This whole race to develop better AI products has shown that companies aren't always respecting intellectual property laws properly. Meta even came clean about using posts from Instagram and Facebook to train their Llama language model. When you add in how big tech giants like Alphabet can tap into huge amounts of data from Google's corner of the internet, it really makes you think about whether users are actually giving their permission and whether copyright laws are being respected when these language models get trained.

The current legal system is struggling to figure out where to draw the line when it comes to "derivative works" - basically, creative stuff that builds on existing work to make something new and original. It's also having a hard time making sense of fair use rules, which let people use copyrighted material without asking permission for things like commenting on it, critiquing it, teaching, or doing research. All of this has really shaken up copyright law. Since AI generated content doesn't have a traditional author behind it, we might see money and benefits flowing away from human creators toward whoever owns or runs the AI tech and platforms instead. This could really hurt artists, writers, and other creative people's ability to make a

living, creating bigger gaps between the haves and have-nots and making society value human creativity less. We need to find the right balance between embracing these amazing new AI technologies and making sure the people who create content can still get paid fairly for their work. That's the key challenge we're facing as we try to navigate copyright issues in this new world we're living in.

While some researchers have tried to figure out what Generative AI means for copyright and intellectual property rights, we're still missing a big-picture understanding of all the challenges that content creators and policymakers are dealing with. The thinking is that if we get different voices involved in looking at how Gen AI, copyright, and intellectual property all connect, we can uncover some really useful insights and create new ways of studying this stuff. So this research wants to take a deep dive into all the different problems that Gen AI brings up. The idea is to get people talking and researching copyright and intellectual property issues in a more forward-thinking way, moving beyond just being engaged scholars to becoming generative scholars who look at potential misuse scenarios to build better theories. This approach should help keep the focus on human values and creativity, making it easier to deal with these tricky issues in an ethical and responsible way[2].

ChatGPT and Open AI:

Generative AI is basically a type of artificial intelligence that can create brand new content on its own - things like pictures, written pieces, songs, and videos. What makes it different from regular AI systems that mostly sort things into categories or make predictions is that generative AI is built specifically to make fresh content by learning from huge amounts of data it's been trained on. These systems usually work with advanced deep learning methods, especially something called generative adversarial networks or autoregressive models, to create stuff that looks real and makes sense.

A major way AI gets better at creating new stuff is by figuring out complicated patterns from all the data it's trained on, then using those patterns to make fresh content. Take text creation, for example - you can teach an AI system using tons of written material, and it picks up on how words and sentences typically work together. After it's done learning, the system can write new

[2] Redefining boundaries in innovation and knowledge domains: Investigating the impact of generative artificial intelligence on copyright and intellectual property rights, By R. Raman, Elsevier España, S.L.U, Journal of Innovation & Knowledge 9 (2024)

text by drawing from all those patterns it discovered, usually creating something that makes sense grammatically and fits the context pretty well.

ChatGPT is basically a type of AI that creates text, and it was made by Open AI. It's built using something called transformer architecture, which is a fancy way of describing a deep learning setup that's really good at understanding and working with language. What ChatGPT does is look at what you write to it and then comes back with responses that make sense based on what you said. It uses a specific version of this transformer setup called GPT, which stands for Generative Pre-trained Transformer. Before it ever talks to users, this model gets trained on massive amounts of text so it can pick up on how people actually communicate and use language.

ChatGPT works by using a few main building blocks that all come together in a pretty clever way. It relies on things like self-attention and multi-head attention systems, plus feed-forward neural networks. Basically, these parts team up to take whatever you type in, turn it into this complex mathematical representation, run it through bunch of different layers, and then spit out a response. The whole system gets really good at what it does because it's been trained on tons and tons of text from all over the internet, and then gets extra fine-tuning for specific jobs. That's how it ends up being able to chat with you in a way that actually makes sense, stays on topic, and sounds pretty natural[3].

Concerns for Copyright and ChatGPT:

There are several copyright issues that come up with ChatGPT that people are worried about. First, there's the fact that AI models get trained using copyrighted material, then there's the problem of these systems actually creating content that might be copyrighted, plus nobody really knows who's responsible when copyright gets violated, and it's hard to figure out who actually owns or created something when AI is involved. When ChatGPT gets trained, it uses enormous amounts of text data, and a lot of that stuff is copyrighted. OpenAI has even said straight up that you can't build something as sophisticated as ChatGPT without stepping on some copyright toes - they've admitted that their GPT-4 model was trained on protected content. Because of this, the AI often spits out material that's copyrighted. So, when copyright infringement happens, figuring out who's to blame gets really messy. It could be the people

[3] Copyright and generative AI, By Laila Barqawi*, Mohammad Abdallah, Journal of Infrastructure, Policy and Development 2024, 8(8), 6253.

who made the AI software, the company hosting the platform, or whoever published the original content. Different countries like the US, South Korea, and Japan have set up their legal systems to shield publishers from AI-related copyright problems. They're trying to encourage innovation while dealing with the tricky business of proving someone actually infringed copyright, especially as the laws around AI generated content keep changing.

Additionally, well-known writers John Grisham and George R. R. Martin are among 17 authors who've filed lawsuits against OpenAI's ChatGPT over copyright violations. These writers claim that ChatGPT has shared their copyrighted material without properly acknowledging or crediting their original work. On top of that, the New York Times has also taken legal action against both OpenAI and Microsoft, arguing that ChatGPT doesn't just use copyrighted content without permission, but also creates false or misleading information that the newspaper never actually published.

Open AI has made licensing deals with some publishing companies, but that hasn't stopped writers and news outlets from taking legal action against them. Right now, Open AI is in talks with publishers for more licensing agreements, and it's turned into quite a bidding war. People are looking at the deal they made with Axel Springer and comparing it to what other publishers might get. What they're really trying to do is make ChatGPT's business legitimate and show people exactly how it creates content. By being proactive about getting these licenses, Open AI wants to tackle the legal and moral problems that come up when you use copyrighted stuff to train AI systems. They're also hoping to build better relationships and trust with the people who actually create content and the companies that publish it[4].

Jurisdictional Issues and Legal Challenges:

The emergence of AI-created content has revealed gaps in our current copyright laws, which have always assumed only humans, can be authors. One of the biggest legal headaches we're facing is figuring out who owns AI-generated works and whether they can even be copyrighted in the first place. While the Berne Convention and other international copyright rules don't explicitly say that only humans can create copyrightable works, many countries like those in the EU and US insist that there needs to be a human behind any work that gets copyright protection. What's more, copyright law has always been built with humans in mind. Just look

[4] Ibid.

at how the Berne Convention handles copyright duration - it lasts for a certain number of years after the author dies. This whole setup obviously assumes the author is mortal, which means they're thinking about human creators. Legal systems around the world are struggling to figure this out, and they're coming up with different approaches. In the US, for example, copyright law doesn't recognize non-human creators at all, which leaves AI-generated content in a legal gray area. The EU takes a more complex view, saying that for something to be original enough for copyright protection, it needs to reflect the author's own intellectual effort and creative choices. This really highlights how tricky it is to fit AI authorship into our existing copyright system[5].

Existence of IPR in the Age of Machine Learning:

It's crucial to understand that GenAI technologies depend heavily on data that already exists, including content protected by copyright, to train their systems. This dependency has led to heated discussions about whether it's legal and ethical to use copyrighted material for AI training without getting clear permission from the copyright holders. There's an ongoing debate about whether AI-generated content should be considered derivative work or completely original under current laws, which leaves creators, users, and AI developers in a confusing position. GenAI systems that can create art or write content need massive amounts of human-made data to learn from. This dependence on existing content brings up legal problems under today's copyright laws. One big issue is that the datasets used are all over the place in terms of what they contain. Some datasets have informational content that isn't protected by copyright, but many likely contain copyrighted materials. You can see this especially in datasets used for text processing, facial recognition, and image recognition, where copyrighted content shows up all the time. These practices naturally lead to legal questions about when and how copyrighted materials can be used legally.

In the US, fair use gives people some wiggle room when it comes to using copyrighted stuff for things like critiques, comments, research, or teaching. This is pretty big deal for AI companies that need tons of data to train their systems. Section 107 of the Copyright Act has been a lifesaver for major projects before - just look at what happened with Google Books. But here's the thing: nobody's really sure yet how fair use applies to all the data that goes into

[5] Redefining boundaries in innovation and knowledge domains: Investigating the impact of generative artificial intelligence on copyright and intellectual property rights, By R. Raman, Elsevier España, S.L.U, Journal of Innovation & Knowledge 9 (2024)

training AI models. We're still waiting for the courts to hash that out in cases that are happening now or coming up soon. Sure, fair use can be pretty flexible, but it's not a free-for-all. Right now, since there aren't clear rules about what's okay and what's not, both the people building AI and the folks who create content are kind of stuck in limbo, not knowing exactly where they stand legally.

The European Union has set up two different exceptions for Text and Data Mining that work in pretty different ways. The first one is really meant to help out researchers and innovators who aren't trying to make money from their work, while the second one casts a much wider net and covers all sorts of uses - as long as the people who own the rights haven't specifically said "no, you can't do this." So, with that first exception, researchers and organizations can basically dig into copyrighted material for their scientific work and innovation projects without having to ask permission from whoever owns the copyright, assuming they meet certain requirements. This has been a game-changer for universities and research centers because now they can crunch through massive amounts of data in ways that just weren't possible before, which means scientific progress moves faster and we see more innovative discoveries. The second exception is much more generous and applies to anyone - whether they're researchers or not - as long as they got their hands on the works legally. This one goes way beyond just scientific research and opens things up for a lot more people and purposes. This exception lets copyright owners back out by making their rights clear. They can do this using formats that are easy to recognize, like machine-readable tags for online content that everyone can access, metadata, or the rules listed on websites and services. On top of that, the new EU AI Act says that companies making AI systems have to create a plan to follow EU copyright rules. This means they need to use smart technology to spot and respect copyright notices. So AI developers have to make sure their systems honor copyright protections by noticing and following what rights holders say. This rule is meant to give creators the key details they need to understand how their work gets used as training data, so they can make smart decisions about keeping their rights for text and data mining. Even though Europe's rules seem straightforward on the surface, they actually bring in complications and tough requirements that make it harder to use copyrighted stuff legally. This creates a weird situation where things look clear but using these rules properly requires walking a tightrope to avoid breaking the law, which might hold back innovation by making it harder to get the important data needed for AI development and other creative projects.

International Copyright Policies Regulating AI Machines:

Worldwide, nations have been hesitant to create rules for artificial intelligence. It's pretty ironic that the developed countries where AI was first developed are being so careful about setting up clear guidelines. Instead, they're mostly using vague rules to handle copyright problems with AI-created content. This reluctance shows just how complicated and fast-changing AI technology is, plus how tricky it is to balance encouraging innovation while protecting people's intellectual property. Take Europe, the US, Canada, and the UK, for example - they've all responded to AI in different ways. The EU has really stepped up to the plate with their new AI Act. This legislation is all about making sure AI use is transparent, particularly when it comes to folks who are building these AI systems. The Act makes developers ensure they're handling copyrighted stuff the right way, which helps keep AI development on the ethical side while making sure people's intellectual property stays protected. It's pretty clear that the EU is serious about putting together rules that deal with the unique problems AI tech creates.

Canada has taken steps to tackle how artificial intelligence intersects with copyright law in much the same way. Back in 2022, they passed the Digital Charter Implementation Act, which includes a part called the Artificial Intelligence and Data Act, or AIDA for short. This creates the rules and regulations that AI has to follow in Canada. What AIDA is trying to do is build confidence in how AI gets developed and used by making sure that powerful AI systems are created and used responsibly. The law has rules built in to reduce risks around potential harm and unfair bias, so AI technology gets used in ways that match up with what's ethically and socially acceptable. But here's the thing - this law won't actually kick in until 2025, which shows Canada is taking things slow and being careful about regulating AI. Meanwhile, over in the United States, the Copyright Office has been busy looking into AI and what it means for copyright law. The U.S. hasn't put in place anything as comprehensive as what Europe has done with their laws, but they've started looking into how copyright law applies to things created with AI's help.

The UK hasn't really taken the lead when it comes to creating policies around AI and copyright issues. They haven't passed any specific laws yet that deal with how AI and copyright work together. But they're not completely ignoring it either - the Intellectual Property Office has been doing some research and asking people what they think about how AI affects intellectual property rights. They're basically trying to figure things out so they can make better policies or

change laws down the road if they need to handle all the tricky situations AI creates with copyright. On the flip side, developing countries seem to be more willing to jump in and create AI laws. Take Jordan, for example - they rolled out their Artificial Intelligence Policy back in 2020. This policy is basically their rulebook for how AI should be developed and used in their country. It sets up a regulatory body to keep an eye on AI development and usage, and it makes sure AI systems are transparent and accountable. The policy also deals with keeping data private and secure, plus it looks at how AI might affect jobs and whether people might lose work because of it[6].

Incorporating Copyright and IP Policies in the AI Era:

In today's changing world, people involved in this space are dealing with a pretty tough situation. When it comes to updating copyright and intellectual property rules for the age of AI, lawmakers and decision-makers have to tackle a complex problem: finding the right middle ground. They need to strike a balance that keeps intellectual property owners protected while still allowing AI innovation to flourish. If current copyright laws stay too strict, they might slow down the development of AI technologies that create content. But if they go too easy on the rules, human creators could get hurt and their creative work might lose its value. Getting this balance just right means really understanding how technology, law, and ethics all come together when we're talking about AI and creativity.

People are starting to realize that our current laws just aren't cutting it when it comes to dealing with AI-created content. There's this growing idea of what some folks are calling a "synthetic society" - basically a world where AI plays a huge role in creating things we used to think only humans could make. This whole concept is making us rethink what creativity and ownership really mean. As AI becomes more involved in making cultural stuff - art, writing, music, you name it - we're going to need some fresh approaches. Maybe we'll see more partnerships between humans and AI, or special licensing deals for training AI systems. The goal is to create a world where AI and human creators work together seamlessly, but we need laws that protect everyone's contributions while still encouraging innovation and respecting our cultural roots. Some ideas floating around include recognizing that both AI developers and the people using AI tools should share credit for what gets created. We might also need brand new copyright

[6] Copyright and generative AI, By Laila Barqawi*, Mohammad Abdallah, Journal of Infrastructure, Policy and Development 2024, 8(8), 6253.

categories just for AI-made works, plus international agreements so countries are all on the same page about AI copyright rules. The whole joint authorship thing between AI developers and human users makes a lot of sense when you think about it. It acknowledges that creating content with AI is really a team effort, and the line between what humans contribute and what machines contribute is getting pretty blurry these days. These efforts aim to make sure copyright and intellectual property laws keep up with how creativity and innovation are changing. When it comes to using copyrighted material to teach AI systems, there are ideas floating around like creating special licenses for machine learning or adding an AI fee to copyright laws. This would help pay back human creators who might lose money or market share because AI-generated content is taking over in creative fields. The whole point is to find a middle ground that lets AI technology grow while still recognizing what human creators bring to the table. It's all about fairly balancing everyone's rights and finding a fair way to pay creators when their work gets used to train AI. This thinking goes hand in hand with new copyright rules that support AI's place in creativity and innovation, making sure creators get properly compensated in this new world where AI plays a big role Redefining boundaries in innovation and knowledge domains: Investigating the impact of generative artificial intelligence on copyright and intellectual property rights, By R. Raman, Elsevier España, S.L.U, Journal of Innovation & Knowledge 9 (2024)[7].

Conclusion:

The rise of Gen AI is really changing how we think about copyright and intellectual property, and it's pretty obvious that our current laws just aren't cutting it when it comes to dealing with AI-influenced creativity. We need to take a hard look at our copyright and IP laws and give them a major overhaul if we want to encourage innovation while still making sure creators get the protection they deserve in this AI-powered world we're living in. This isn't just something for lawyers to figure out on their own - it's really something we all need to work on together, and it's going to take a lot of talking, teamwork, and thinking outside the box. As we're feeling our way through this whole new landscape, what we should be shooting for is building legal and ethical guidelines that work with how human creativity and AI are starting to mesh

[7] Redefining boundaries in innovation and knowledge domains: Investigating the impact of generative artificial intelligence on copyright and intellectual property rights, By R. Raman, Elsevier España, S.L.U, Journal of Innovation & Knowledge 9 (2024)

together, making sure that having AI in the mix actually adds to our culture and intellectual accomplishments instead of taking away from them[8].

[8] Ibid.