Introduction

Since 2011, the world is currently in the category of Industrial Revolution 4.0. The term "industrial revolution" was first used in 1799 by a French envoy in Berlin who said his country had entered the industrial revolution. Since then, the term has been used to refer to rapid progress in one industry at one particular place and time (Gordon, 2023, p. 118). The development of the concept of this revolution has entered its fourth, the Industrial Revolution 4.0, which began in the early 21st century and was introduced in 2011(Ivanova, 2022) -with a focus on the merging of digital, physical, and biological technologies, thus distinguishing it from previous revolutions in terms of its speed, scale, and impact on production, management, and governance systems. One of the key characteristics of this revolution is Artificial Intelligence (AI), which plays a crucial role in shaping the technologies and processes that drive this revolution, alongside other technologies such as robotics, the Internet of Things (IoT), driverless cars, and nanotechnology (Levente & Péter, 2023, p. 4).

AI combines various technologies that enable software and machines to sense, understand, act, and learn independently or with human collaboration (Chaurasia et al., 2024, p. 362). The working stages of AI go through several steps, namely data collection, data processing (cleaning, manipulation, and preparation), model training, data testing, and data optimization (Pragyna Karmakar et al., 2024, p. 79). AI uses inductive and deductive logical processes in data processing, where patterns from specific data are used to make general conclusions, while deduction uses existing premises to make specific conclusions (Collecchia & De Gobbi, 2024). In summary, AI collects, processes, and learns from data to perform tasks and make decisions.

Data is like "fuel" for AI, playing a crucial role in its development and operation. AI cannot be trained to make predictions or make relevant decisions without data. Pankaj Kumar Verma and Lakhbir Kaur's research shows that AI can only work effectively with high-quality and sufficient data.

The more diverse and accurate the data used, the better the performance of AI systems in recognizing patterns and making more accurate predictions. Data is used not only to train models but also to test their performance. This allows AI developers to evaluate the model's accuracy before it is used in a real-world context (Verma & Kaur, 2024, pp. 111–116).

However, using data as the object of AI is crucial from a legal perspective. Data has become a highly valuable asset, but not all data can be used freely. Some of it is protected by copyright, giving rise to significant legal implications in its use. Copyright is regulated by law and granted to creators of literary works, plays, music, art, cinematographic film producers, and sound recordings (Vibha S, 2022). Copyright infringement can result in severe legal sanctions, including fines and imprisonment. Some of the international copyright infringement cases that have occurred are Robin Thicke & Pharrell Williams vs. Marvin Gaye, Queen & David Bowie vs. Vanilla Ice, and Art Rogers vs. Jeff Koons (Jindal, 2024; Manar, 2024).

There is an interesting case on copyright infringement involving AI, which is the case of Thomson Reuters vs. ROSS Intelligence. This case is a significant legal dispute regarding the use of intellectual property in AI development. Thomson Reuters, the owner of Westlaw, a widely used legal research platform, alleged that ROSS Intelligence, a startup that developed an AI-based legal research platform that allows users to ask legal questions and get answers from court opinions, used copyrighted headnotes from Westlaw (a product of Thomson Reuters) to train ROSS Intelligence's AI system. Headnotes are summaries of court decisions created and organized by Westlaw as part of its legal database (Delman, 2023). The main issue in this case is whether using copyrighted data (such as headnotes) to train an AI can be considered copyright infringement.

Some previous studies that discuss copyright infringement in technology are as follows. *First*, Ujang Badru Jaman et al. examine the importance of legal protection of digital copyrighted works in the modern era influenced by technological developments and the internet. This study aims

to explain the form of legal protection of digital copyrighted works and analyze the role of the government in overcoming illegal duplication of copyrighted works in the digital era. Digital works include electronic books (e-books), music, videos, software, images, and other digital content. These works have advantages regarding ease of access, distribution, and storage. Legal protection of digital works requires a special approach that differs from physical works, especially regarding distribution and copy control. A case in point is the piracy of Andrea Hirata's "*Laskar Pelangi*" e-book by an illegal website, showing how printed works can be duplicated and disseminated without the author's permission. The government is important in providing legal protection for digital copyrighted works through strengthening regulation and supervision (Ujang Badru Jaman et al., 2021).

Second, Fenny Wulandari discusses the problem of copyright infringement in the digital era, where technological developments massively expand access and distribution of digital content. The main objective of this research is to identify the causes of copyright infringement, explore the legal system's role, and offer concrete solutions for strengthening monitoring and enforcement. Copyright infringement in the digital era is increasingly complex, mainly due to digital content's ease of distribution and reproduction. There are several types, such as music and movies, illegal software, plagiarism of literary and graphic works, and unauthorized distribution of photography and cinematography. In addition, copyright infringement negatively impacts creators and holders, including financial losses, loss of control, and weakening of creative industries(Wulandari, 2024).

Third, Lidwina Dope Nyadjroh Gabsa explores the phenomenon of copyright infringement in the rapidly growing era of information and communication technology (ICT). This research aims to provide insight to copyright owners and related stakeholders regarding the concept of copyright protection, copyright infringement in the context of ICT, the basics of copyright infringement, and copyright infringement in literary works, art,

and neighboring rights. According to him, copyright protects creators of works that are original and embodied in tangible form, including books, music, visual arts, computer programs, and audio-visual broadcasts. Copyright infringement in the ICT context includes music and film piracy, unauthorized duplication, illegal downloading, unauthorized distribution, and unlicensed use of images, videos, or music. In principle, copyright infringement occurs when a third party uses a work without the copyright holder's permission, making it subject to civil or criminal prosecution, depending on the applicable legal jurisdiction (Gabsa, 2024).

Fourth, Ziyan Yan and Zhao Hao examine the issue of copyright infringement arising from using artificial intelligence (AI) technology in creating paintings. Since 2022, AI has developed rapidly in the art field, enabling innovative and stylish artwork creation. However, creating AI artwork often involves using materials from other people's artworks, potentially infringing copyright. This research identifies three main types of copyright infringement by AI paintings: reproduction right infringement, adaptation right infringement, and attribution right infringement. This research shows that there are difficulties in identifying and proving copyright infringement by AI due to the difficulty in identifying the act of infringement, the difficulty in determining the liable subject, and the difficulty in the lawsuit process (Yan & Hao, 2024).

Fifth, Lijie Ai addresses the legal issues of copyright protection on the internet that arise with the development of network technology. The main focus of this research is to identify challenges, elements in need of legal protection, and measures to improve copyright legal protection in the digital environment. Some of the challenges of copyright protection on the Internet are easy and rapid dissemination, high cost of monitoring and enforcement, and non-exclusive nature. The elements of copyright legal protection on the Internet include protection of network transmission rights, technical measures, and rights management information. Improving the legal protection of copyright can be done through strengthening laws and

regulations, strengthening judicial protection, and optimizing administrative protection (Ai, 2015).

Table 1. Comparison of Previous Research

No.	Researcher	Research Focus	Research Objectives
1	Ujang Badru Jaman et al. (2021)	Legal protection of digital copyrighted works in the modern era influenced by technology and the internet.	Explains the form of legal protection of digital copyrighted works as well as the role of the government in monitoring and strengthening regulations.
2	Fenny Wulandari (2024)	Problems of copyright infringement in the digital age, focusing on access and distribution of digital content.	Identify the causes of violations, explore the legal system's role, and offer supervision and enforcement solutions.
3	Lidwina Dope Nyadjroh Gabsa (2024)	Copyright infringement in the context of Information and Communication Technology (ICT) and the protection of copyrighted works.	Provide insight to copyright owners and stakeholders regarding the concept of copyright protection in the context of ICT.
4	Ziyan Yan and Zhao Hao (2024)	The issue of copyright infringement is due to AI's use in creating paintings.	Identify types of copyright infringement in creating AI paintings and examine legal challenges in the lawsuit process.
5	Lijie Ai (2015)	The legal issue of copyright protection on	Identify challenges, elements of legal

		the internet includes challenges and elements that require legal protection.	protection, and measures to improve copyright legal protection on the internet.

Based on the comparison of previous research, it seems that no research specifically discusses the legal implications of using copyrighted data in Artificial Intelligence (AI) training. Most of the previous research focuses on copyright protection in general in the digital context, such as the protection of e-books, music, videos, software, and other digital content. While some studies address copyright infringement in the digital age, the focus is more on infringement of illegal distribution, piracy, and unauthorized copying on digital platforms or internet networks. In addition, previous studies have mostly explored the role of the government in strengthening regulations and supervision in preventing copyright infringement. No research has examined the use of copyrighted data as AI training material and its impact on the regulation of copyright law policy in the AI era. The approach used in previous research is also more normative and descriptive, with no concrete case study approach related to this legal issue.

Based on this description, this research aims to identify the legal implications of using copyrighted data in AI training. This research has the advantage of specifically analyzing the Thomson Reuters vs. ROSS Intelligence case, which deals directly with using copyrighted data (headnotes from Westlaw) in AI training. The case raises an urgent and contextual legal question: Can using copyrighted data to train AI be classified as copyright infringement? No previous research has discussed this case in depth, let alone in relation to the legal setting in Indonesia and its comparison with other jurisdictions, such as the United States and the European Union. Therefore, this research fills a gap in the legal literature (research gap) related to legal arrangements and policies governing the use of copyrighted data in the AI development process, especially by emphasizing the perspective of the legal system in Indonesia and other

countries' legal systems. This research is relevant in terms of academic contribution and provides more specific policy recommendations to policymakers, AI developers, and copyright holders.