

A Dual-System Approach for Integrating AI Inventorship into Law

Xin Shao

Share:



©2024. Published in *Landslide*, Vol. 17, No. 1, September/October 2024, by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association or the copyright holder.

The rapid development of artificial intelligence (AI) has led to a growing number of AI-driven activities that challenge traditional notions of inventorship, authorship, and intellectual property rights. Among these challenges, one contentious issue is whether AI can be recognized as an inventor.

The current legal landscape faces a dilemma regarding the recognition of AI as an inventor. While AI plays an increasingly crucial role in various fields, such as drug development, ¹ the lack of definite legal recognition of AI's contributions raises significant regulatory and compliance concerns. ² The growing capability of AI to "invent" suggests that failing to recognize its role as an inventor could possibly limit the full exploration of its potential and hinder technological progress. The diminishing human involvement in AI-driven innovations will likely create a legal vacuum in rights assignment and liability, leaving the legal system unable to address the ensuing legal and ethical implications. ³ Specifically, the absence of well-defined liability standards for AI-generated creations and actions creates a legal void that could lead to the misuse of AI technology without proper accountability. One example is in situations where it is unclear to what extent responsibility should be attributed to humans, AI systems, or the companies that develop and deploy them, such as in cases of accidents involving autonomous vehicles or the generation of harmful content by AI. This ambiguity necessitates guidance to determine when and how to pierce the shield or "aegis" of AI. ⁴

However, a recognition of AI-driven activities, particularly in the context of inventorship, may present significant legal challenges. The traditional legal system aims to incentivize human

contribution in inventive activities, and this objective justifies distinguishing between human and AI contributions. (5) While making such a distinction offers the advantage of potentially ensuring equal consideration of inventive outcomes in terms of patentability, (6) it may not help assign legal rights and liabilities effectively. But a more inclusive approach, such as recognizing AI systems as inventors, (7) could introduce significant legal and ethical challenges, potentially risking the stability and consistency of the existing legal framework.

“To elude a storm, you can either sail into it or around it, but you must never await its coming.” (8) In the face of rapid and transformative technological changes, it may be desirable for the legal system to take proactive measures rather than waiting for the arrival of, for example, superintelligent models that will eclipse human inventors in some areas. (9) To address these challenges in a proactive way without risking the stability of the established legal system, this article proposes a dual-system approach that integrates AI-driven activities, including AI-driven inventive activities, into the legal framework, providing a new perspective regarding the legal recognition of AI inventorship.

Inspiration from Software Deployment Strategies and AI Alignment Ideas

The proposed dynamic dual-system approach draws inspiration from software deployment strategies and alignment concepts in AI model development.

Blue/Green Deployment

Blue/green deployment is a strategy aimed at reducing downtime and risk during software updates.

(10) It involves two identical environments: blue (current) and green (new). Initially, user interactions are handled by the blue environment, while the green environment is updated and tested in parallel. Once the green environment is fully tested and confirmed operational, the user traffic is shifted from blue to green. The major advantage of this method is its fail-safe nature: The blue environment is kept intact. Modifications performed on the green environment do not influence the blue environment. If issues occur post-switch, the system can quickly revert to the blue environment, minimizing the impact on the integrity and stability of the original model.

Canary Deployment

Canary deployment is an alternative strategy in software deployment that manages the rollout of new versions by gradually shifting traffic from the old version to the new one. (11) Unlike blue/green deployment, where the switch is instantaneous and affects all users at once, canary deployment aims to minimize risk by targeting a small group of users initially. This method involves

deploying the new version of the software (the “canary”) to a limited subset of servers, which then serves a small percentage of users. Traffic is incrementally shifted toward the new version, allowing for detailed monitoring and adjustments based on performance and user feedback. This incremental rollout helps minimize impact on the original system.

AI Alignment

AI alignment is a concept that has gained significant attention in the development and deployment of AI systems. It refers to the idea of ensuring that an AI system’s goals, decisions, and actions are consistent with human values, intentions, and ethical principles. ¹² The primary objective of AI alignment is to guarantee that AI systems behave in ways that are beneficial to humans and do not cause harm or act in ways that are contrary to human interests.

Superalignment

Superalignment goes beyond basic alignment. It aims to ensure that advanced AI systems, potentially surpassing human intelligence, stay aligned with human values across various domains.

¹³

The Dual-System Design

Guided by these concepts, this article proposes a dual-system design that introduces two complementary systems: the traditional legal system for human-driven activities and a new AI-specific legal system for AI-driven activities. Under this dual-system approach, the established legal system continues to function as it traditionally has, designed to address all legal matters except for the unique challenges presented by AI-driven activities. The AI-specific system, on the other hand, will introduce rules that are distinct yet aligned with traditional legal standards, upholding fairness and human values. The alignment mechanism will serve as a key component, establishing a connection between the two systems by precisely mapping legal standards from the traditional system to the AI context. This may ensure that the new rules are adapted and aligned to meet the unique challenges posed by AI while maintaining a clear and coherent set of principles that integrate seamlessly with the existing legal structure.

To illustrate the application of the dual-system approach, consider an AI system that invents a new drug. Under the proposed framework, a pharmaceutical company could file a patent application listing the AI as the inventor. The AI-specific system would address such matters by applying principles consistent with traditional legal norms. This may include defining an AI entity (in terms of its architecture and set of parameters) to be aligned with the legal concept of a person and defining

an ordinarily skilled AI (in terms of a predetermined range of quantifiable metrics related to its architecture and set of parameters) to be aligned with a person having ordinary skill in the art. This ensures consistent legal interpretation during patent prosecution and other legal proceedings within the AI-specific system.

The bifurcated design of the dual-system approach facilitates the smooth incorporation of AI into the legal framework, giving the legal system the latitude to handle legal challenges without jeopardizing the established legal system. This dual-system approach may provide the following advantages, analogous to the strategic benefits observed in the above software deployment methods

14 :

- **Preservation of the established legal system:** Similar to how blue/green deployment maintains operational continuity without compromising current functionalities, this approach protects the integrity of the existing legal framework. It ensures that the incorporation of AI-driven activities does not destabilize well-established legal principles and practices.
- **Flexibility and adaptability:** Reflecting the adaptable nature of the software deployment strategies, this approach allows for fine-tuning and adjustments within the foundational legal framework. This flexibility enables the tailored integration of AI-specific considerations, ensuring that the legal system remains relevant and responsive to technological advancements.
- **Dynamic evolution:** This approach enables a legal framework that evolves alongside technological advancements, incorporating continuous improvement principles similar to those in software updates. It allows the legal system to adapt to emerging AI-driven challenges without compromising its foundational values while facilitating a two-way dynamic interaction between legal rules and scientific advancements.
- **Scalability:** The proposed design introduces scalability, echoing the strategic benefits of the two software deployment strategies. It allows for a focused start with AI inventorship in specific sectors like drug development before broadening to encompass AI inventorship across more extensive technological domains. Subsequently, it can extend to cover AI authorship and even AI-driven infringement activities. This scalability ensures that the legal framework can evolve and expand in response to the expanding scope of AI-driven activities, maintaining its effectiveness and relevance as new challenges arise.

Achieving Inter-System and Intra-System Coherence

At first glance, the concept of a “dynamic dual system with alignment mechanism” might appear familiar within the legal domain, where adapting to emerging challenges is standard practice. However, the rationale for developing a dual-system framework, with one system aligned to the other instead of merely formulating individual principles for various AI scenarios, is rooted in the unique nature of AI. Unlike conventional technologies, AI embodies a form of intelligence.

The legal implications of AI technology may be fundamentally different from previous technologies due to the remarkable similarity between the capabilities and performance of AI systems and human cognition. While there is no single universally accepted definition of “artificial intelligence,” ¹⁵ several prominent definitions emphasize the ability of AI systems to produce outputs or actions that are indistinguishable from those of humans. The Turing test, for instance, defines AI as a computer system capable of engaging in communication that is indistinguishable from human communication. ¹⁶ This indistinguishability based on the system’s output suggests that AI systems can generate responses or behaviors that are functionally equivalent to those of humans, to the point where they may not be reliably differentiated.

The inherent coherence in AI-driven activities, based on fundamental similarities in their underlying architectures and functionalities, suggests that a unified set of principles is crucial. These principles should not only align with established legal norms but also maintain coherence to ensure that AI-specific regulations are consistent across different legal domains. For example, if rules are derived for adjudicating AI copyright cases, these rules should maintain consistency with other rules derived for governing liability in autonomous vehicle accidents, considering that these two types of cases share the same important facts of computer vision.

Therefore, the dual-system approach strives for two-dimensional coherence: inter-system and intra-system. Inter-system coherence means that AI-driven activities are governed by principles harmonious with those for human activities. Accordingly, intra-system coherence demands consistent and cohesive legal principles for AI-driven activities, from copyright infringement to antitrust issues. Recognizing AI as a distinct form of intelligence means that the legal system must be consistent with fundamental laws and reflect an internal coherence suited to AI-driven contexts. The dual-system approach is promising in achieving both inter-system and intra-system coherence.

Adjudicating AI-Driven Activities beyond Human Comprehension

The dual-system design aims not only to develop aligned legal rules for AI-driven activities within human understanding but also to provide a framework for those beyond it. As AI systems continue to advance and potentially surpass human cognitive abilities, the legal system will face the challenge

of adjudicating AI-driven activities that may be difficult for humans to fully comprehend. How will such actions be adjudicated while ensuring alignment with human values?

The alignment mechanism in the dual-system design may provide a solution. To achieve superalignment, the idea of “weak-to-strong generalization” has been proposed to iteratively align AI with human values. ¹⁷ This weak-to-strong generalization can be illustrated by the following example: Initially, humans align weaker AI models with fundamental human values and ethics. Once these weaker models are reliably aligned, they are used as a benchmark to align more advanced AI models. This sequential approach ensures that even as AI systems grow in complexity and capability, they maintain an alignment with the foundational human values instilled in the initial models. Therefore, the process involves using human-aligned AI to subsequently align more sophisticated AI systems, creating a chain of alignment from weak to strong AI models. In other words, initially, this framework establishes legal principles that tackle simple, well-defined AI-related legal issues, thus providing “weak labels.” As the complexity of AI-driven activities increases, the system evolves, utilizing the aligned AI to generate “strong labels,” thereby broadening the legal principles to encompass a wider array of intricate issues.

To further illustrate the application of the weak-to-strong generalization method for achieving superalignment in the dual-system, consider the challenge of adjudicating whether a super AI model’s claimed invention is “obvious” in view of prior super AI models’ inventions. In other words, the legal system is tasked with differentiating between an “obvious” invention and a “nonobvious” invention from the perspective of AI systems, while both inventions are developed by advanced AI systems and nonobvious from the perspective of humans. In this scenario, the legal standard of “obviousness” within the AI-specific system needs to be aligned through weak-to-strong generalization. This alignment ensures that even when AI becomes significantly superior to humans in terms of cognitive abilities, ¹⁸ the dual system can still use the aligned obviousness standard to adjudicate whether a patent application listing an AI as an inventor is patentable in view of prior AI inventions. ¹⁹ The process of weak-to-strong generalization in this context could involve the following steps:

- ¹ Judges or lawyers initially provide labels of “obviousness” or “nonobviousness” to a dataset of case law.
- ² The labeled dataset is used to train a weak machine learning model.
- ³ The trained weak model generates AI-generated labels for an expanded dataset. At this step, the AI-generated labels start to differ from human-provided labels. That is, what is “obvious” in the eyes of an aligned AI is different from what is obvious in the eyes of humans.

- 4 The expanded dataset is used to train a stronger model.
- 5 The process of refining labels and models is repeated iteratively, resulting in a stronger machine learning model that is aligned with human values, such as the notion that an invention needs to be nonobvious to satisfy the requirements of 35 U.S.C. § 103.
- 6 The ongoing superalignment research can contribute to the development of aligned legal concepts and provide metrics to measure the alignments.

As AI inventors become more sophisticated and surpass human inventive capacities, the trained stronger AI model can be employed to generate labels for even more advanced AI models. In this manner, the dual-system design may be able to adjudicate—based on well-defined human values—AI-driven activities beyond human comprehension. In short, the dual-system design aims to ensure that when the players are AI, the judges are humans; and even when the judges are AI, the judges of the AI judges are humans.

The Future

Two divergent possibilities emerge regarding the future scope and impact of AI-driven activities. One possibility involves AI's influence being constrained by factors such as data scarcity or language model limitations, restricting AI to specialized functions within the proposed dual-system legal framework. The other possibility sees AI attaining remarkable sophistication that transcends human-level capabilities across various domains. In this latter scenario, AI systems could potentially drive a wide array of innovative activities, from conception and design to production and distribution, rendering AI as the primary instigator deserving of legal rights and responsibilities. Regardless of which possibility unfolds, the dual-system approach provides a framework to adjudicate AI-driven activities in alignment with human values, without disrupting the existing legal landscape.

Endnotes

1. See Catrin Hasselgren & Tudor I. Oprea, *Artificial Intelligence for Drug Discovery: Are We There Yet?*, 64 ANN. REV. PHARMACOL. & TOXICOL. 527 (2024) (underscoring the transformative impact of AI on drug discovery; detailing how AI applications in generative chemistry, machine learning, and multiproperty optimization are propelling several compounds into clinical trials; and highlighting the significant benefits of AI in reducing development costs, minimizing animal testing, and accelerating the creation of effective treatments).
2. See Jens-Ulrich Stegmann et al., *Trustworthy AI for Safe Medicines*, 22 NATURE REV. DRUG DISCOVERY 855 (2023).
3. See Ryan E. Long, *Artificial Intelligence Liability: The Rules Are Changing*, LSE BUS. REV. (Aug. 16, 2021), <https://blogs.lse.ac.uk/businessreview/2021/08/16/artificial-intelligence-liability-the-rules-are-changing/> (discussing the evolving landscape of AI liability, pointing out that the law has been

relatively slow to regulate AI, and highlighting the challenges and changes in legal norms as AI technologies become more integrated into various sectors).

4. See Timothy T. Hsieh, *The Aegis of AI* (May 11, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4824463.
5. See John Villasenor, *Reconceptualizing Conception: Making Room for Artificial Intelligence Inventions*, 39 SANTA CLARA HIGH TECH. L.J. 197 (2023), <https://digitalcommons.law.scu.edu/chtlj/vol39/iss2/2/>.
6. See *id.* at 230 (“Because AI systems possess intelligence, an attribute that patent law has historically considered only in relation to humans, complexities are inevitable when AI systems produce inventions. . . . [I]nventions that would be patentable if they were not created using artificial intelligence (AI) should not become unpatentable by virtue of the use of AI in their creation.”).
7. See, e.g., Russ Pearlman, *Recognizing Artificial Intelligence (AI) as Authors and Inventors under U.S. Intellectual Property Law*, 24 RICH. J.L. & TECH., no. 2, 2018, <https://jolt.richmond.edu/recognizing-artificial-intelligence-ai-as-authors-and-inventors-under-u-s-intellectual-property-law/>.
8. See *House of the Dragon: The Rogue Prince* (HBO television broadcast Aug. 28, 2022) (quoting the master of ships’ advice to the king).
9. See NICK BOSTROM, *SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES* (2016).
10. See AMAZON WEB SERVS., INC., *BLUE/GREEN DEPLOYMENTS ON AWS* (2024), <https://docs.aws.amazon.com/pdfs/whitepapers/latest/blue-green-deployments/blue-green-deployments.pdf>.
11. See Danilo Sato, *Canary Release*, MARTIN FOWLER (June 25, 2014), <https://martinfowler.com/bliki/CanaryRelease.html>.
12. See Eliezer Yudkowsky, *The AI Alignment Problem: Why It’s Hard, and Where to Start*, MACH. INTEL. RSCH. INST. (May 5, 2016), <https://intelligence.org/stanford-talk/>.
13. See Collin Burns et al., *Weak-to-Strong Generalization: Eliciting Strong Capabilities with Weak Supervision*, arXiv:2312.09390 [cs.CL] (2023), <https://arxiv.org/abs/2312.09390>.
14. See AMAZON WEB SERVS., *supra* note 10.
15. See Jonas Schuett, *A Legal Definition of AI*, arXiv:1909.01095v1 [cs.CY] (2019), <https://arxiv.org/abs/1909.01095v1>.
16. See *id.*
17. See Burns et al., *supra* note 13.
18. See BOSTROM, *supra* note 9 (examining the potential future development of superintelligent AI, capable of performing general cognitive tasks at a level of intelligence that parallels how an

average human surpasses a beetle or a worm in cognitive abilities).

19. *See id.* (discussing the challenges in designing control mechanisms that could effectively manage a superintelligent AI; emphasizing the importance of getting it right on the first attempt, as a superintelligent AI could potentially outmaneuver human attempts to control it afterward; and highlighting the importance of taking a proactive approach).

ENTITY:

SECTION OF INTELLECTUAL PROPERTY LAW

TOPIC:

TECHNOLOGY, INTELLECTUAL PROPERTY

The material in all ABA publications is copyrighted and may be reprinted by permission only. Request reprint permission [here](#).

Author

Xin Shao

Michael Best & Friedrich LLP

Xin Shao is an associate at Michael Best & Friedrich LLP. He prepares and prosecutes patents in areas including artificial intelligence, machine learning, and cloud computing.