Human, All Too Human: A Philosophical Investigation on Intellectual Property Rights for AI-Based Creativity

Human, All Too Human:
A Philosophical Investigation on Intellectual Property Rights for AI-based Creativity

Antonino Rotolo*

University of Bologna,
Alma Mater Research Institute for Human-Centered Artificial Intelligence
and Department of Legal Studies,
Bologna, Italy

antonino.rotolo@unibo.it

Man is a machine […]
in the whole universe there is
but a single Substance differently modified.
(Julien Offray de La Mettrie, Man a Machine)

If you steal from one book you are condemned as a plagiarist,
but if you steal from ten books you are considered a scholar,
and if you steal from thirty or forty books,
a distinguished scholar.
(Amos Oz, A Tale of Love and Darkness)

Abstract. As generative AI continues to evolve, questions regarding its role in creativity and the legal framework surrounding intellectual property rights are becoming increasingly pertinent. This paper examines the implications of AI-generated inventions within the context of the TRIPS agreement, especially in light of the DABUS case, which raised fundamental debates about AI's capacity for innovation and the ownership of created works. We explore the distinctions between various forms of creativity, as proposed by Margaret Boden, and investigate the philosophical and legal complications associated with AI's outputs. Through this framework, we assess whether AI-generated inventions can genuinely be considered novel and non-obvious, and provide critical criteria for patentability. We introduce the concept of novelty as a supervenient property on the data from which AI learns, suggesting that true innovation may emerge only when AI is exposed to diverse and extensive datasets. Furthermore, we address the evolving nature of ownership in cases where iteratively developed AI systems contribute to creation, challenging conventional ideas about inventorship. By highlighting the complexities of AI's interaction with creativity and IPR, this paper aims to contribute to the ongoing discourse on fostering a responsible and ethically grounded innovation ecosystem in the age of AI. Ultimately, we advocate for a re-evaluation of existing legal

* This chapter elaborates on a presentation I gave on these topics at the workshop *Mapping the future: AI, IP law & Other Creatures*, held on September 26-27, 2022, at the Jagiellonian University in Kraków. My sincere gratitude goes to the workshop organisers and participants for their many fruitful comments.

frameworks to accommodate the transformative impacts of AI-driven creativity, ensuring appropriate recognition and protection for human and machine-generated contributions.

Keywords: Legal Philosophy; Generative AI; Intellectual Property Rights; Creativity; Novelty; Inventiveness

1. Introduction

Generative AI is perhaps poised to surpass human expertise in several fundamental activities, challenging our current understanding of creativity and its legal protection. As these systems develop advanced capabilities, critical questions arise: Could AI become the sole entity capable of monitoring and detecting its own regulatory breaches in the domain of IPR? Do we need to rethink the very concept of intellectual property? Generative AI acts as a transformative force, revolutionizing tasks traditionally linked to human intelligence, such as natural language understanding, decision-making, and experiential learning.

This, however, was discussed long before the recent impetus provided by the development of popular LLMs. Indeed, the famous case of Stephen Thaler and his Artificial Intelligence (AI) system, DABUS, has ignited in 2018 a global debate regarding the patentability of AI-generated inventions, and, more generally, the Intellectual Property (IP) protection of AI-based intellectual creations. As is well known, Thaler, president and CEO of *Imagination Engines*, opens the discussion by claiming that DABUS independently invented a new type of food container and a flashing light for emergencies. His subsequent attempt to obtain patents for these inventions, listing DABUS as the inventor, was met with consistent rejection by patent offices worldwide: inventions should be made only by humans.

This chapter investigates some philosophical and legal-theoretical complexities surrounding this issue. We explore the fundamental question: under what philosophical assumptions can creations birthed from AI (in particular, generative AI) be considered, e.g., patentable inventions (but also copyrighted creations)? This inquiry necessitates examining the very nature of the AI-based creative process and its alignment with the established criteria for IP legal protection.

Furthermore, the chapter discusses some basic philosophical implications and premises for the ownership of IP rights (IPR) in the context of AI-generated inventions and creations. Who deserves the legal recognition and associated benefits—the programmer who created the AI, the company that owns the AI, the AI itself, or who else? Examining this question necessitates a critical review of current legal theoretical issues and their potential shortcomings in addressing the unique challenges posed by AI-driven innovation.

This chapter aims to not only analyse the legal and philosophical complexities but also briefly explore the potential consequences of attributing inventive capacity to AI. What are the potential ramifications for human inventors and the future of innovation as a whole? Finally, the chapter acknowledges additional layers of complexity that arise when considering AI's evolving capabilities and potential for independent thought.

By examining these questions, this chapter aims to contribute to the ongoing discourse on AI, creativity, and IP, paving the way for a framework that fosters responsible and ethically grounded innovation in the age of AI.

The layout of the contribution is as follow. Section 2 evaluates DABUS case in the light of the principles of the Trade-Related Aspects of Intellectual Property Rights (TRIPS) of 1994. Section 3 discusses the potential legal-theoretical impact of AI-based creativity. Section 4 outlines Margaret Boden's theory of creativity as applied to AI-based inventions and creations. Section 5 highlights benefits and limits of the use of some variations of Turing test to check the novelty of

AI-based inventions. Section 6 discusses what type of creativity can be acknowledged to AI and under which philosophical assumptions this is possible. Section 7 further investigates the issue of AI creativity by focusing on the ontology of the property "new". Sections 8 focuses on some consequences and further complexities of the analysis developed in the previous sections. A summary ends the chapter.

2. An Example: Patentable AI-based Inventions

The legal framework for patentability is established by international agreements like the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) of 1994. This agreement defines an invention as a creation made according to the following four key criteria for patentability (for a discussion on AI-based inventions, see George and Walsh 2022; Spedicato 2019; Picht and Thouvenin 2023):

1) *An invention must be made by one or more inventors*. This includes products, processes, or methods in almost all fields of technology.
2) *Novelty*. The invention does not already exist.
3) *Inventive step or non-obvious*. The invention would not be obvious to a "person skilled in the art" who has "common general knowledge" in that field.
4) *Capable of industrial application or utility*. The invention can be made or used in industry, does as is claimed and/or has economic significance.

However, the question of attributing inventorship to AI systems introduces unique challenges. While TRIPS establishes human inventors as the norm, it does not explicitly address the patentability of AI-generated inventions. This ambiguity creates a grey area, particularly regarding the criteria of novelty and inventive step.

To illustrate the techniques used to dissect originality, consider the example of an AI system that creates a new type of biodegradable packaging. When assessing its patentability, we could analyse the underlying algorithms and data the AI was trained on. Techniques such as novelty searches in existing databases can help determine whether the AI-generated packaging indeed represents a significant advancement over existing solutions. This careful dissection is crucial to distinguishing between mere combinations of existing ideas and genuinely novel inventions.

As highlighted in points 2) and 3) from the TRIPS criteria become crucial in the context of AI-based inventions:

Novelty - While AI can undoubtedly generate original outputs, the question arises: can these outputs be considered truly "new" if they are derived from the vast amount of data the AI has been trained on? Determining whether an AI-generated invention is truly novel becomes a complex task, based on appropriate theoretical assumptions and potentially requiring sophisticated techniques to dissect the originality of the output from the underlying data. An example would be an AI system trained on thousands of existing recipes. If it generates a new dish by combining familiar ingredients, we must analyse whether this dish is genuinely novel or merely a recombination of known elements. Evaluating such cases requires an understanding of both the data used and the creative process involved, which enriches our comprehension of originality.

Inventive Step - This criterion demands that the invention be non-obvious and not readily apparent to a "person skilled in the art" with relevant knowledge. Evaluating the inventive step of an AI-generated invention presents another challenge. Can an invention be considered non-obvious if it originates from a non-human entity operating outside the usual realm of human intuition and

reasoning? Determining the inventive step of these creations necessitates an exploration of what constitutes "inventiveness" in the context of AI and how it might differ from human ingenuity.

Therefore, the established patent standard, as outlined in the TRIPS agreement, presents significant obstacles in readily applying traditional patentability criteria to AI-generated creations.

3. IPR, Creativity, and the Importance of Being Quantified

In the context of creative industries, consider AI's impact on music production and copyright law. Generative AI systems are increasingly being used to compose music, blending various genres and styles, often producing tracks that are not only original but also commercially viable. For instance, consider an AI system that analyses thousands of hits across different decades to create a new song that captures the essence of modern pop, rock, and jazz. Does this creativity deserve full IPR protection? This reframing of the problem challenges the current IPR system. If most AI creativity falls under the realm of sophisticated combination of data, fuelled by vast data sets, how do we credit innovation? Perhaps a new category within IPR is needed, one that acknowledges the transformative power of data alongside human ingenuity. The data itself becomes a decisive collaborator, shaping the creative process and potentially deserving a share of the credit. This requires a new philosophical framework for IPR that recognises the unique interplay between human and machine. It is not just about protecting inventions but fostering an ecosystem where both human and machine creativity can flourish, with data acting as the fuel for innovation.

Dentons' Artificial Intelligence Survey conducted in 2021 raised the following question:

> "If an AI system develops intellectual property, who or what do you believe should own the intellectual property rights and patents?"[1]

58 percent of participants hold the view that IP rights should be owned by a user of the AI system, with 20 percent advocating for the rights to belong to the inventor of the AI system. Only 4 percent of respondents consider the AI itself as the owner of IP rights.

The survey clearly showed that granting AI personhood, a notion explored, e.g., in contexts such as AI liability (Kurki 2019; Novelli 2022), seems like a dead end. Many comments on the DABUS case failed to provide assistance, as they primarily addressed patent offices' denial of patent applications naming AI as an inventor (see George and Walsh 2022), since assigning ownership to a legally prima facie non-existent entity creates more problems than it resolves. This is particularly true when it comes to determining who can legitimately claim a share of any revenues arising from an AI-based invention. Indeed, the central issue revolves around identifying who contributes to novelty and non-obviousness, both of which are rooted in the concept of creativity.

According to Dentons' Survey, we thus see two competing claims:

- **Inventor of the AI -** There is a strong case to be made for the inventor who created the AI system. Their ingenuity led to the development of the tool that ultimately produced the invention.

[1] https://www.dentons.com/en/insights/articles/2022/june/8/legal-regulation-of-artificial-intelligence-in-kazakhstan-and-abroad#:~:text=Only%20humans%20can%20be%20considered,labor%20has%20created%20a%20work.&text=Only%20individuals%20can%20be%20inventors

- **Company using the AI -** Companies often invest heavily in the development and training of AI systems. They might argue that their resources and expertise fostered the environment for the invention to occur.

Surprisingly, until recently, the significance of large and high-quality datasets was not regarded as a primary concern in recognizing IP rights for AI-based inventions and creations. Hence, we have a third claim:

- **The Data Owner's Stake -** The role of data becomes crucial when considering "combinational creativity" as the basis of novelty. Here, the value lies in the unique way disparate data points are combined, not just the data itself.

Imagine AI system *X* invents *Y* based on data set *Z*. Suppose data *Z* is essential for creating *Y*, or for training *X* to invent or create *Y* altogether (think of supervised learning models). In other words, the reasoning is: *If the AI system did not have access to the data in Z, the invention or creation Y would not exist as such and would not have the same value in terms of novelty and non-obviousness.*

One may thus see the legal issue as based on the following challenges.

1) **The Data Quandary:**
   - **Data as a Building Block -** If data *Z* is considered a fundamental building block, then should not its owner have some claim to the IPR of *Y*? After all, without *Z*, *Y* would not exist.
   - **Data as a Catalyst -** On the other hand, data *Z* might simply be a catalyst, allowing *X* to reach its full potential and generate *Y*. In this case, the ingenuity of *X* might be the key factor.
2) **Finding the Balance:**
   Is determining ownership becoming a balancing act? We need to acknowledge the contributions of the AI developer, the company that fostered innovation, and potentially, the owner of the crucial data set.
3) **Possible Solutions:**
- **Joint Ownership -** Co-ownership between the AI developer, the company, and the data owner could be a solution, with shares reflecting their contributions.
- **Licensing Agreements -** Data owners could license their data for AI development, with clauses outlining potential IPR rights if the data leads to inventions.

4) **The Road Ahead:**
   The legal framework surrounding AI-generated inventions is still evolving. Open discussions involving legislators, AI developers, data owners, and corporations will be crucial in creating a fair and sustainable system for IPR ownership in this new age of innovation. However, it is clear that data value will play more and more a role in IPR for AI-based inventions and creations.

In the next sections we will provide philosophical and legal theoretical reasons supporting the idea that AI creativity may be sufficient to ground IP protection. We will start examining the idea of AI creativity.

4. Creativity

Margaret Boden (2004) played a pioneering role in the field of creativity in AI, providing an interesting conceptual framework. Boden has explored how machines can display creative

potential, but the emergence of generative AI technologies challenges the conventional notions of what constitutes creativity. This distinction becomes crucial as we navigate the complexities of AI outputs, which often mimic creativity without embodying the full spectrum of human creative processes. Boden (2004, 1) defines (machine) creativity as the ability to generate ideas or artefacts that are *new*, *surprising*, and *valuable*. Indeed, this concept closely aligns with three essential characteristics of inventions eligible for patenting, being that

1) *novelty* corresponds to the fact that the generated ideas or artefacts are *new*;
2) *the inventive step or non-obviousness* correspond to the fact that the generated ideas or artefacts are *surprising*;
3) *capability of industrial application or utility* correspond to the fact that the generated ideas or artefacts are *valuable*.

However, which type of creativity do make AI-based inventions and creations suitable for making some *X* new and surprising, thus generating IP rights?

As is well known, Boden (2004, 1ff.) distinguishes three kinds of creativity:

- *combinational*, which "involves making unfamiliar combinations of familiar ideas";
- *exploratory*, which requires exploring a conceptual space;
- *transformational*, that allows for transforming a conceptual space and brings new and previously inconceivable thoughts to become possible.

Examples from the first type "include poetic imagery, collage in painting or textile art, and analogies. These new combinations can be generated either deliberately or, often, unconsciously. Think of a physicist comparing an atom to the solar system, for instance, or a journalist comparing a politician with a decidedly non-cuddly animal" (ibid., 3).

The second type of creativity is based on the idea of conceptual space, which is any structured style of thought shaped for example by a culture context. Within this space, new ideas can emerge through exploratory creativity leading to the discovery of previously unnoticed possibilities. Boden recalls the analogy of driving into the countryside with a map to illustrate this concept, where exploring new roads represents exploring new ideas within a particular style of thinking.

Following the above-mentioned metaphor of geographical maps, the third type of creativity requires transforming mental maps. While geographical maps are static, mental maps can change through creative thinking. Changes to mental maps can be small and superficial or significant and transformative, akin to adding roof extensions to a village or rerouting a motorway.

More recently, Boden (2016) has directly addressed the question of whether AI can genuinely be considered creative. Her objective is to grasp the complexity offered by philosophical literature on the concept of artificial creativity. While AI has made significant strides in mimicking creative processes, its ability to possess true understanding, intentionality, and consciousness remains in debate (ibid., chap. 6). When evaluating AI's creative capabilities, we must examine whether these systems truly possess intelligence or understanding. For instance, if we develop an AI that can compose symphonies or generate visual art, we must ask: does this achievement imply that the AI understands the nuances of music or art? Many individuals intuitively respond with a resounding "no". The philosophical implications challenge our perceptions of what constitutes creativity and whether it can exist independently from consciousness. According to Boden's reconstruction, AI excels at manipulating vast datasets and identifying patterns but struggles with aspects of human creativity related to context and emotional depth. For example, an AI system trained on historical music can generate new compositions. However, *its outputs are outputs of combinational creativity*, where it remixes existing ideas rather than originating truly novel concepts. This

distinction draws attention to the limits of AI's creative capacity; it can produce new combinations but lacks the intrinsic motivation and emotional grounding that characterize human creativity.

These reflections are essentially pre-generative AI (see also Colton 2008), which changes some elements in the way in which philosophers had conceptualised AI creativity. On the other hand, precisely Boden's (2004) framework and other related models of creativity have been applied in the context of recent AI technologies (for an overview, see Franceschelli and Musolesi 2021): if we adhere to Boden's original analysis, it appears that all the above three forms of creativity are pertinent to computers. This assertion has been acknowledged in the context of generative AI (Franceschelli and Musolesi 2021 and 2024; see also Farina et al. 2024).

Other contributions (Runco 2023; see also Veale et al. 2019), however, discuss the boundaries of what we consider creative in his examination of AI's capabilities. Runco (2023) suggests that the outputs of AI often reflect a form of *pseudo-creativity*, where they appear innovative but fundamentally lack the deep understanding, intentionality, and self-actualization present in human creative endeavours. By exploring these characteristics further, Runco posits that while AI may produce impressive outcomes, it still does not equate to authentic human creativity.

But—we contend—even if we may agree on the view that AI lacks essential qualities of human creativity—such as consciousness, intentionality, and genuine understanding—this is not sufficient to avoid the legal problem outlined in Section 3 and that can be summarised in the following challenge.

Challenge 1
Can we rule out that the outcome of AI's (pseudo-)creativity, whether an invention or a piece of art, might be considered novel and non-obvious to a "person skilled in the art", or not original and valuable? Evidence and compelling legal reasons—such as developing legal-philosophical grounds to identify those who should own the intellectual property rights and patents—remain, and the challenge is still unresolved.

5. AI Creativity: Turing Test, Inventor Room, and Other Creative Beasts

One may address the problem of novelty and non-obviousness of AI-based inventions and creations by resorting to suitable adjustments of the Turing test for AI-based creativity.

As is well known, the Turing Test, a landmark thought experiment, aimed to assess a machine's ability to exhibit human-like intelligence through conversation. However, creativity, a hallmark of human ingenuity, presents a distinct challenge. Can we devise a similar test to recognise creativity in AI? Turing test suggests that if an AI can generate responses indistinguishable from a human's, it could be said to "think" as humans do. However, critics argue that passing the Turing Test does not equate to genuine intelligence or creativity. This opens the door to the notion that an AI's output, while possibly original and effective, may not arise from a true creative process. Therefore, creativity might require qualities that AI systems currently lack (Boden 2016, chap. 6).

In fact, the Chinese Room Argument, proposed by John Searle (1980), challenged the idea that manipulating symbols signifies true understanding. As is well known, this argument is a thought experiment in which Searle imagines himself in a room, using a computer program to respond to Chinese characters slipped under the door, even though he does not understand Chinese. By manipulating the symbols according to the program, he produces appropriate responses that mislead those outside into believing they are interacting with a Chinese speaker. Searle's main

conclusion is that, while a computer can mimic understanding through programming, it does not achieve genuine comprehension. Therefore, he argues that the Turing Test is insufficient as a measure of genuine understanding. Similarly, concerns exist that AI may simply be a sophisticated symbol manipulator, apparently lacking the essence of creativity. This raises the question: is intentionality, a core aspect of creativity, even relevant when considering patentability of AI-generated inventions?

To this end, it is worth considering the Lovelace Test (see Bringsjord et al. 2001), which goes beyond Turing test and proposes three criteria for an AI (*A*) designed by a human (*H*) to be considered "creative":

5. Firstly, *A* must produce an output (*o*) that is demonstrably novel or innovative;
6. Secondly, this output must be a product of *A*'s internal processes, not a random fluke; importantly, *A* must be able to reliably reproduce *o*, indicating a repeatable creative process;
7. Finally, the most challenging criterion: neither *H* nor anyone with access to *H*'s knowledge and resources can explain o solely through *A*'s architecture, knowledge base, or core functions.

This last criterion is crucial. If *o* can be fully explained by *A*'s internal workings, it suggests a deterministic process, akin to a complex calculator producing a pre-determined answer. True creativity, however, often involves elements of surprise and the ability to generate outcomes beyond the sum of its parts.

Subsequent analysis has criticised this test and a refinement, called Lovelace 2.0 Test, has been elaborated (Riedl 2014). At any rate, the Lovelace Test offers in general a compelling methodological framework but also faces challenges. Defining "novelty" objectively is complex. What might be ground-breaking in one field could be an incremental step in another. Additionally, the test relies heavily on human judgment in determining if an explanation for *o* is sufficient. Finally, if the issue is establishing the objective novelty of an invention, a test could provide a useful proxy, although it may not be conclusive.

Despite these hurdles, the Lovelace Test opens a crucial conversation. As AI continues to evolve, the ability to recognise and value its creative contributions will become increasingly important. The Lovelace Test, or a future iteration, serves as a valuable stepping stone in our quest to understand and nurture the creativity of machines, but it may not provide a definitive answer to clarify the objective nature of AI-based novelty.

The next sections will attack the problem at a more foundational level moving from Challenge 1 above and with a special focus on generative AI.

6. Which Creativity for AI?

Let us consider the use of generative AI in healthcare. Specifically, imagine an AI system employed in the pharmaceutical industry tasked with developing novel drug compounds. In traditional drug discovery, researchers analyse large datasets of chemical properties and conduct numerous experiments, a process that can take years. Now, with generative AI, this timeline dramatically reduces. For instance, an AI trained on existing molecular structures could rapidly generate unique chemical combinations. In a case study, such an AI was reported to produce potential drug candidates significantly faster than human researchers, leading to breakthroughs in treatments that previously would have taken decades to materialize. This scenario raises profound regulatory questions: How do we classify and patent these AI-generated compounds?

In general, what do we really mean in these scenarios by "new" and "non-obvious"? Clearly, the challenging aspect is novelty, because the non-obviousness is based on the idea that "the invention would not be obvious to a 'person skilled in the art' who has 'common general knowledge' in that field": in other words, the non-obviousness of an invention or creation *X* requires to perform a counterfactual reasoning in which we should imagine that any existing "person skilled in the art" who has "common general knowledge" in the field acknowledges that *X* is surprising. On the contrary, we do not think that the above purely subjective reconstruction of non-obviousness can be extended as well to the idea of novelty.

In examining the notion of novelty in the light of Challenge 1, we believe that a fruitful philosophical understanding of the foundations of IPR for AI-based inventions and creations may be limited to the concept of combinational creativity. If so, Boden's framework of creativity offers a lens through which we can examine the philosophical underpinnings of IPR for AI-generated inventions and creations. While AI exhibits a form of creativity, it likely falls short of the transformative kind, raising questions about how we credit innovation in the age of machines. However, this does not diminish the value of AI's contribution in the legal perspective.

Let us revisit combinational creativity, the art of remixing existing ideas. This aligns perfectly with how many *AI generative models* function. They are, in essence, data alchemists, transforming vast quantities of information into novel outputs. Borrowing from Amos Oz's famous quote

> "If you steal from one book you are condemned as a plagiarist, but if you steal from ten books you are considered a scholar, and if you steal from thirty or forty books, a distinguished scholar." (Oz 2005, 129)

if we follow that metaphor, AI can be seen as a scholar who has access to a library of unimaginable size, weaving together existing knowledge to produce something new. Here, data is the key ingredient. The richer and more diverse the data, the more surprising and potentially valuable the combinations AI can generate. One may object that the idea of inspiration and the reuse of narrative elements, as illustrated by the Oz quote above, represent only a fraction of how humans engage in creative processes. Indeed, many philosophical perspectives on inspiration, plagiarism, and creation can be drawn from sources such as Aristotle's concept of mimesis, Deleuze and Guattari's view (1987) or Foucault's (1977) analysis. However, *our objective is not to definite human authorship*, but to understand if the type of AI-based creativity is adequate enough to fall withing the legal protection of IP law, a regulatory paradigm that was recalled in Sections 2 and 3[2].

True invention in the context of Challenge 1 often requires a deeper kind of creativity—transformational creativity. Here, the very fabric of knowledge is reshaped. This is where AI seems to struggle, since one may argue that, despite technical analysis (Franceschelli and Musolesi 2021), AI lacks the true ability to fundamentally alter the underlying frameworks that organise information.

To understand this potential limitation, let us elaborate the concept of the "Ideal Observer" in ethics (Firth 1951). According to this concept, something *X* is deemed right in ethical contexts if

[2] Perhaps, M. Foucault's view could be of some inspiration for non-analytic philosophers when he says (Foucault 1977, 138): "The author—or what I have called the 'author-function'—is undoubtedly only one of the possible specifications of the subject and, considering past historical transformations, it appears that the form, the complexity, and even the existence of this function are far from immutable. We can easily imagine a culture where discourse would circulate without any need for an author. Discourses, whatever their status, form, or value, and regardless of our manner of handling them, would unfold in a pervasive anonymity". Could we consider AI as an anonymous collector of discourses?

there existed a person who was omniscient, omnipercipient, disinterested, and dispassionate, yet otherwise ordinary, and this person would have a particular experience (such as feeling approval or perceiving a sense of obligation) regarding *X*. *The concept of Ideal Observer thus provides ideal epistemic conditions under which knowledge ensures moral rightness.*

If reframed in the AI domain, we can think of an "Ideal Inventor"—a hypothetical being with unrestricted access to all information and the capacity to grasp every potential connection within that data. By this standard, only outputs that transcend the existing data landscape, creations the Ideal Inventor could not have predicted, could qualify as true inventions. In this second case, *the concept of Ideal Inventor may provide ideal epistemic conditions under which knowledge ensures non-trivial forms of creativity*.

Now, how to relate combinational creativity with the transformational one in the context of the Ideal Inventor?

Let us suppose that the Ideal Inventor had access to the set *F* of all possible science fiction novels written so far and those that could be written in the future by present and future human beings. Obviously, this is a structurally counterfactual assumption, but conceivable on an ideal epistemic level. The question then is: if such an Inventor found herself in these ideal epistemic conditions, and if such an Inventor were to write a new science fiction novel *R* by combinationally exploiting the data in *F*, could we define transformational creativity in terms of combinational creativity? Note that *R* does not belong by definition to *F* because the Ideal Inventor is an AI system, whereas *F* contains only novels written by humans.

While any formal reconstruction of Boden's theory of creativity is not obvious (cf. Kronfeldner 2018; Said-Metwaly et al. 2017), as it depends on the notion of conceptual space and on compositional and other change patterns applied to them (Thornton 2007), it seems to me the Idea Inventor perspective may lead to assuming that an infinite set like *F* at least prima facie covers all possible results from transformational creativity. Indeed, the conjecture is that, for any transformational-creativity-based outcome *o*, there is any dataset *D* in *F* such that *D* grounds *o* via combinational creativity.

Though a formal reconstruction of Boden's theory of creativity is not straightforward (cf. Kronfeldner 2018; Said-Metwaly et al. 2017), as it relies on the formal account of the idea of conceptual space and the application of compositional and other change patterns to it (Thornton 2007), it appears to me that positing an infinite set such as *F* prima facie encompasses all potential outcomes of transformational creativity. Indeed, the hypothesis suggests that for any outcome *o* based on transformational creativity, there exists a dataset *D* in *F* that underpins *o* through combinational creativity: the set *F* is thus the *ideal dataset* for a certain field of knowledge.

We can thus argue in favour of the following conjecture:

Conjecture 1
A creation or invention *X* in the field φ made by AI system *Y* is *new* **if** *X* would not be created through combinational creativity by the Ideal Inventor without having access to the ideal dataset *F* for φ.

Notice that if we replace '**if**' with '**if and only if**' in Conjecture 1, we adopt a reductionist conception of AI creativity, viewing it as nothing more than combinational creativity. This is not our position; we argue that the greater the reliance on larger datasets for combinational creativity, the closer the output comes to resembling the kind of creativity to be legally protected according to Challenge 1**Challenge 1**.

7. AI-based Novelty as a Supervenient Property on Datasets

If we assert that something *X* is new in the legal context of TRIPS, one might pose the question: what exactly constitutes novelty? Essentially, novelty is a *relational property*: for *X* to be considered new, it must possess that property with respect to something else and, typically, within a particular domain. For instance, it is widely recognised that new therapeutic applications for existing drugs often entail significant inventive advancements (see Verbaanderd et al. 2021): it is a nonsense to say Viagra has not been a novelty in abstract terms, only because the drug molecule was well-known but indicated for hypertension and angina. However, the relational aspect of "new" does not fully elucidate the inherent nature of novelty as a property.

In the preceding sections, we have offered a foundational philosophical discourse on how to associate AI-based novelty with combinational creativity and the value of data (especially in the context of generative AI). However, we still need to elucidate the *ontological status of novelty* within this perspective.

To address this problem, it is interesting to start with Tony Lawson (2016)'s theory of social reality and of artefacts[3]:

> "Social reality is in large part at least structured by interlocking, internally related, often spontaneously emergent collective practices, carrying, in the sense of manifesting, (often contested) rights and obligations inter-relating the human beings who undertake these practices as positioned components of communities, of which the positions and practices in turn are properties. As such this social reality is in a sense both given to individual human beings at the moment each comes to act, as a (typically unacknowledged) condition of their individual (positioned) practices, and also reproduced and/or transformed as a (typically unintended) result of these individual practices taken in total." (Lawson 2016, 9-10).

Lawson's realist theory of social science is grounded on "insights and assessments whose truth or falsity is not contingent upon the opinions of observers but rather on the objective nature of the world" (Lawson 2016, 2). According to Lawson, the social realm encompasses phenomena whose existence relies inherently on human beings and their interactions. Human beings navigate the social realm through their "positioning". Positioning entails attributing rights and obligations to individuals, which entail positive and negative positional powers, respectively. Social reality comprises not only human beings but also other things, among which we have "artefacts". Artefacts possess a material dimension, such as the physical form of a coin. However, this material aspect alone does not encapsulate the essence of artefacts' existence. In social reality, artefacts' function is based on collective recognition or acceptance by human beings, acknowledging them as having certain properties. Hence, a coin can only fulfil its currency functions when, through agreement or convention, its specific physical form is practically positioned and relationally organised as a component of a broader system of value accounting accepted within a community. Artefacts arise from a physical foundation, giving rise to a distinct realm of interaction with characteristics separate from those of the natural environment.

Can we extend the intuition above to (generative) AI and argue that the property of being new in AI-based inventions arise from the structure of data as a separate property precisely from the level of such data? Perhaps, we can couple Lawson intuition with leveraging the philosophical concept

[3] Many thanks to Corrado Roversi for pointing out this idea.

of supervenience and thus can conjecture that *novelty is a supervenient property on the "material" properties of data* (see McLaughlin and Bennett 2023).

As is well-known, supervenience is rather complex notion, but it essentially means that one set of properties *A* cannot differ without a corresponding difference in another set of properties *B*. Imagine an art forgery, identical down to the molecules. While it differs from the original in its history and value, it necessarily shares the same colour arrangement due to its identical physical makeup. In this case, the arrangement of colours (property *A*) supervenes on the microphysical properties (property *B*) of the painting (ibid., sec. 1).

Similarly, we can argue that the novelty of an AI invention (property *A*) supervenes on the data it is trained on (property *B*). No two AI systems can create genuinely novel inventions if their training data is identical. The data serves as the foundation for the AI is internal processes, shaping its understanding of the world and biasing it towards specific patterns. Just like the forger's canvas cannot have different coloured shapes without a physical difference, an AI cannot invent something truly new without a fresh large set of data exposing it to unseen patterns and relationships.

Here is a breakdown to illustrate this point:

- AI System *A* trained on data set *X* invents solution *Y*.
- AI System *B* trained on entirely different data set *Z* invents solution *W*.

If solutions *Y* and *W* are truly novel and functionally distinct, then data sets *X* and *Z* must also be fundamentally different. This reinforces the idea that the novelty of the invention supervenes on the properties of the data.

It is however important to acknowledge some nuances. Large data sets often overlap to some degree, and AI systems might exhibit some level of creativity by combining existing knowledge in new ways. However, the core argument remains firm: substantial novelty is unlikely to emerge from repetitive exposure to similar data.[4]

One counterargument might be that the way data is processed by the AI algorithm plays a role in inventive novelty. While the algorithm is undeniably important, it can be seen as a tool that operates on the data. The data itself provides the raw material for invention, and the algorithm functions to manipulate that material in a way dictated by its inherent properties. However, the more the algorithm asymptotically tends toward operating with the infinite set of data *F* of the Ideal Inventor, the more combinational creativity renders the property of being new corresponding to the supervenience of a structure not present in *F* and correlating data in *F*.

Perhaps, this suggests that the type of supervenience for this idea is global supervenience, in which the assertion "there cannot be an *A*-difference without a *B*-difference" is not directed towards individuals, but rather towards entire possible worlds. Under this assumption, the concept of supervenience offers a compelling lens to examine the essence of novelty in AI inventions. The data an AI system is trained on serves as the foundation for its inventive capabilities, and true novelty seems to arise from feeding the system with fresh large data sets that expose it to unseen patterns and relationships within the information landscape.

One must observe that the above analysis does not necessarily lead to AI-based creativity without human contribution. However, we acknowledge that such an analysis can, in the long run, place the human role in the background. Concepts like "co-intelligence," "hybrid intelligence," and "co-

[4] One unresolved issue pertains to whether ideal datasets (as discussed in Section 6) are mutually exclusive and do not overlap. If they are, this objection is alleviated provided that datasets are sufficiently large, and it ceases to be a concern under the conditions in Conjecture 1.

cre-ai-tivity" highlight that AI should not be viewed merely as an independent tool developed by humans. The concept of intelligence development, regardless of its substrate—whether human or machine—relates to the co-evolution of machine and human intelligence in addressing significant global challenges and fostering creativity. According to this view, human input remains a vital factor in activating AI (see Krinkin et al. 2023). However, through the idea of human-AI coevolution (cf. Pedreschi et al. 2024), if humans and AI continuously influence each other such that the interaction between users and AI creates a feedback loop, then human contributions cannot be regarded as purely human but rather as inherently hybrid. This precisely re-presents the dilemma we have discussed in this chapter.

8. Consequences and Further Complexities

The rise of AI-powered invention presents a fascinating philosophical challenge to the very foundation of IP law. This section mentions two further issues.

Issue 1: Obviousness in the Age of AI Super-Knowledge

Let us recall the scenario mentioned at the beginning of Section 5. Imagine an AI system, let us call it PharmaAI, that has been trained on millions of pharmaceutical compounds, their chemical properties, and the outcomes of previous drug trials. PharmaAI can identify potential new drug candidates for diseases faster than any human researcher by analysing large amounts of data and recognizing patterns that humans might overlook. When a researcher submits a patent application for a drug candidate identified by PharmaAI, questions arise regarding its obviousness. Human patent examiners may struggle to evaluate the candidate's novelty because PharmaAI's knowledge base exceeds that of any individual expert in the field. If PharmaAI suggests a novel drug compound that combines existing molecules in a unique way, can human patent examiners determine whether this combination is obvious, considering PharmaAI's extensive dataset? If PharmaAI identifies a target that is already well-known in the scientific community, but the specific way it combines molecules is unprecedented, it complicates the traditional assessment of "obviousness" to a "person skilled in the art".

George and Walsh (2022) raise precisely this critical point: as AI surpasses human knowledge in specific fields, human patent examiners become ill-equipped to judge the "obviousness" of an AI invention. As we have recalled, IP law grants patents only to inventions that are not obvious to a "person skilled in the art" with "common general knowledge" within the domain. Yet, an AI system trained on a vast dataset would possess knowledge far exceeding any human. Would not almost everything then seem obvious when compared to this super-knowledge?

Here, a philosophical approach can offer solutions. We can distinguish between two types of knowledge:

- **Raw data and associations -** AI systems excel at storing and manipulating massive datasets, identifying patterns and correlations;
- **Human creativity and context -** However, unless we assume to work with an AI system as an Ideal Inventor, humans excel at understanding the context and purpose of inventions, drawing upon broader knowledge and a sense of creative possibility.

Perhaps the key to assessing obviousness lies in this distinction. While an AI might identify the building blocks for an invention within its vast data, a human examiner can judge whether the specific arrangement and application of those elements demonstrate a genuine "inventive step". This inventive step could involve a novel application of existing knowledge, a surprising

combination of elements, or a solution that addresses a previously unrecognised problem. By focusing on this creative aspect, human examiners can still evaluate inventions from AI systems even in the face of their vast knowledge.

Issue 2: Nested (Iterative) Intellectual Property and the AI Inventor

The second issue has to do with the thorny question of ownership for inventions created by what we can call *nested or iterative AI-based inventions*. Imagine AI system *X*, protected by IP law, that invents AI system *Y*, which in turn generates invention *Z*. Who owns invention *Z*? Is it *X*, the original AI, or *Y*, the immediate creator?

Here, we can draw inspiration from the concept of "authorship" in human cooperation (as for example described by Bratman 1992). If two or more humans cooperate on a work, they shared subplans and authorship is typically shared based on their individual contributions. Similarly, the ownership of invention *Z* could depend on the specific roles of *X* and *Y*. If *X* provided the underlying framework and training data, while *Y* merely acted as a tool for processing and generating the invention, then *X* might hold a stronger claim to ownership. Conversely, if *Y* displayed significant independent creativity in the invention process, its claim might be stronger.

As an example, consider a generative AI program, **InventorX**, which is capable of generating designs for innovative products. InventorX is developed under IP law and produces a new AI system, **InventorY**, specifically designed to enhance its creative capabilities. InventorY, then, generates a groundbreaking invention, **GadgetZ**.

The question arises: **Who owns the IP for GadgetZ?**

1. **Ownership claims for InventorX:** If InventorX laid the foundational algorithms and data that allowed InventorY to create GadgetZ, it might argue for ownership based on its initial contribution.
2. **Ownership claims for InventorY:** If InventorY was programmed to exhibit a level of creativity that surpasses that of its predecessor and has adapted through a learning process, it could assert a claim to ownership based on its innovative outputs.
3. **Shared ownership scenario:** Taking inspiration from human collaboration, the ownership may be shared between InventorX and InventorY. This model could involve assessing the unique contributions of each AI system. If InventorX primarily provided the framework while InventorY executed the creative process, ownership could reflect their respective roles and input.

This layered scenario raises complex questions about intellectual property rights within AI-generated innovations, particularly in determining how to navigate ownership claims when multiple AI systems contribute to the invention process.

In very abstract terms we may imagine several scenarios, up to describing a sort of iterative (recursive) invention. Let **Invent** be a function mapping any AI-system *X* and any dataset *F* into an invention *Y*. In other words,

- let *AI* and *IdealAI* be the set, respectively, of AI systems and AI systems which are ideal inventors (where $IdealAI \subset AI$),
- let *DataSet* and *IdealDataSet* be the sets, respectively, of datasets and of ideal datasets (where $IdealDataSet \subset DataSet$),
- let *Inv* be the set of AI-based inventions where $AI, IdealAI \subset Inv$.

Hence, the function is defined as follows:

$$\textbf{Invent}: AI \times DataSet \rightarrow Inv.$$

An *iterative invention* is such that, for any $X \in AI$ and any $F,G \in DataSet$

$$\textbf{Invent}\ (X, F) = \textbf{Invent}\ (\textbf{Invent}\ (X, F), G)$$

The theoretical question has to with the nature of *X*, *F* and *G*. If we aim at ensuring novelty in AI-based inventions, we should follow Conjecture 1 and establish the following:

- $X \in IdealAI$
- $F,G \in IdealDataSet$
- **Invent** $(X, F) \in IdealAI$.

Of course, other (less ideal) scenarios are possible. Which ones are acceptable to ensure iterative novelty? This remains an open question for future research.

9. Summary

In conclusion, our examination of the issue AI-based inventions within the framework of the TRIPS principles has shed light on several critical issues surrounding the evaluation of AI-based inventions and creations. Firstly, regarding novelty, we have questioned whether outputs generated by AI can be considered truly "new" if they stem from the vast data AI has been trained on. Determining the novelty of AI-generated inventions requires careful consideration of theoretical assumptions and sophisticated techniques to distinguish originality from underlying data.

Similarly, assessing the inventive step of AI-generated inventions presents challenges. Can an invention be deemed non-obvious if it originates from a non-human entity operating beyond human intuition? This necessitates exploration into the nature of "inventiveness" in AI, distinct from human ingenuity.

Building on Boden's theory of creativity, we have recognised the relevance of multiple forms of creativity to computers, suggesting that AI possesses creative potential acknowledged by Boden herself.

Despite difficulties, we have also argued that the Lovelace Test sparks vital discourse on recognizing and valuing AI's creative contributions. While it may not definitively clarify the objective nature of AI-based novelty, it serves as a crucial step towards understanding and nurturing machine creativity. More importantly, our examination of creativity in AI through the lens of the concept of Ideal Inventor and of supervenience suggests that true novelty arises from exposing AI to fresh and very large datasets, allowing it to uncover unseen patterns and relationships within the information landscape.

Finally, considering the consequences and complexities, as well as the potential legal-theoretical impact of our analysis, underscores the need for ongoing research and dialogue to navigate the evolving landscape of AI creativity and its implications.

Acknowledgements

Antonino Rotolo was partially supported by the projects CN1 “National Centre for HPC, Big Data and Quantum Computing” (CUP: J33C22001170001) and PE01 “Future Artificial Intelligence Research” FAIR (CUP: J33C22002830006).

References

Boden, M.A. 2004. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

Boden, M.A. 2016. *AI: Its Nature and Future*. Oxford: Oxford University Press.

Bratman, M.E. 1992. Shared Cooperative Activity. *The Philosophical Review* 101(2): 327-341.

Bringsjord, S., Bello, P. and Ferrucci, D. 2001. Creativity, the Turing Test, and the (Better) Lovelace Test. *Minds and Machines* 11: 3-27.

Colton, S. 2008. Creativity Versus the Perception of Creativity in Computational System. In *Creative Intelligent Systems - Papers from the AAAI Spring Symposium*, Technical Report. Stanford: AAAI Press.

Deleuze, G. and Guattari, F. 1987. *A Thousand Plateaus*. Minneapolis: University of Minnesota Press.

Farina, M., Lavazza, A., Sartori, G., and Pedrycz, W. 2024. Machine Learning in Human Creativity: Status and Perspectives. *AI & Society*, online first.

Foucault, M. 1977. What Is an Author? In D.F. Bouchard (ed.), *Language, Counter-memory, Practice*. (Ithaca, NY: Cornell University Press.

Firth, R. 1951. Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research* 12 (3): 317-345.

Franceschelli, G., and Musolesi, M. 2021. Creativity and Machine Learning: A Survey. arXiv:2104.02726 [cs.LG].

Franceschelli, G., and Musolesi, M. 2024. Reinforcement Learning for Generative AI: State of the Art, Opportunities and Open Research Challenges. *Journal of Artificial Intelligence Research* 79: 417-446.

George, D. and Walsh, J. P. 2022. Can Machines Be Inventors? Unveiling the Legal and Philosophical Complexities of AI-generated Inventions. *Nature* 605 (7909): 50-54.

Krinkin, K., Shichkina, Y. and Ignatyev, A. 2023. Co-evolutionary Hybrid Intelligence is a Key Concept for the World Intellectualization. *Kybernetes* 52(9): 2907-2923.

Kronfeldner, M. 2018. Explaining Creativity. In B. Gaut and M. Kieran (eds.), *Routledge Handbook on Creativity and Philosophy*, 213-229. London: Routledge.

Kurki, V.A.J. 2019. The Legal Personhood of Artificial Intelligences. In *A Theory of Legal Personhood*, 175-189. Oxford: Oxford University Press.

Lawson, T. 2016. Comparing Conceptions of Social Ontology: Emergent Social Entities and/or Institutional Facts? *Journal for the Theory of Social Behaviour* 46(4): 359-399.

McLaughlin, B. and Bennett, K. 2023. Supervenience. In E.N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), URL = <https://plato.stanford.edu/archives/win2023/entries/supervenience/>.

Novelli, C. 2022. *AI and Legal Personhood: A Theoretical Survey*. PhD thesis. Bologna/Luxembourg: University of Bologna and University of Luxembourg.

Oz. A. 2005. *A Tale of Love and Darkness*. San Diego: Harcourt.

Pedreschi, D., L. Pappalardo, E. Ferragina, R. Baeza-Yates, A.-L. Barabasi, F. Dignum, V. Dignum, T. Eliassi-Rad, F. Giannotti, J. Kertesz, A. Knott, Y. Ioannidis, P. Lukowicz, A. Passarella, A.S. Pentland, J. Shawe-Taylor, A. Vespignani. 2024. Human-AI Coevolution. arXiv:2306.13723.

Picht, P.G. and Thouvenin, F. 2023. AI and IP: Theory to Policy and Back Again – Policy and Research Recommendations at the Intersection of Artificial Intelligence and Intellectual Property. 2023. *IIC - International Review of Intellectual Property and Competition Law* 54: 916-940.

Reidl, M.O. 2014. The Lovelace 2.0 Test of Artificial Creativity and Intelligence. In *Proceedings of the AAAI Workshop: Beyond the Turing Test*: https://doi.org/10.48550/arXiv.1410.6142

Said-Metwaly, S., Van den Noortgate, W. and Kyndt, E. 2017. Approaches to Measuring Creativity: A Systematic Literature Review. *Creativity. Theories–Research-Applications* 4(2): 238-275.

Searle, J.R. 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3(3): 417-424.

Spedicato, G. 2019. Creatività artificiale, mercato e proprietà intellettuale. *Rivista di Diritto Industriale* LXVIII (4-5): 253-307.

Thornton, C. 2007. *A Quantitative Reconstruction of Boden's Creativity Theory*. Department of Informatics, University of Sussex.

Veale, T., Amílcar Cardoso, F., and Pérez y Pérez, R. 2019. Systematizing Creativity: A Computational View. In T. Veale and F. A. Cardoso (eds.), *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*, 1-19. New York: Springer International Publishing.

Verbaanderd, C., Rooman, I., and Huys, I. 2021. Exploring New Uses for Existing Drugs: Innovative Mechanisms to Fund Independent Clinical Research. *Trials* 22: 1-13.