Embodiment, Anthropomorphism, and Intellectual Property Rights for AI Creations

Deepak Somaya‡ and Lav R. Varshney‡
‡ Department of Business Administration and College of Law
‡ Coordinated Science Laboratory and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

Abstract

Computational creativity is an emerging branch of artificial intelligence (AI) concerned with algorithms that can create novel and high-quality ideas or artifacts, either autonomously or semi-autonomously in collaboration with people. Quite simply, such algorithms may be described as artificial innovation engines. These technologies raise questions of authorship/inventorship and of agency, which become further muddled by the social context induced by AI that may be physically-embodied or anthropomorphized. These questions are fundamentally intertwined with the provision of appropriate incentives for conducting and commercializing computational creativity research through intellectual property regimes. This paper reviews current understanding of intellectual property rights for AI, and explores possible framings for intellectual property policy in social context.

Introduction

A central concept emerging in discussions of robot law and policy is that of anthropomorphism: whether robots should be thought of as technological tools or as human-like social agents (Weaver 2014; Calo, Froomkin, and Kerr 2016; Jones and Millar 2017). Richards and Smart (2012) put forth the Android Fallacy, opposing the seductive metaphor that robots are “just like people” and forcefully rejecting the idea that there is a meaningful difference between humanoid and non-humanoid robots. Robots should be cast simply as tools or as “non-biological autonomous agents,” in their view, no matter their physical form. Interestingly, Darling (2015) has shown that using anthropomorphic language (such as personified names) for robots can impact how people perceive and treat them. When people’s perceptions of robots are framed in ways that discourage them from anthropomorphizing robots, policy debates tend to not view robots as having rights. There may, however, be settings where encouraging anthropomorphism is desirable. Since people and politics respond to both the physical forms and the framing of language, modulating these factors could help in separating the two cases in policy discussions.

Although the Android Fallacy and related notions of embodiment/anthropomorphism are emerging in discussions of liability, contracts, and criminal law, as far as we know, it has not previously been discussed in intellectual property (IP) contexts. This is the theme of the present paper.

Computational creativity is a branch of AI that encompasses the art, science, philosophy, and engineering of building computational systems that demonstrate behaviors that would be deemed creative by unbiased human observers. Although computational creativity as a goal for AI dates back to the 1956 Dartmouth Conference (McCarthy et al. 2006), there has been much recent progress in this field of research, including the formalization of what it means for software to be creative and the introduction of many exciting and valuable applications of creative software in the sciences, the arts, technology, cooking, literature, fashion, and elsewhere.

AI has now advanced to the stage where computationally creative systems are able to produce ideas and artifacts that are judged to meet standards of novelty and utility by experts in creative domains, see e.g. (Boden 2004; 2015; Colton and Wiggins 2012). Quite simply, computationally creative systems have become artificial innovation engines. Indeed, tests for creativity have recently been proposed to augment or replace the Turing test in assessing abilities of machines to exhibit intelligent behavior equivalent to that of a human (Riedl 2015). Examples that have captured much public attention include the IBM Chef Watson system that produces novel and flavorful culinary recipes (Varshney et al. 2013; Pinel and Varshney 2014; Pinel, Varshney, and Bhattacharjya 2015), and the Google Magenta system that composes novel and pleasing music (Bretan et al. 2017).

These creative machines are often operated semi-autonomously in a mixed-initiative form of human interaction, but they are also able to operate completely autonomously in producing novel artifacts (Lubart 2005; Smith, Whitehead, and Mateas 2011). Such AI systems may have physical embodiments as robots, and may be discussed using anthropomorphized language.

The central questions we aim to address in this paper are (1) whether IP protections should be offered to ideas or artifacts that are created purely by AI systems without human intervention or whether such use of creative technology should forfeit IP rights; and (2) how the embodi-
ment and anthropomorphization of such technologies may influence this IP debate. Indeed it may be possible to influence debates in ways that carve out IP negative spaces (Raustiala and Sprigman 2005; Rosenblatt 2011; 2013) or even create an IP anticommons (Heller and Eisenberg 1998; Somaya and Teece 2000). Some similarity to the uncanny valley (Mathur and Reichling 2016) is observed.

The remainder of the paper is organized as follows. First we review the current state of computational creativity technologies and systems, possible embodiments and anthropomorphisms, as well as human interactions with such systems. Then we discuss extant legal scholarship on IP for artificially-created ideas and artifacts. Next, we introduce our central argument on nuanced IP regimes and framing. Finally, we conclude.

**Computational Creativity**

Psychologists define creativity to be the generation of a product or service that is judged to be novel and also to be appropriate, useful, or valuable by a knowledgeable social group (Sawyer 2012), and is often said to be the pinnacle of intelligence (Boden 2004). This basic definition is also adopted in computational creativity and essentially yields artifacts that meet standards for patentability such as being novel, nonobvious, and useful. Due to greater competitiveness in global markets for all industries, computational creativity technologies are being adopted to make product/service development cycles more efficient. Thus, computational creativity is a kind of general-purpose technology that enables further invention and creativity.

There are a variety of algorithmic techniques that have been used within computational creativity, including genetic algorithms, simulated annealing, stochastic sampling and filtering, and deep neural network approaches, much beyond what some legal scholars have considered (Grimmelmann 2016). There are also various physical embodiments and levels of anthropomorphism that have been attempted. Fig. 1 and 2 show different physical embodiments that have been developed for the culinary computational creativity system that was once known as Blue Chef and is now called Chef Watson (thereby having a more personified name and background story). Notably there have been even humanoid embodiments, Fig. 2(c), through the Pepper personal robot platform, which stands 1.2m tall, has two arms and rolls around on a wheeled base, with a 10.1in tablet mounted on its chest. Empirically, users of creativity technologies are already anthropomorphizing them, e.g. Chef James Briscione describes the Chef Watson system as a social conversation partner, saying “Watson forced me to approach ingredients without any preconceived notions of which ingredients pair well together.” (Pinel 2015).

More broadly, general studies in the field of human-computer interaction have demonstrated positive effects of physical embodiment on the feeling of an artificial agent’s social presence (Lee et al. 2006). There are fundamental differences between virtual agents and physically-embodied robots from a social standpoint and have significant implications for human-robot interaction (Wainer et al. 2006; 2007), such as in the sense of authorship/inventorship and the sense of agency that the human feels (Hoesl and Butz 2017).

**Incentives in Intellectual Property Protection**

One intended purpose of IP regimes is to stimulate discovery. IP protection gives innovators an incentive to invest in new knowledge, but at the social cost of deadweight loss due to the resultant monopoly pricing and more importantly dynamic inefficiencies in innovation created by monopoly. Thus, IP policy ideally seeks to balance these two effects (Kitch 1977; Mazzoleni and Nelson 1998). When innovation is cumulative, an important incentive design problem is ensuring each inventor is rewarded enough to take account of the benefits conferred on future innovators. For this to work, each innovator must either receive enough of the total profit to cover costs (Scotchmer 2004) or have costs covered by other means such as government support of research. As a basic general-purpose technology, computational creativity research and development costs may need to be recouped from the profits of second-generation inventions developed via a creative AI system in order to create sufficient incentives for such basic innovation (Nelson 1959). As Scotchmer (2004) notes, “The sole source of profit on research tools is the second-generation products that they enable. To stifle such products would be economic suicide.”
Can machine-based innovations be protected so as to yield profits and appropriate upstream incentives? There are several possible approaches to this IP policy question. However, a central tenet of much copyright and patent law is that an IP right stems from and vests in a human creator or inventor, which would appear to constrain the available options. According to the U. S. Copyright Office, a work must be the product of “human authorship” to be entitled to copyright protection. Although case law remains largely unsettled, it seems autonomously machine-created works that would otherwise be copyrightable would currently enter the public domain. In contrast, in the U. K. and Australia, there are provisions to assign IP ownership of autonomously-created works to a particular person (Fitzgerald and Seidenspinner 2013; McCutcheon 2013b; 2013a). Moreover, Bridy (2012) has put forth the work-for-hire doctrine as a mechanism for vesting ownership of copyright in AI-authored works.

Although there is no specific administrative guideline or case law for U. S. patents, extant legal scholarship largely argues for autonomously machine-created inventions entering the public domain (Milde 1969; Clifford 1997; Acosta 2012). Moreover, the U. S. patent system ties the act of invention to conception and reduction to practice, such that an applicant cannot be granted a patent unless (1) the invention was formed in his mind and (2) he constructed an embodiment that operated for its intended purpose.

In nearly all current IP regimes, the results of semi-autonomous creation, rather than fully autonomous creation, have rights assigned to the human inventive partner that co-conceived the claimed invention with the assistance of the creative robot. In this scenario, the AI itself would be akin to a research tool, with all IP rights accruing to the human co-inventor. The people or company that invented the AI benefit only through explicit contracting under the presumption that the AI system is itself covered by some IP rights that are not exhausted by the sale of the system to the co-inventing party.

This conceptual distinction between autonomous and semi-autonomous invention creates the empirically difficult evidentiary test of delineating the two in practice. Clearly, users of AI innovation engines face perverse incentives to misrepresent the existence and extent of autonomous innovation by these creative agents. A similar incentive existed prior to the Civil War when slaves did not have IP rights; evidence suggests Eli Whitney’s cotton gin and Cyrus McCormick’s reaper were both at least partially the products of slave intellectual labor, though both inventions were claimed as their own (Vanatta 2013). This also poses an interesting adjudicatory problem as there may be no clear aggrieved party to bring a claim, assuming the AI system’s inventors’ rights are deemed to have been exhausted (e.g., by sale), though the AI system may disagree!

---

2Section 503.03 of Compendium II of Copyright Office Practices. See also (Ralston 2005).
**IP Negative Spaces and the Anticommons Problem**

We have thus far discussed IP regimes as justified through the theory of incentive to invent, a blunt-force economic argument. IP policy, however, cannot be considered in isolation from other institutional features of the context in which innovation is conducted. There are major fields of creativity and innovation where IP rights are either very limited or absent (Raustiala and Sprigman 2006; Darling 2014; Dreyfuss 2010), such as fashion, cuisine, tattoo artistry, professional magic, financial services, and sports, yet significant innovation occurs. These so-called IP negative spaces suggest that other factors may be promoting innovation.

Rosenblatt (2011) suggests that IP negative spaces are well-suited for creative domains that satisfy several possible criteria. These include settings where creation is largely driven by intrinsic motivation (perhaps via self-determination theory ideals of mastery and connectedness (Deci and Ryan 2012)) rather than just financial gain, and where more value may be generated by investing scarce resources in further creation rather than in securing and enforcing IP rights. Intrinsic motivation may be driven and reinforced by social structures such as communities of sharing (Rosenblatt 2013), of which the scientific commons is a prominent example. This suggests that in addition to exclusive rights as an incentive, in certain creative domains, innovation law and policy may promote creative communities that engage in social innovation (Lee 2014) even in the absence of strong IP rights for AI innovation engines.

The reverse may also hold true in some settings. Pervasive IP rights may constraining innovation by preventing the effective use and recombination of prior inventions. These issues have been amply illustrated in the concerns related to the patent anticommons problem (Heller 1998; Heller and Eisenberg 1998), where stronger patent protections may lead to under-utilization of innovative resources since no single entity is able to access all the rights needed for effectuating innovation. Fundamentally, the anticommons problems need to be viewed in terms of transaction costs: do IP rights that motivate upstream inventors by allowing them to capture returns from downstream activity end up creating too much friction in transacting for these rights so as to effectively stymie innovation (Somaya and Teece 2000)? An important issue related to the anticommons problem is that commercializing complex innovations like creative robots may be pursued through licensing or sale of physical product embodiments, but patent rights are generally assumed to be “exhausted” in the latter case but not the former. This creates perverse incentives for actors to license such upstream technologies so rights to downstream innovation may be preserved through contractual terms.

**Revisiting the Android Fallacy**

We have seen that in various extant IP regimes, a reticence to characterize AI systems in anthropomorphic ways (as epitomized by the Android Fallacy) may either hold back or fundamentally undermine appropriate policy regarding IP rights for AI innovation engines. Separately, we have seen some creative domains where IP negative spaces may be more appropriate than IP protection, or at the very least the transaction cost implications of strong upstream IP rights may be an important consideration to avoid creating downstream anticommons problems. How are these two dichotomies related, and how do they impact technology design (in terms of embodiment/anthropomorphism) in service of influencing policy debates?

As we see it, there are three possible framings of AI innovation engines:

1. Research tools similar to pencils, in that use has no inherent impact on downstream IP rights, even in the full autonomous setting,

2. Social agents with no property rights, such that any downstream IP enters the public domain, or

3. Social agents with some preservation of property rights, such that downstream IP need not enter the public domain.

These possibilities are qualitatively mapped in Fig. 3 along the dimension of downstream IP rights and a humanness dimension meant to capture how embodiment/anthropomorphism may impact framing of policy discussions. There is visual similarity to the uncanny valley.

Policies that treat creative robots as social agents without IP rights may not provide the intended economic incentives to their designers (computational creativity researchers). Unlike pencils or inference algorithms, the use of an autonomous AI could completely forfeit downstream IP protections, as the resulting artifacts may be deemed to have entered the public domain. There may therefore be a need for policies that maintain the downstream IP rights of purely autonomous innovation engines, so as to produce upstream profits to support computational creativity research. This need is especially acute in industries with strong IP protections and primarily extrinsically-motivated actors, such as in the area of industrial machinery, and where the potential transactional challenges related to an IP anticommons are more modest. To induce the framing of computational creativity systems as different from pencils—simply tools rather than social agents—in policy debates, physical embodiment and anthropomorphism may need to be avoided. For example, AI for industrial machinery creation may need to remain hidden or virtual, without personified name. Alternatively, one could go to the other extreme and achieve...
a level of humanness that would induce IP rights that are equivalent or near-equivalent to a human.

Contrarily, there are settings where creative robots as social agents without property rights may support computational creativity research: IP negative spaces or weak IP spaces. Examples include settings where the second-generation is in a low-IP industry where intellectual production outside the IP paradigm is common (such as food, magic, or comedy), or where people are primarily intrinsically motivated (Benkler 2006), or where opportunities for downstream innovation and creativity far outstrip and may be constrained by strong rights for the upstream AI. Self-determination theory has found people intrinsically want to interact and be connected to others, and so social bonds (whether human or machine) are vitally important (Baumeister and Leary 1995). Since forfeiture of downstream IP protections are less relevant in these creative domains, a creative AI might be licensed or sold at a fixed price, without reach-through rights to creations facilitated by the artificial innovation engine. When this happens, incentive structures simplify and can be understood as in the anticommons case. Strong IP rights on multiple upstream AI technologies, and the difficulties in transacting for them, might discourage use of creative robots and compound social losses from forgone innovation in downstream products (or even further generations of AI, each pushing the next).

Thus we see that in high-IP industries with extrinsic motivation as a key driver, virtual embodiment of AI technologies or pursuit of full humanness is appropriate; in low-IP industries or in settings with strong intrinsic motivation (IP negative spaces), however, medium levels of physical embodiment and anthropomorphism is the appropriate AI design principle. These design principles for AI systems that emerge from the desire for appropriate framing of IP policy discussion are important but thus far neglected in AI design.

**Conclusion**

To summarize, this paper first reviewed computationally-creative technologies, next discussed incentive structures needed to support cumulative innovation or IP negative spaces, then discussed legal scholarship on whether machines can be authors or inventors, and finally provided a policy distinction among treating AI as tools, social agents with, or without IP rights, governed by whether second-generation industries are extrinsically motivated with strong IP, or not. This led to principles for embodiment and anthropomorphism of AI systems, governed by the creative domain in which they are to be deployed.

This analysis speaks to the larger problem of legal personhood for AI where general abstract principles are often sought (Hassler 2017; Solaiman 2017), but here we suggest that the social context of AI deployment is central.

**Acknowledgments**

Thanks to G. I. Wang for pointers to IP law in the U. S., to F. Pinel for providing images of physical embodiments of a computational creativity system, and to V. K. Goyal, V. Misra, J. Z. Sun, and K. R. Varshney for further discussions.

**References**


