



CREATe Working Paper 2015/01 (February 2015)

# Monkeying Around with Copyright – Animals, AIs and Authorship in Law

---

## Authors

David Komuves  
University of Edinburgh  
s1268366@sms.ed.ac.uk

Jesus Niebla Zatarain  
University of Edinburgh  
j.niebla@ed.ac.uk

Burkhard Schafer  
University of Edinburgh  
b.schafer@ed.ac.uk

Laurence Diver  
University of Edinburgh  
Laurence.diver@ed.ac.uk

---

This paper was originally presented at the *Internationales Rechtsinformatik Symposium (IRIS)*, Salzburg, 26 – 28 February 2015. It won third place in the LexisNexis best paper award.

CREATe Working Paper Series DOI: [10.5281/zenodo.16113](https://doi.org/10.5281/zenodo.16113).

This release was supported by the RCUK funded *Centre for Copyright and New Business Models in the Creative Economy (CREATe)*, AHRC Grant Number AH/K000179/1.

# MONKEYING AROUND WITH COPYRIGHT – ANIMALS, AIs AND AUTHORSHIP IN LAW

David Komuves<sup>1</sup>, Jesus Niebla Zatarain<sup>2</sup>, Burkhard Schafer<sup>3</sup>,  
Laurence Diver<sup>4</sup>

<sup>1</sup> PhD Fellow, CREATE and University of Edinburgh, SCRIPT Centre for IT and IP Law  
Old College, EH8 9YL Edinburgh, UK  
s1268366@sms.ed.ac.uk; <http://www.create.ac.uk>

<sup>2</sup> PhD Researcher, University of Edinburgh, SCRIPT Centre for IT and IP Law  
Old College, EH8 9YL Edinburgh, UK

j.niebla@ed.ac.uk; <http://www.law.ed.ac.uk/research/students/viewstudent?ref=264>

<sup>3</sup> Professor of Computational Legal Theory and Director, SCRIPT Centre for IT and IP Law, University of Edinburgh  
Old College, EH8 9YL Edinburgh, UK

b.schafer@ed.ac.uk; <http://www.law.ed.ac.uk/people/burkhardschafer>

<sup>4</sup> Research Assistant, CREATE and University of Edinburgh, SCRIPT Centre for IT and IP Law  
Old College, EH8 9YL Edinburgh, UK

laurence.diver@ed.ac.uk; <http://www.law.ed.ac.uk/people/laurencediver>

**Keywords:** *copyright, artificial intelligence, computer art*

**Abstract:** *Advances in artificial intelligence have changed the ways in which computers create “original” work. Analogies that may have worked sufficiently well in the past, when the technology had few if any commercially viable applications, are now reaching the limit of their usefulness. This paper considers particularly radical thought experiment in relation to computer generated art, challenging the legal responses to computer generated works and discussing their similarity to works by animals.*

## 1. Hey, hey, we're the (thousand) Monkees

### 1.1. Daydream believers

In September of 2014, a number of technology blogs reported a potential “copyright apocalypse”.<sup>1</sup> A Russian company, Qentis, claimed to have found a way to use computer technology not just to create *some* new and original works – itself a considerable challenge – but *every* possible text that can be written (in a range of languages), every possible piece of music that can be composed, and a significant number of all paintings that it is possible to paint.<sup>2</sup> By using statistical and evolutionary algorithms that combine the smallest building blocks of language, individual letters, they claim to have

“generated and deployed 97.42% of all possible useful texts of ten to 400 words in length (the remaining 2.58% has already been deployed in the last 2000 years).

They state further that

“[...] Qentis is responsible for over 97 percent of all feasible text that can be created in English, German, French, Russian, Polish, Portuguese, Italian and Spanish. Qentis aims to

---

<sup>1</sup> See for example <https://torrentfreak.com/copyright-apocalypse-trolls-attack-the-net-from-the-future-140928/>, last accessed 21 January 2015.

<sup>2</sup> <http://www.qentis.com>, last accessed 21 January 2015.

create 99.2% percent of all target-length Internet text, making it by far the largest copyright holder in the world.”

The business model behind this idea is simple: becoming the world’s largest copyright troll. Or in their own words:

“The Qentis Corporation works with a powerful network of international law firms that represent our clients. The law firms notify authors, bloggers, news corporations, publishers and website owners whenever we feel they have breached the copyrights of our clients. As Qentis approaches 100 percent of content generation, all content owners will eventually have to pay royalties to our clients or face massive lawsuits.”<sup>3</sup>

If these claims were credible – and Qentis claims in particular to have generated the lyrics to Lady Gaga’s “Applause” four years before she did – then the consequences would be dramatic. No genuinely “new” works could be created in the future because everything that can possibly be expressed in the languages covered by Qentis’ technology has already been said. Nor is this approach restricted to text. Musical works follow essentially the same idea. Even more ambitious is Qentis’ claim to have generated images and even 3D objects. It claims that, since 2007, it has generated 3.23% of all possible images with dimensions up to 1000×800 pixels. By the end of 2020, its stated aim is to have generated all conceivable images.

The good news for artists, photographers, painters and writers everywhere, is that Qentis is “just” a work of art itself, created by the Vienna-based performance artist Michael Marcovici.<sup>4</sup> As with most of his work, it explores concepts of creativity and copyright through the lenses of technology. Despite its playfulness, the project asks some important questions about copyright, computer generated works and business models in the creative economy. It challenges our conceptions of creativity and the economic value that we attach to creative work, and exposes concerns about the logical limits of our legal vocabulary when attempting to conceptualize and resolve the legal tensions that arise from disruptive technologies. In what follows, we will briefly discuss the science behind “Qentis”. This will allow us to learn more general conceptual lessons about the interrelation between creative computers and copyright law. This discussion will involve an introduction to probability theory’s “infinite monkey theorem”, which underpins Qentis. There is perhaps only a short leap from these metaphorical monkeys to real AI “monkeys” causing the kind of copyright apocalypse Qentis points towards.

## 1.2. A barrel full of monkeys

Qentis is based on a simple idea: although the recursive nature of natural language allows for an infinite number of sentence constructions (Chomsky 2002), if we limit the permissible length of a piece of text then only a finite set of strings or texts can be generated. This idea has been famously generalised in the “infinite monkey theorem”, first used explicitly by the French mathematician Émile Borel (Borel 1913). The well-known thought experiment illustrates a special case of the Borel–Cantelli lemma in probability theory that is useful for the proof of the Law of Large Numbers: a thousand monkeys, hitting keys at random on a typewriter keyboard for an infinite amount of time, will “almost surely” produce the complete works of William Shakespeare at some point. In some variations of the theorem, larger numbers of monkeys are used (also approaching infinite monkeys), but a single monkey with infinite time is all that is needed. For illustration, if we assume that a typewriter has 50 keys, and that every one of them has the same chance of being pressed by the monkey, then after 35,977,876,623 hits ( $35,977,876,618 + 5$ ), there will be a 90% chance that the monkey has typed the word “Hamlet”. The longer the monkey types, the higher the

---

<sup>3</sup> <http://www.qentis.com/work/work-13/>, last accessed 21 January 2015.

<sup>4</sup> <http://www.artmarcovici.com/BIOGRAPHY>, last accessed 21 January 2015.

probability becomes, approaching 1 in the case of an infinite number of keystrokes. Paradoxically, this does not mean that it is impossible for the monkey *not* to type “Hamlet”, even with infinite time – it could, for example, hit the letter “g” for all eternity. But the nature of infinity is such that the probability of it eventually typing “Hamlet” is nevertheless “almost surely”. The same holds true for any string of letters of finite length, including the entirety of the play Hamlet. From a legal perspective, this also means that the owner of the monkey’s output could prove, for probabilistic reasons – that is, not just on the balance of probabilities but beyond a reasonable doubt – that a copy of Hamlet is in his possession, without any need to actually check the output of the typing simians. This idea is key to Qentis’ “business model”.

The infinite monkey theorem has influenced the artistic imagination more than most mathematical theorems. In “The Total Library”, Jorge Luis Borges followed the history of the argument, tracing it back to Aristotle’s critique of the atomism of Leucippus and his notion that the world is the result of random combination of atoms. We note in passing that Aristotle presents this as a reduction *ad absurdum*: because “brute force” combination of constitutive elements (the atoms) is so unlikely to create any meaningful result, the infinite amount of time that is needed but is not available in the physical world means that, for him, it is extremely unlikely that the world arose in this way. In a similar vein, Cicero argued 300 years later in *De natura deorum* that

“He who believes this may as well believe that if a great quantity of the one-and-twenty letters, composed either of gold or any other matter, were thrown upon the ground, they would fall into such order as legibly to form the Annals of Ennius. I doubt whether fortune could make a single verse of them.” (Cicero 1961)

Borges notes that by the 19<sup>th</sup> century the emphasis and vocabulary of the argument had changed. From essentially a negative statement about *impossible* outcomes, it became a statement about *likely* results. Falsely attributing the idea to Huxley and his defence of Darwin, he states that instead of throwing a golden letter once, for “Darwin’s Bulldog”, “a half-dozen monkeys provided with typewriters would, in a few eternities, produce all the books in the British Museum” (Borges 2007, p. 215). This idea inspired Borges to imagine a “Total Library” (Borges 2007, p. 219):

“Everything would be in its blind volumes. Everything: the detailed history of the future, Aeschylus’ The Egyptians, the exact number of times that the waters of the Ganges have reflected the flight of a falcon, the secret and true nature of Rome, my dreams and half-dreams at dawn on August 14, 1934, the proof of Pierre Fermat’s theorem, [...], [...] Everything: but for every sensible line or accurate fact there would be millions of meaningless cacophonies, verbal farragoes, and babblings. Everything: but all the generations of mankind could pass before the dizzying shelves — shelves that obliterate the day and on which chaos lies — ever reward them with a tolerable page.”

Borges returns to this idea in his short story “The Library of Babel”. In this novel, a vast library is made from interlocking hexagonal rooms, each of which has four walls with books and the bare amenities a human needs to survive. According to its inhabitants, it contains every possible volume that could be composed from the letters of the alphabet. We do not learn in this novel if monkeys are responsible for the books, only that “A blasphemous sect suggested [...] that all men should juggle letters and symbols until they constructed, by an improbable gift of chance, these canonical books”. The outcome however is the same – the majority of the books are just meaningless strings of letters. They are not ordered in any logical fashion but are seemingly assigned to rooms randomly. Thus, even if the library *necessarily* contains all useful information in existence, including predictions of the future, the sheer amount of unstructured information means that it remains useless for the readers, leaving the librarians in a state of despair.

Qentis is a modern day version of the library of Babel, the monkeys replaced by parallel computing processors. Unlike Borel’s use of the infinite monkey theorem, however, rather than its core aim

being to find an *existing* work, the goal is instead to create all possible works that have not yet been written. As we shall see below, some of the shortcomings that the users of Borges' library experience are solved by Qentis through a business model that illustrates the perversion of copyright when it is separated from any meaningful use of the creative works. Or put differently: even though the library is useless for readers, copyright law gives it commercial value.

While the infinite monkey theorem originates in mathematics and centres around the mathematical concept of infinity, many researchers have speculated if it is possible to construct finite but physically realistic models. Richard Dawkins, for example, uses a similar idea in *The Blind Watchmaker* to illustrate the power of evolutionary processes. Dawkins' "Methinks it is a weasel" program starts from a randomly typed string of symbols, the "parent". New "generations" are created by replacing letters randomly, keeping only those letters that match the "methinks" target phrase from Hamlet.

Dawkins' approach remains a purely abstract realisation of the monkey theorem for finite applications. Unlike Qentis, it too requires "selection" towards a predefined goal, which speeds up the process of generation of meaningful strings by several orders of magnitude. A very similar approach by Jesse Adamson, which uses Amazon's EC2 cloud computing system to increase the number of monkeys into the millions, has already typed up the whole of *A Lover's Complaint*.<sup>5</sup> The system generates random strings of nine characters which are then matched against Shakespeare's oeuvre and kept when match is found. Both Dawkins' and Adamson's approaches thus require human intervention and ingenuity to model the "information creating" aspect of natural selection. In this crucial respect they differ from Qentis' "brute force" approach. Below we discuss the copyright status of the works these approaches create.

An implementation closer to the original idea is "The Monkey Shakespeare Simulator" described in Wershler-Henry's *The Iron Whim*. This software simulation started with 100 simulated monkey employees. In the simulation, time passes more than 86,000 times faster than in the real world (Wershler-Henry 2005, p. 192). In addition, between typing (at 60 letters per minute), the monkeys procreate, thus steadily adding new employees to the pool of typists. In 2004, one simulation which was running for the equivalent of 42,162,500,000 billion billion monkey years, produced the 19 letter string "valentine cease to" that matches, reasonably closely, the sentence "Valentine: cease to persuade, my loving Proteus".

Finally, we mention an attempt to implement the theorem not with figurative computer simulations of monkeys, but with real simians. In 2002, six Sulawesi macaque monkeys contributed to the exhibition GENERATOR by the University of Plymouth's MediaLab Arts course. The monkeys produced five pages consisting mainly of the letter "s" before they started destroying the typewriters with a stone and using it as toilet.<sup>6</sup>

Where infinitely many, thousands of, or six, monkeys failed, one has succeeded. In the same year in which Qentis made its "announcement" to commercialise the infinite monkey theorem, a solitary monkey photographer stimulated the legal imagination of copyright lawyers. An Indonesian black macaque monkey grabbed the camera of a British wildlife photographer and took several pictures of itself – some of them of quite astonishing quality. When one of the photos was posted on Wikimedia, the photographer threatened to sue for copyright infringement.<sup>7</sup> But is the owner of the

---

<sup>5</sup> <http://www.jesse-anderson.com/2011/09/a-few-million-monkeys-randomly-recreate-shakespeare/>, last accessed 21 January 2015.

<sup>6</sup> The outcome and images of the monkeys at work can be seen at [https://web.archive.org/web/20130120215600/http://www.vivaria.net/experiments/notes/publication/NOTES\\_EN.pdf](https://web.archive.org/web/20130120215600/http://www.vivaria.net/experiments/notes/publication/NOTES_EN.pdf), last accessed 21 January 2015.

<sup>7</sup> <http://www.create.ac.uk/blog/2014/08/07/quit-playing-around-monkey-stirs-up-copyright-controversy-with-selfie-guestpost-by-emily-goodhand/>, last accessed 21 January 2015.

camera the author, and with that the copyright holder? Or are “works without human author” outside the scope of copyright law? In the discussion of the monkey “selfie”, the analogy to computer generated works was frequently made. As our discussion shows, the similarities between the two modes of production are quite strong; stronger even than commentators realised. But can we distinguish for copyright purposes the infinite monkeys simulated on a computer from a single, real monkey? Can a case be made to extend copyright law to either, or both, if not as a principle of existing law, then of some future copyright law? Qentis can be read as a dire warning against pursuing such a route.

## 2. Monkey business

It should be clear why Qentis’ idea does not, for *technological* reasons, pose a feasible challenge to our copyright regime. Borel used the infinite monkey theorem to prove what is *not* possible within the lifetime of this universe as an indirect proof of the laws of thermodynamics. Similarly, it would need several million times the life span of our universe to make even a dent in the number of possible permutations of letters. For images, the number is even higher – calculations show that to generate even a small subset of the images Qentis aim to create by 2020, it would take  $3.13 \times 10^{240800}$  lifetimes of the universe so far. For Qentis, this fact is arguably a blessing in disguise, because not only would its software create every possible work that has commercial value, it would also create every possible defamatory statement, every possible image of child pornography, and every possible incitement to violence. But if we assume for argument’s sake that brute force methods could generate new works, is Qentis’ business model sound, and would copyright law promote its economic success?

In Borges’ novel, even though they have all the library at their disposal, including all correct future predictions, its inhabitants fail to benefit from the works in it – either because they cannot find the ones they need, or they cannot determine which of the texts is a truthful account of the external world. Both problems would also arise for Qentis, if its business model were to produce content that someone actually wants to read or find useful. Since the overwhelming majority of the works created by their algorithm would be unintelligible, it would take many lifetimes for a user to find the right text for a given task, even if we could know *a priori* that it must be somewhere in the collection. But Qentis’ business model is different. It does not claim to produce useful works. Rather, all the computer generated work is sold in bulk to commercial outfits aiming to use it for solely for copyright litigation. They need not check so much as a single document in their collection: probability theory means that a given work is “almost surely” in their collection, and this is all that is needed to discharge the civil burden of proof that requires them to have acquired the rights to that work. Qentis is the platonic ideal of the copyright troll: a right holder who cannot even in theory access the work they own, let alone use it for a meaningful propose, but who nevertheless can use its knowledge of its ownership to extract monetary benefit from its purely passive ownership.<sup>8</sup> This is the political message the Qentis thought experiment delivers: copyright is now so far removed from the notion of societal usefulness that, at least in principle, something like Qentis has become conceivable.

A greater problem for Qentis is directly related to copyright law. Even if Qentis were to own the copyright in all possible future works, this alone does not of course mean that they could prevent people from writing and selling the product of their own creative work. Unlike patent law, copyright law permits parallel creation, so in addition to have the copyright in all possible works, they would have to show in each individual instance that the other author actively copied from their database. But as Borges’ story teaches us, it is not possible in the library of Babel to find any

---

<sup>8</sup> For more messy real life instantiations of this platonic ideal of copyright trolling as a business strategy for law firms, see for example *DeBriyn* (2012) or *Curran*, (2013–2014).

specific work. The noise always drowns out the signal. For the infinite monkey theorem, we can rigorously prove this using probability theory. Thus, just as pure mathematics tells us *a priori* and beyond reasonable doubt that Qentis has indeed created a copy of the specific work that is the subject of litigation, the defendant can also prove *a priori*, for the same reasons and beyond a reasonable doubt, that he never saw *this* work in Qentis' database, since it would take many times the lifetime of the universe to locate it there.

While these two obstacles are straightforward and apply uniformly in all current copyright systems, a more complicated question arises around whether Qentis would have any copyright in the works that its software generates. In the paradigmatic case of copyright protected work, a human, the author, takes an abstract idea or thought (not protected by copyright) and transforms it into a concrete, embodied form. On the one hand, the transition from idea to concrete, syntactic embodiment is closely linked with the concept of creativity, and on the other it requires the type of effort and investment that copyright law also aims to reward. These key components are not present in the case of Qentis. It lacks a human author, and the way in which it generates work means that the concept of an "idea" is absent. Instead, it deals only in expressions, but expressions of what? This objection against the very concept of AI has been raised in particularly poignant form by the philosopher John Searle in his famous "Chinese Room" thought experiment. His argument, in a nutshell, is that the mere simulation of the knowledge of Chinese by someone who does not speak the language but simply replaces one set of (for him) unintelligible scribbles with another set according to a predefined set of rules, is the best an artificial intelligence can ever be expected to achieve. Computers lack intentionality and thus they do not truly communicate; they merely "ape" communication (Searle 1980). Long before Searle, and directly referencing the infinite monkey theorem, the historian and philosopher Robin Collingwood drew inferences for the theory of literature. For him, anything generated by mere random processes, even if it looks exactly like a text, should not be considered an artistic work: there is an ontological difference between "the works of Shakespeare" – which for Collingwood are an abstract, imaginary object closely tied to the idea of emotional expression – and a particular physical embodiment of letters on a medium. He attacks, in scathing terms, those who think random processes can generate works of art (Collingwood 1958, p. 126, footnote 1):

"But the interest of the suggestion lies in the revelation of the mental state of the person who can identify the 'work' of Shakespeare with the series of letters printed on the pages of a book bearing that phrase as its title: and thinks, if he can be said to think at all, that an archaeologist of 10,000 years hence, recovering a complete text of Shakespeare from the sands of Egypt but unable to read a single word of English would possess Shakespeare's dramatic and poetic works."

Note that Collingwood does not reject the idea that monkeys could, physically, generate a text that is letter by letter identical to Shakespeare's – rather, his claim is that this would nevertheless not qualify as "Shakespeare's works" because of a lack of intentionality in its creation. Most legal systems so far embrace this line of argument. There cannot be a copyright work without a (human) author, and no protected expression without a concomitant idea that it embodies. Consequently, copyright for computer generated works has generally been rejected.

The US Copyright Office states:<sup>9</sup> "[a]s discussed in Section 306, the Copyright Act protects "original works of authorship, 17 U.S.C. § 102(a)". To qualify as a work of "authorship" a work must be created by a human being. It states further that "[t]he Office will not register works produced by nature, animals, or plants" and gives as examples "a photograph taken by a monkey" or "a mural painted by an elephant". The guidance continues by saying the Office "will not register works produced by a machine or mere mechanical process that operates randomly or automatically

---

<sup>9</sup> <http://copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf>, last accessed 21 January 2015.

without any creative input or intervention from a human author”. It gives as an example “[a] claim based on a mechanical weaving process that randomly produces irregular shapes in the fabric without any discernible pattern”.

At least for US law, therefore, the news is bad for both infinite and individual monkeys. We note however that the example given by the US Copyright Office mentions outputs that “lack discernible patterns”. However, the infinite monkey theorem states that random processes do produce discernible patterns, and Dawkins’ application in particular aims to dispel the lay perception that random processes are incapable of creating order. Furthermore, his implementation of the theorem does involve human intervention to add an element of “natural selection” to the process. Is this sufficient to argue that this is not a process without *any* human intervention?

The situation is clearer in the UK. It is one of the few legal systems that explicitly created a norm that protects computer generated art (for a comparative discussion see McCutcheon 2013). Section 9(3) of the Copyright, Designs and Patents Act 1988 (c. 48) specifies that “[i]n the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.” Thus, despite the other obstacles discussed above, Qentis would be the copyright holder under UK law. The sole simian from Sumatra fares less well. In an interview about the story, the Intellectual Property Office stated that animals cannot own copyright, although the possibility was left open that a human who significantly contributes to such a work could claim it.<sup>10</sup> This seems a counter-intuitive result: why should works created by simulated monkeys be treated more favourably than those created by a real monkey, especially when we consider that biologically inspired computing is moving towards the creation of computers that match exactly in structure and substance that of organic beings?<sup>11</sup> Why should a work generated by cells taken from a monkey brain encased in a machine be treated differently from a work created by identical cells “in situ”, especially given the observation in the above-mentioned University of Plymouth experiment that the monkeys demonstrated creative processes far closer to human behaviour than computer behaviour?

### **3. Conclusion: throwing a monkey wrench into copyright’s machinery**

The idea of computer creativity and computer generated works is not new (see for example Gelernter 1994 and Shank and Cleary 1994). Questions regarding the copyright status of such works are almost as old as the first prototypes of computer creativity (for a discussion see Bridy 2012). What has changed over recent years, however, is that we now have viable business models that are able to utilise computer generated works. While earlier works aimed at “high art”, contemporary applications focus on lesser examples of human creativity; the “small coinage” of German IP law. Short, technical articles and notes for online publication, data-driven journalism and summarisation services are most likely to avail themselves of this technology. A typical application could harvest customer reviews about their holiday in city X from the Internet and rewrite the information into a Wikipedia entry on X, or take business data and statistics and turn these into a report for shareholders. Services such as *Narrative Science*<sup>12</sup> or *Automated Insight*<sup>13</sup> focus on this segment of the market (for a scientific discussion see, for example, Lee *et al* 2012). Qentis’ most realistic aspect is its restriction to texts of 400 words or fewer – it is indeed this size of article that is most likely to become generable by computers in the near future. This technology

---

<sup>10</sup> <http://www.theguardian.com/technology/2014/aug/22/monkey-business-macaque-selfie-cant-be-copyrighted-say-us-and-uk>, last accessed 21 January 2015.

<sup>11</sup> In, for example, the field of DNA-based computing. See for example Kahan *et al* 2008.

<sup>12</sup> [http://www.slate.com/articles/technology/future\\_tense/2012/03/narrative\\_science\\_robot\\_journalists\\_customized\\_news\\_and\\_the\\_danger\\_to\\_civil\\_discourse\\_.single.html](http://www.slate.com/articles/technology/future_tense/2012/03/narrative_science_robot_journalists_customized_news_and_the_danger_to_civil_discourse_.single.html), last accessed 21 January 2015.

<sup>13</sup> <http://towcenter.org/blog/automated-stories-using-algorithms-to-craft-news-content/>, last accessed 21 January 2015.



threatens established business models in the creative economy and will devalue certain forms of human creativity.<sup>14</sup> It also disrupts the legal regulatory machinery. As we have seen, core concepts of copyright law fail to express adequately the issues that are at stake. The focus in the academic debate has been on the concept of “author”, but our discussion indicates that the “idea vs expression” dichotomy is at least as problematic.

Unlike Qentis, these systems will not simply generate random texts, but learn from and incorporate text written by others. The need for such an approach was also recognized in the analysis of the infinite monkey theorem. Qentis’ n-gram approach fails for the same reason that Dawkins’ monkeys fail to be a proper analogue to evolution. In either case, only one letter at a time is typed/changed, independently of the other letters and without an evaluation of past experience. Hugh Petrie argues that, similarly, the evolution of written ideas requires to follow biological evolution in accounting for this historical context, and argues that we should equip the monkey with not just a typewriter, but what we would today call an expert system that incorporates “whole Elizabethan sentences and thoughts. It would have to include Elizabethan beliefs about human action patterns and the causes, Elizabethan morality and science, and linguistic patterns for expressing these” (Petrie 1981, p.132). This approach comes much closer to what working text generation systems attempt. They combine rules of composition distilled from past experience with texts and text fragments written by others. From a copyright perspective, we therefore face not just one but two questions: who owns the IP in the computer generated work and were they permitted to use the work of others to generate it? The second question does not apply to Qentis’ probabilistic *ex nihilo* creation, but it is also not a straightforward question of impermissible copying: the input texts are not reproduced in any recognizable form; rather, their logical structure is analyzed and through purely mechanical ways reconstructed beyond recognition. The process is similar to that of a human reader who grasps the idea underlying a text and then expresses it entirely in her own words – an unproblematic and legally permissible process. But the distinction between idea and expression, *per* Searle or Collingwood, does not apply to computers, demonstrating how these technologies disrupt established regulatory ideas.

Secondly, we remember how Qentis fails as a business because the readable works it generates are drowned out by excess noise. Dawkins addresses this in his simulation of the infinite monkey theorem by introducing a pre-defined goal. Applications of computer generated text will also typically require a user to define permissible outcomes and to act as the equivalent of “natural selection”. This insight allows us to think of the reader as a co-creator of computer generated works, an idea proposed from the perspective of literary theory in Garcia’s discussion of the infinite monkey theorem (Garcia 1996, pp. 122–125). Selecting goals, finding the right answer amongst the noise and acting as “natural selection” may well fulfil the minimum requirement of “creative human input” required by US law, though the type of creativity is different from a traditional writer and includes creative search strategies that are absent in Qentis.

In conclusion: automated text generation has matured to the level of commercial viability. As a disruptive technology it poses dangers for some creative writers. The current legal system and its vocabulary is ill-suited both to the protection of the legitimate interests of human writers on whose work these machines still rely, and to encouraging the creation of efficient business models that use these technologies. UK law potentially leads the way by recognizing IP in computer generated work, but by framing the issue in traditional copyright vocabulary it fails to express how the “input” for these machines ought to be classified in law. At the same time, “brute force” approaches such as Qentis illustrate potential dangers if we fail adequately to regulate these technologies. Where other legal systems may prematurely block computer generated works, UK law arguably brings us too close to a Qentis-style copyright apocalypse. What therefore is needed is a *sui generis* concept for

---

<sup>14</sup> <http://www.cbsnews.com/news/this-post-was-written-by-a-human/>, last accessed 21 January 2015.

copyright in computer generated works that is expressed independently of the traditional author/reader and expression/idea terminology and is at the same time sensitive to the very different technologies that computer generated works can utilise. This would allow for a proper grading of protection according to the creative approach chosen, recognizing that computer generated works typically require and deserve much less protection than human generated work. Our biological cousins, and other members of the animal world, are likely to be the indirect beneficiaries of such an approach that can accommodate their creativity.

## 4. References

*Borel, Émile*, Mécanique Statistique et Irréversibilité. *Journal Phys.* 5e série, 3, (1913), pp 189–196.

*Borges, Jorge Luis*, *The Total Library*, Penguin (2007).

*Bridy, Annemarie*, Coding Creativity: Copyright and the Artificially Intelligent Author, *Stan. Tech. L. Rev.* (2012), pp 1–28.

*Chomsky, Noam*, *Syntactic structures*. Walter de Gruyter, (2002).

*Cicero, Marcus Tullius*, *De natura deorum*, *Academica*. Vol. 268. Harvard University Press, (1961).

*Collingwood, Robin George*, *The Principles of Art*, Galaxy Books (1958).

*Curran, Luke S.*, Copyright Trolls, Defining the Line between Legal Ransom Letters and Defending Digital Rights: Turning Piracy into a Business Model or Protecting Creative from Internet Lawlessness *Marshall Rev. Intell. Prop. L.* [v] (2013–2014), pp 170–202.

*DeBriyn, James*, Shedding Light on Copyright Trolls: An Analysis of Mass Copyright Litigation in the Age of Statutory Damages, *UCLA Entertainment Law Review* 19.1 (2012).

*Gracia, Jorge*, *Texts: Ontological Status, Identity, Author, Audience*. SUNY Press (1996).

*Gelernter, David*. *The muse in the machine: Computerizing the poetry of human thought*. Simon and Schuster (1994).

*McCutcheon, Jani*, Curing the Authorless Void: Protecting Computer-Generated Works Following IceTV and Phone Directories. *Melbourne ULR* 37 (2013), pp 46–232.

*Petrie, Hugh G.*, *The dilemma of enquiry and learning*. Living Control Systems Publ (2011).

*Schank, Roger C., Cleary, Chip*, Making Machines Creative. In: S Smith, T B Ward & R A Finke (eds.) *The Creative Cognition Approach*. MIT Press (1995), pp 229–247.

*Searle, John*, Minds, Brains and Programs, *Behavioral and Brain Sciences* 3 (3) (1980), pp 417–457.

*Wershler-Henry, Darren Sean*, *The iron whim: A fragmented history of typewriting*. Cornell University Press (2005).



RCUK Centre for Copyright and  
New Business Models in the  
Creative Economy

College of Social Sciences / School of Law  
University of Glasgow  
10 The Square  
Glasgow G12 8QQ  
Web: [www.create.ac.uk](http://www.create.ac.uk)

