

Forthcoming book chapter in in Abbott R (ed), *Research Handbook on Intellectual Property and Artificial Intelligence* (Edward Elgar, 2022)

CAN ARTIFICIAL INTELLIGENCE INFRINGE COPYRIGHT? SOME REFLECTIONS

Enrico Bonadio, Plamen Dinev, Luke McDonagh*

* Dr Enrico Bonadio is Reader in Law City, University of London; Dr Plamen Dinev is Lecturer in Law at Goldsmiths, University of London; Dr Luke McDonagh is Assistant Professor at LSE Law, London School of Economics.

INTRODUCTION

Artificial intelligence (AI) encompasses a range of computer algorithms which mimic the human mind—behaviour that would be regarded as ‘intelligent’ if performed by a human.¹ AI powers self-driving cars; Tesla’s conceptual humanoid robot Optimus (designed to help humans with ‘unsafe, repetitive or boring’ tasks and deadlift up to 150 lbs);² multilingual neural machine translation services such as Google Translate; and DeepMind’s AlphaGo—which recently forced a former world champion to retire from professional play, after declaring AI invincible: ‘[e]ven if I become the number one, there is an entity that cannot be defeated.’³ AI is also capable of generating various *creative* outputs—works traditionally protected by copyright. OpenAI’s GPT-3 language model, for instance, can write poems and other forms of literature.⁴ Other creative AI systems can produce music and visual art,⁵ often attracting considerable media attention; the *Portrait of Edmond Belamy* sold for US\$432,500 through a Christie’s auction in 2018.⁶ In addition to creative works, AI is now commonly utilised to generate inventions essential to products (which may be patentable), ranging from kitchen appliances to drug synthesisers and other more sophisticated inventive outputs.⁷

¹ See Russell S and Norvig P, *Artificial Intelligence: A Modern Approach* (Pearson, 2016) at p. 2 for a full range of definitions of “AI” (dividing these into four broad categories, including thinking humanly; acting humanly; thinking rationally; and acting rationally).

² See Tesla, ‘Artificial Intelligence & Autopilot’ <https://www.tesla.com/en_GB/AI> (accessed 1 March 2022); see also Maynard A, ‘Elon Musk’s Tesla Bot raises serious concerns – but probably not the ones you think’ (2021) <theconversation.com/elon-musks-tesla-bot-raises-serious-concerns-but-probably-not-the-ones-you-think-166714> accessed 1 March 2022 (‘[H]ow responsible is Musk’s vision? Just because he can work toward creating the future of his dreams, who’s to say that he should? Is the future that Musk is striving to bring about the best one for humankind, or even a good one? And who will suffer the consequences if things go wrong?’).

³ Vincent J, ‘Former Go champion beaten by DeepMind retires’ <theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat> accessed 28 July 2020.

⁴ Asnen A, ‘Dear Science’ (Medium, 25 October 2021) <medium.com/the-bad-influence/dear-science-4e1c549e4f80> accessed 1 March 2022 (GPT-3 when asked to write an essay on the future of humanity (‘[t]here was a time when the future was certain. That time is now reaching its conclusion... We are on the brink of a technological revolution that has the potential to eradicate human suffering while simultaneously bringing an end to our existence as a species’); Tang D, ‘The Machines Are Coming, and They Write Really Bad Poetry’ <lithub.com/the-machines-are-coming-and-they-write-really-bad-poetry> accessed 28 July 2020.

⁵ Baraniuk C, ‘Computer Paints ‘New Rembrandt’ after Old Works Analysis’ (6 April 2016), available at <<https://www.bbc.com/news/technology-35977315>> accessed 18 January 2022.

⁶ See Christie’s, ‘Is artificial intelligence set to become art’s next medium?’ (2018) <<https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>> accessed 18 January 2022; for a detailed discussion on the protection of AI-generated works, see Bonadio E and McDonagh L, ‘Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity’ (2020) *Intellectual Property Quarterly*, 2, pp. 112-137.

⁷ Yanisky-Ravid S & Liu X, ‘When Artificial Intelligence Systems Produce Inventions: The 3A Era and an Alternative Model for Patent Law’ (2017) 39 *Cardozo Law Review* 2215, p. 2219; see also, more broadly, Hartmann C et al, ‘Trends and Developments in Artificial Intelligence Communications Networks, Content and Technology Final Report Prepared by: Challenges to the Intellectual Property Rights Framework’ (25 November 2020) <<https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>> accessed 8 December 2021 (examining, inter alia, the issue of IP protection for

These examples clearly raise a wide range of intellectual property (IP) issues, ranging from questions of copyright subsistence and core patentability requirements to issues concerning IP theory more broadly. Rather than examining all pertinent IP issues, this chapter will focus on copyright and machine-induced *infringement* in particular. AI is capable of consuming large amounts of creative works and data as part of its learning process, which clearly raises serious risks of infringement. At the same time, this is an area which has not been fully explored yet (as most academic works have focused on the issue of protecting AI-generated works). In terms of scope, we will examine key jurisdictions including the UK, EU and US which are leading the academic debate (and scientific research) in this area. Before looking at the critical issues of infringement and exceptions to infringement, we will first provide a brief overview of AI creativity, which will help to explain why and how infringement may occur.

CAN AI INFRINGE COPYRIGHT?

Creative AI uses substantial amounts of input data—images, videos, text and other artistic content—as part of its learning process.⁸ Music-generating AI, for instance, utilises significant amounts of source material to find patterns and create new melodies based on various elements including tempo, chords and length.⁹ Similar rules apply in the context of visual art—*The Next Rembrandt* project involved 350 scanned images and over 150 gigabytes of data.¹⁰ While the final output was not universally acclaimed (labelled by some as ‘fan-fiction’),¹¹ it clearly demonstrated that modern AI can produce sophisticated creative output that resembles the works of professional (human) artists. Large amounts of source text are required to generate literature and creative writing too. Deep learning language models such as GPT-3—AI that produces human-like text across a range of categories, including creative writing, parodies and

AI-assisted output in the fields of science (focusing on meteorology), media (journalism) and pharmaceutical research); see also University of Surrey, ‘World’s first patent awarded for an invention made by an AI could have seismic implications on IP law’ <<https://www.surrey.ac.uk/news/worlds-first-patent-awarded-invention-made-ai-could-have-seismic-implications-ip-law>> accessed 1 March 2022 (patent officials in South Africa have granted a patent that names AI as the inventor); see also Bonadio E, McDonagh L and Dinev P, ‘Artificial Intelligence as Inventor: Exploring the Consequences for Patent Law’ (2021) *Intellectual Property Quarterly* 1 48-66, p. 48.

⁸ Sobel B, ‘Artificial Intelligence’s Fair Use Crisis’ (2017) 41 *Colum. J.L. & Arts* 45, p.1.

⁹ Deahl D, ‘How AI-Generate Music is Changing the Way Hits are Made’ (2019), <theverge.com/2018/8/31/17777008/artificial-intelligence-taryn-southern-amper-music> accessed 19 February 2022.

¹⁰ Yanisky-Ravid S, ‘Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era’ (2017) *Mich. St. Law Review* 659, 669.

¹¹ Schjeldahl P, ‘A Few Words About the Faux Rembrandt’ (8 April 2016), *The New Yorker*, <<https://www.newyorker.com/culture/culture-desk/a-few-words-about-the-faux-rembrandt>> accessed 19 February 2022.

storytelling—are now used in hundreds of different apps.¹²

The key point here is that creative AI cannot function without source material. It needs to learn from existing works, many of which could be protected by copyright owned by another party. This inevitably raises the risk of infringement, both in relation to the AI's inputs and outputs.¹³ Feeding source material (inputs) into the AI and processing this data may violate the right to reproduction.¹⁴ Likewise, the final product (outputs) could be regarded as an adaptation of pre-existing works.¹⁵ With regards to outputs, however, any finding of infringement will depend on whether pre-existing elements can be recognised in the final product. While output which has been subject to substantial change may escape infringement, works that contain clearly identifiable elements are likely to violate the adaptation right.¹⁶

EXEMPTING AI INFRINGEMENT

Having briefly outlined the various circumstances where the use of creative AI may constitute infringement, we need to consider the key question of whether such use may be exempt from liability. We will first look at the fair use doctrine in the US before moving to UK and EU law.

The US 'fair use' doctrine and AI—expressive vs non-expressive use

While certain mechanical (non-expressive) uses of protected material may be exempt from infringement under the 'fair use' doctrine in US law, this exception will generally not be available where the use 'conveys expression', i.e. where the final output is artistic and creative in nature.¹⁷ There is little jurisprudence specifically dealing with AI as such, though existing principles from cases such as *Kelly*¹⁸ and *Perfect10*¹⁹ (both concerning the use of image search

¹² See OpenAI, 'GPT-3 Powers the Next Generation of Apps' (25 March 2021) <<https://openai.com/blog/gpt-3-apps>> accessed 20 December 2021; see also Branwen G, 'GPT-3 Creative Fiction' (1 July 2021) <<https://www.gwern.net/GPT-3>> accessed 20 December 2021; see also Aalho J, 'I Wrote a Book with GPT-3 AI in 24 Hours — And Got It Published' (12 June 2021) <<https://medium.com/swlh/i-wrote-a-book-with-gpt-3-ai-in-24-hours-and-got-it-published-93cf3c96f120>> accessed 20 December 2021.

¹³ See also Dee C, 'Examining Copyright Protection of AI-Generated Art' (2018) 1 *Delphi – Interdisciplinary Review of Emerging Technologies* 31, 36.

¹⁴ See e.g. s. 17(2) UK CDPA 1988 (infringement includes reproducing, inter alia, a literary, artistic or musical work in any material form, and other European jurisdictions offer similar provisions).

¹⁵ Sobel B, 'Artificial Intelligence's Fair Use Crisis' (2017) 41 *Colum. J.L. & Arts* 45, p. 16; see also e.g. s. 21(1) CDPA ('The making of an adaptation of the work is an act restricted by the copyright in a literary, dramatic or musical work').

¹⁶ Deltorn J, 'Deep Creations: Intellectual Property and the Automata' (2017), <<https://www.frontiersin.org/articles/242911>> accessed 19 January 2022.

¹⁷ Sobel B, 'Artificial Intelligence's Fair Use Crisis' (2017) 41 *Colum. J.L. & Arts* 45; see also Lim D, 'AI & IP: Innovation & Creativity in an Age of Accelerated Change' (2018) 52 *Akron Law Review* 813.

¹⁸ *Kelly v. Arriba and Perfect10 v. Amazon (Kelly v. Arriba Soft Corp)*, 336 F.3d 811 (9th Cir. 2003).

¹⁹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007).

engines) could certainly be applied in this context. In both cases, the plaintiffs owned copyright in various images which had been copied as thumbnails. They were hosted on Arriba and Amazon's (the defendants) servers and then made available to their customers. As the search engines acted as mere 'tools', the court found that there was no expression being conveyed here. Therefore, this use was regarded as transformative and 'fair'.²⁰

*Authors Guild*²¹ concerned Google Books—a popular service based on scanning text and converting it via optical character recognition, a process which allows end users to search scanned collections and quickly locate relevant information. As the process behind the service inevitable requires copying, the court had to consider whether this use could be exempt from infringement. It found in favour of Google, arguing that reproducing text with the aim of merely making it more easily searchable was transformative and thus 'fair'. A key consideration here was the fact that the service does not offer a substitute for the actual books as it does not allow users to read them in full; it merely facilitates their search. Similarly, in *Vanderhye*,²² the court found that *Turnitin*—a plagiarism-detection service based on scanning submissions against a large database of existing text and materials—could rely on this exception as the creative value of the scanned papers is irrelevant here.²³ What is more, despite the fact that *Turnitin* makes copies of documents in their entirety, these full reproductions are not available to end users who merely see snippets of specific data relevant to plagiarism.

The above cases thus indicate that certain non-expressive uses by AI may be regarded as fair use. Consider the example of training a facial recognition system (matching a face from an image against a large database), a process which typically involves making copies of substantial amounts of (protected) images. What is actually used by the AI here does *not* concern the artistic merit of the photos or the expressive choices made by the artist. This technical process is solely focused on comparing facts about one's personal identity with facts about their appearance.²⁴ These images are reproduced for the sole (non-artistic and non-

²⁰ See *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

²¹ *Authors Guild, Inc. v. Google, Inc.*, 954 F. Supp. 2d 282 (SDNY 2013); *Authors Guild v Google Inc.*, 804 F (3d) 202 (2nd Cir. 2015).

²² *A.V. ex rel. Vanderhye v. iParadigms, L.L.C.*, 562 F.3d 630 (4th Cir. 2009).

²³ Sartor G et al, 'The Use of Copyrighted Works by AI Systems: Art Works in the Data Mill' (2018) available at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3264742> accessed 20 December 2021.

²⁴ See Sobel's interview with Ipwatch, Sobel B, 'The Dilemma Of Fair Use And Expressive Machine Learning: An Interview With Ben Sobel' (23 August 2017) available at <<https://www.ip-watch.org/2017/08/23/dilemma-fair-use-expressive-machine-learning-interview-ben-sobel/>> accessed 20 December 2021. The same would apply in the context of a caretaker robot who helps a blind person find an exhibition, taking photos inside the building and taking her to specific paintings based on her own preferences; see Schafer B et al, 'A Fourth Law of Robotics? Copyright and the Law and Ethics of Machine Co-Production' (2015) 23 *Artificial Intelligence and*

expressive) purpose of pattern recognition. Likewise, using an AI algorithm which goes through large volumes of computer folders containing copyright music for the purposes of organising and sorting albums by genre is likely to be regarded as non-expressive and thus fair.

In contrast, the exception is unlikely to be available where the AI consumes existing copyright works for the purposes of making new creative works, i.e. where the purpose is clearly *expressive*. This would be applicable in cases such as the Next Rembrandt project discussed above²⁵ and the *Portrait of Edmond Belamy* (15,000 portraits painted over seven centuries were fed into the system)²⁶ where the specific aim was to generate artistic output.²⁷ Given that creative AI—by definition—aims to convey expression, it is therefore unlikely to benefit from ‘fair use’ in the vast majority of cases. But this also raises some important normative questions. Do we need to distinguish between human- and machine-induced infringement in this context?

Special treatment for AI?

If AI owners or users are to claim copyright ownership over new creative works generated by an algorithm, it is certainly arguable that they should also be held accountable where the AI commits infringement.²⁸ If there is liability where a human performs a certain act, why should we give a machine more favourable fair use treatment when it does the same (especially as AI can do this on a much larger scale)?²⁹ The law should not facilitate a *binary* system where humans are disadvantaged when carrying out the same task and should stay ‘technology

Law 217.

²⁵ However, please note that the works in question were already in the public domain here (the Dutch maestro died in 1669).

²⁶ See Christie’s, ‘Is artificial intelligence set to become art’s next medium?’ (2018) <<https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>> accessed 18 January 2022.

²⁷ See also Lim D, ‘AI & IP: Innovation & Creativity in an Age of Accelerated Change’ (2018) 52 Akron Law Review 813, p. 850 (commenting on Jukedeck: ‘[t]here is no doubt that AI-generated, royalty-free sound recordings would jeopardize the market for recordings that are composed and performed by humans in a traditional fashion. Jukedeck’s rates are lower than what it would cost to license a conventional sound recording, and its output is not limited by the constraints human composers or recording artists face’).

²⁸ Gervais D, *The Machine as Author* (2019) 2015 Iowa Law Review, pp. 36-38 on SSRN (noting that ‘with rights comes responsibilities’); Senftleben M & Buijtelaar L, ‘Robot Creativity: An Incentive-Based Neighbouring Rights Approach’ (2020) European Intellectual Property Review, 42 (12), p. 810-811; Carlisle S, ‘Should Music Created by Artificial Intelligence Be Protected by Copyright?’ (2019) at p. 134

<<http://copyright.nova.edu/ai>> accessed 19 January 2022; for a more detailed overview of ownership in this context, see also see Bonadio E and McDonagh L, ‘Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity’ (2020) Intellectual Property Quarterly, pp. 112-137.

²⁹ Sobel B, ‘Artificial Intelligence’s Fair Use Crisis’ (2017) 41 Colum. J.L. & Arts 45, p. 34; see also Lim D, ‘AI & IP: Innovation & Creativity in an Age of Accelerated Change’ (2018) 52 Akron Law Review 813, p. 851 (‘AI producing commercially valuable art, prose, or music trained on copyrighted works chafes uncomfortably against interests that normally attract infringement liability if done by humans’).

neutral’ insofar as possible.³⁰ This risk of a double-standard framework could also be explained with reference to the terminology used in this context—source material used by humans is typically referred to as ‘works’ whereas the same material used by AI is called ‘data’.³¹

Furthermore, there are genuine concerns that granting AI more favourable treatment may have a negative impact on human creativity in the long run. Indeed, if creative AI benefits from more lenient fair use rules, this could contribute to a gradual automation or ‘roboticization’ of entire industries, ultimately displacing humans as creators or at least diminishing their involvement to a significant extent.³² AI could increasingly dominate fields including art, music, film and literature as there would be no need to clear IP rights—a scenario that is certainly worrying and unlikely to have a net positive effect on social welfare.³³

On the other hand, when making the assessment of whether AI should benefit from more lenient fair use treatment, we also need to take into account the issue of algorithmic bias. As noted by Amanda Levendowski, copyright law imposes strict limits on access to training data (i.e. only public domain materials can be used freely) and the amount of participants who can use specific outputs.³⁴ Most public domain works were created before the twentieth century when the art world was ‘wealthier, whiter, and more Western’.³⁵ If creative AI is primarily able to access such input material, there is a clear risk that underprivileged communities—

³⁰ See also Grimmelmann J, ‘There’s No Such Thing as a Computer-Authored Work’ (2016) 39 Colum. J. L. & Arts 377; Annemarie Bridy, ‘The Evolution of Authorship: Work Made by Code’ (2016) 39 Colum. J.L. & Arts 395 pp. 658 and 674-675 (‘By valorizing robotic reading, copyright doctrine denigrates human reading. A transformative fair use test that categorically exempts robots means that a digital humanist can skim a million books with abandon while a humanist who reads a few books closely must pay full freight for hers’... ‘Copyright’s expressive message here—robots good, humans bad—is the exact opposite of the one it means to convey’).

³¹ Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579, p. 625 (‘[a] best-selling novel becomes data about how humans use language; a selfie becomes data about the features of the human face; a conversation from a film becomes data about human voices’).

³² Lim D, ‘AI & IP: Innovation & Creativity in an Age of Accelerated Change’ (2018) 52 Akron Law Review 813, p. 850.

³³ See also Abbot R, *The Reasonable Robot: Artificial Intelligence and the Law* (CUP, 2020), pp. 1-17, on the issue of ‘AI legal neutrality’ (‘...there needs to be a new guiding tenet to AI regulation, a principle of AI legal neutrality asserting that the law should not discriminate between AI and human behavior. Currently, the legal system is not neutral. An AI that is significantly safer than a person may be the best choice for driving a vehicle, but existing laws may prohibit driverless vehicles. A person may be a better choice for manufacturing goods, but a business may automate because it saves on taxes. ... In all these instances, neutral legal treatment would ultimately benefit human well-being by helping the law better achieve its underlying policy goals.’).

³⁴ Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579, p. 589.

³⁵ Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579, p. 589.

including women, people of colour and the LGBTQ community—may be ignored.³⁶ Granting AI a more relaxed fair use treatment where it can access protected content (which effectively shapes the final outputs) could thus contribute to a more modern, diverse and tolerant body of works being produced. Moreover, under the current framework, the existing threat of infringement could certainly force AI companies to keep their datasets private, which makes it difficult for third-party observers (including academics, journalists and NGOs) to monitor potential biases.³⁷ If, on the other hand, rules are relaxed, this may encourage developers to make these datasets public and therefore subject to more scrutiny.

As a result, a strong argument could be made that a more lenient fair use regime for AI may help to avoid bias and outdated social norms—promoting fairer, more transparent and accountable AI that reflects today’s attitudes.³⁸ While this point is clearly convincing and in line with modern views on what IP should aim to achieve, a key challenge remains. Namely, balancing this potential expansion of fair use with the need to avoid a binary regime which, as explored above, could have serious consequences for human creativity.

Exceptions under EU and UK law: transient copies

Article 5(1) of the Information Society Directive—known as the ‘transient copy’ exception—may be applicable in relation to some uses of creative AI where the reproduction is merely temporary.³⁹ In order to rely on this exception, the copying must: i) be incidental or transient; ii) form an essential part of the technological process; iii) enable the lawful use of a work; iv) and have no independent significance.⁴⁰ The criteria are *cumulative* and will be interpreted strictly by the court.⁴¹ The exception generally permits the copying of a protected work for the

³⁶ Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579, p. 589.

³⁷ Matsakis L, ‘Copyright Law Makes Artificial Intelligence Bias Worse’ (2017) <https://motherboard.vice.com/en_us/article/59ydmx/copyright-law-artificial-intelligence-bias> (owing to copyright ‘major AI companies keep the data they use to train their products a secret, preventing journalists and academics from uncovering biases, as well as stifling competition’).

³⁸ See again Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579, p. 630.

³⁹ Directive 2001/29 on the harmonisation of certain aspects of copyright and related rights in the information society (Information Society Directive); see Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* EU:C:2009:465; *The Newspaper Licensing Agency v Meltwater Holding BV and others* [2010] EWHC 3099 Ch.

⁴⁰ This is satisfied where the relevant act does not result in additional profit beyond what is obtained from the lawful use, or in modifying the same work; see Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* EU:C:2009:465; Case C-302/10 *Infopaq International A/S v Danske Dagblades Forening* (‘Infopaq II’); see also Margoni T, ‘Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?’ (2018) *CREATe Working Paper* 2018/12 <<http://eprints.gla.ac.uk/175022/>> accessed 30 May 2020.

⁴¹ Case C-302/10 *Infopaq International A/S v Danske Dagblades Forening* (‘Infopaq II’), para. 27.

purposes of performing mechanical tasks which have no autonomous value (e.g. web browser data stored in a cache). The CJEU has previously considered this provision in the context of data capture, which is arguably not too different from some of the steps involved in modern machine learning.⁴²

Consider the example of an AI application scanning through data on weather forecasts, aiming to assist users with scheduling holidays. In this case, some data may be temporarily stored so that it can be transmitted through a network between third parties.⁴³ Here, the reproduction is clearly an essential part of the technological process and it is not necessary to keep the data once it has been run through the AI, i.e. this may be regarded as non-expressive use.⁴⁴ Provided that there is no economic harm for the rightholder, this use would also satisfy the three-stage test in the Directive (which states the provision will only apply in circumstances which ‘do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder’).⁴⁵

Exceptions under EU and UK law: text and data mining

Creative AI could also be exempt under the Digital Single Market (DSM) Directive’s text- and data-mining exception.⁴⁶ Text and data mining (TDM) concerns the extraction and use of large amounts of data for the purposes of finding patterns, discovering relationships, providing valuable information for research and other activities.⁴⁷ In order to rely on this exception, first, the relevant activity must be performed by cultural heritage and research institutions in the context of scientific research.⁴⁸ Given the lack of a broader fair use doctrine in Europe, the

⁴² Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* EU:C:2009:465; Case C-302/10 *Infopaq International A/S v Danske Dagblades Forening* (‘Infopaq II’); see Margoni T, ‘Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?’ (2018) *CREATe Working Paper* 2018/12 <<http://eprints.gla.ac.uk/175022/>> accessed 30 May 2020.

⁴³ This example is given by Schafer B et al, ‘A Fourth Law of Robotics? Copyright and the Law and Ethics of Machine Co-Production’ (2015) 23 *Artificial Intelligence and Law* 217.

⁴⁴ Schönberger D, ‘Deep Copyright: UP- and Downstream Questions Related to Artificial Intelligence (AI)’ in *Droit d’auteur 4.0 / Copyright 4.0* (2018), pp.16–17.

⁴⁵ Information Society Directive, Article 5(5); see also *Football Association v QC* (emphasising the need for an appropriate balance between the needs of users and rightholders in this context; see also *Football Association Premier League Ltd v QC Leisure* (C-403/08) EU:C:2011:631; [2012] Bus. L.R. 1321; and *Karen Murphy v Media Protection Services Ltd* (C-429/08) EU:C:2011:631.

⁴⁶ See also Directive 2019/790 on copyright and related rights in the Digital Single Market arts 3 and 4. Article 3 in particular exempts from copyright infringement the reproduction of copyright material.

⁴⁷ See Directive 2019/790, Article 2(2) (‘any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations’); See Flynn S, ‘WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)’ (2020) Working Papers 43, available at <https://digitalcommons.wcl.american.edu/fac_works_papers/43> accessed 20 June 2020.

⁴⁸ See Directive 2019/790, Article 3

more limited protection offered under this exception may encourage AI companies to engage in mixed partnerships with public entities.⁴⁹ However, this provision will be of limited significance where companies wish to exploit the final output commercially (s. 29A of the UK CDPA, for instance, specifically restricts the application of the exception to reproduction for the purposes of non-commercial research).⁵⁰ While Article 4 of the Directive seemingly permits TDM to be performed by business entities and for any purpose, there is an important caveat.⁵¹ The provision is inapplicable where rightsholders reserve the right to mine, which significantly limits its usefulness in practice.⁵² Finally, one key unanswered question here—which should hopefully be addressed by the CJEU—is whether non-profit entities could also utilise the exception where the mined data is used in an expressive manner.

Overall, it is clear that the scope of the relevant EU and UK exemptions is quite narrow compared to US law and its broad fair use doctrine, which may be a concern for AI businesses based in Europe. Thomas Margoni, for instance, argues that EU copyright law is ‘falling behind’ other jurisdictions due to its cumulative and narrow interpretation of Article 5(1) ISD, thus contributing to a less favourable and ‘innovation-oriented’ environment for advancing AI.⁵³ Similarly, it is evident that the text- and data-mining exception only offers limited protection with regards to commercial activities.

INFRINGEMENT LIABILITY

How should liability be allocated where AI performs an infringing act and there are no available exceptions? While courts have not clearly addressed this question yet, the threat of infringement and lack of legal certainty could cause harm to the advancement of AI if developers are discouraged from creating and distributing important products and there is no timely guidance on the issue.⁵⁴

⁴⁹ See also Directive 2019/790, Recital 11 (referring to EU policies which facilitate ‘universities and research institutes to collaborate with the private sector’).

⁵⁰ CDPA, s. 29A.

⁵¹ See Directive 2019/790, Article 4.

⁵² Rosati E, ‘Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity’ (12 September 2019), *Asia Pacific Law Review*, p.21, SSRN, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452376> accessed 19 January 2022.

⁵³ See Margoni T, ‘Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?’ (2018) *CREATe Working Paper* 2018/12, in ‘Final remarks’ <<http://eprints.gla.ac.uk/175022/>> accessed 30 May 2020.

⁵⁴ Schaal E, ‘Infringing a Fantasy: Future Obstacles Arise for the United States Patent Office and Software Manufacturers Utilizing Artificial Intelligence’ (2004) 11 *Jeffrey S. Moorad Sports Law Journal* 173, 201.

Allocation of liability

AI software is typically not the work of a single individual. A company which develops AI typically employs a team of developers (and also assigns and manages their projects and schedules).⁵⁵ For instance, there are multiple different parties who contribute to the programming of IBM's Watson (while a certain company may be responsible for selling the system to consumers, another party might be responsible for maintaining it and applying updates).⁵⁶ There are also end users and consumers who may have considerable control over the AI and the input data fed into the system once the product is sold. Therefore, considering all the parties that may be involved in the process, there are a number of possible 'candidates' for liability when AI infringes. This includes the end user; the seller or developer/programmer (in a broad sense); and the AI system itself.

The idea of granting AI legal personality (and potentially holding machines or robots liable) has already been discussed by various commentators and high-profile institutions. In 2017, the European Parliament raised the issue of 'creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons'.⁵⁷ It is evident that the EU is not only concerned about the possibility of human displacement as a result of the proliferation of AI, but also about accountability when robots cause damage to people or property.⁵⁸ It has essentially considered placing intelligent machines on equal footing with corporations which already have 'legal personhood', an approach which may pave the way for liability in the long run.⁵⁹ While some commentators have reported that Saudi Arabia recently granted Sophia (a humanoid robot) citizenship, the accuracy of these reports has been disputed and some have described this as nothing more than a publicity stunt.⁶⁰ As of 2021, no major jurisdiction (e.g. US, EU and UK)

⁵⁵ Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14; Watson B, 'A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence' (2017) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65, 81.

⁵⁶ See Watson B, 'A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence' (2017) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65.

⁵⁷ European Parliament, 'Motion for a European Parliament Resolution: with recommendations to the Commission on Civil Law Rules on Robotics' (2015/2103(INL)) <https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html?redirect>; Parviainen J and Coeckelbergh M, 'The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market' (2020) *AI & Society*.

⁵⁸ See Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14.

⁵⁹ Delcker J, 'Europe divided over robot 'personhood'' (2018) *Politico* <<https://www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood/>> accessed 5 February 2022.

⁶⁰ Vincent J, 'Pretending to give a robot citizenship helps no one' (2017)

has recognised AI or robots as entities capable of having legal status. But while machines have no legal personality and cannot be held accountable at present, developers and/or users could certainly face liability.⁶¹

Consider the example of *PaintsChainer*, an AI-powered colorization software programme.⁶² The project helps artists colorize their works by allowing them to apply colours to each drawing with very few instructions and with little human involvement, saving them considerable time and allowing the AI to handle the colorization process largely on its own. However, this creates the risk of copyright infringement. Users may, for instance, apply colours to copyright protected characters (this may involve making a pink Pikachu, which is unlikely to be covered by fair use or similar provisions).⁶³ The crucial point here is that the company merely provides the tool to artists, i.e. the user has complete control over what is fed into the system. The developers effectively provide the necessary code which is then automatically adjusted as per the user's needs. They neither encourage infringement, nor do they have any control over the process.⁶⁴ Moreover, they do not monitor or deal with the AI-generated outputs.

Even though the end user may be liable for primary infringement here, under existing (e.g. US) law,⁶⁵ and as per the landmark case of *Grokster*, secondary liability will only be found where the seller encourages users to commit infringement by using their AI products.⁶⁶ In other words, the developer/seller will not be liable in instances where the AI has substantial lawful use and the product or service is appropriately marketed and advertised.⁶⁷ As a result, companies such as *PaintsChainer* (which merely produce and distribute AI that has substantial lawful use) will generally avoid liability, despite the clear possibility that the AI might be misused.⁶⁸

<<https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia>> accessed 5 February 2022.

⁶¹ On the issue of legal personality and 'punishing AI' more generally, see also Abbot R, *The Reasonable Robot: Artificial Intelligence and the Law* (CUP, 2020) pp. 111-133.

⁶² See Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14, p. 36.

⁶³ See Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14, p. 36. This may be regarded as a derivative work.

⁶⁴ See Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14.

⁶⁵ While this analysis focuses on US law, similar principles apply under UK and EU law.

⁶⁶ See e.g. *Metro-Goldwyn-Mayer Studios v Grokster*, 545 U.S. 913, 919 (2005) ('one who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties.') (US law).

⁶⁷ See e.g. *Sony Corp. of Am. v. Universal City Studios*, 464 U.S. 417, 422-23 (1984), at 440 ('a staple article or commodity of commerce suitable for substantial non-infringing purposes' is not contributory copyright infringement'); see also Naqvi Z, 'Artificial Intelligence, Copyright, and Copyright Infringement' (2020) 24 Marq. Intellectual Property L 14, pp. 35-38.

⁶⁸ See *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005).

Policy considerations—who *should* be liable where AI acts autonomously?

While the analysis so far has focused on who would be liable where AI is primarily viewed as a ‘tool’ (which is largely unproblematic from a doctrinal and policy perspective) and at least one of the parties has some awareness and control over the process, it is far less clear how liability *should* be allocated where, for instance, AI alters its own programming (e.g. IBM’s Watson is capable of doing this through machine learning) to such an extent that neither developers nor end users are able to appreciate the risk of infringement.⁶⁹ This inevitably entails normative considerations and raises issues of fairness and—as courts have not specifically addressed the issue yet—legal certainty.

As illustrated by the European Parliament in a recent Motion for a Resolution on the regulation of robotics, traditional rules on liability do not apply neatly to scenarios where robots or machines make highly independent decisions.⁷⁰ Are developers or end users more blameworthy? Who is in a better position to appreciate the risk? Should anyone be held accountable at all? Holding end users liable (where they are not necessarily aware that infringement may occur) could be particularly unfair,⁷¹ especially as consumers/users are often legally unsophisticated individuals. Adopting a punitive approach towards users may also discourage the use of (otherwise helpful) AI and, in any event, rightsholders tend to sue companies that develop and/or sell products, as opposed to their users (even where users are sued, they are often indemnified through contracts).⁷² Programmers and developers are arguably in a better position to appreciate the risk (as the AI continuously learns and alters its programming) and they are also more likely to acquire economic value from the AI by, for instance, selling it or licensing it. Nevertheless, given the very nature and purpose of processes such as machine learning, expecting developers to foresee the risk will not always be realistic, and some degree of case-by-case assessment is likely to be required (and desirable).

⁶⁹ There is further uncertainty with regards to instances involving open-source development and multi-collaborative work involving programmers, users and various other parties. In other words, there are a number of possible scenarios where someone may not be able to appreciate the risk of infringement, which further complicates the issue of allocating liability.

⁷⁰ European Parliament, ‘Report with recommendations to the Commission on Civil Law Rules on Robotics’ (2015/2103(INL)) <https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html> accessed 28 June 2020 (see ‘Liability’); see also Mendis D et al, ‘Artificial Intelligence and Intellectual Property: The View of The British and Irish Law, Education and Technology Association’ (BILETA) (November 22, 2020) <<https://ssrn.com/abstract=3752956> or <http://dx.doi.org/10.2139/ssrn.3752956>> at p. 3.

⁷¹ Watson B, ‘A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence’ (2017) 58(1) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65.

⁷² See Watson B, ‘A Mind of Its Own - Direct Infringement by Users of Artificial Intelligence Systems’ (2017) 58(1) IDEA: The Journal of the Franklin Pierce Center for Intellectual Property 65, p. 85.

Until this issue receives further attention by policymakers, Watson suggests adopting an interim contractual solution (although this is discussed in the context of patents, a similar approach may be considered in the context of copyright).⁷³ Under this proposal, for the developer or seller to hold the purchaser harmless for infringement, i) the buyer must regularly apply software updates provided by the developer; ii) the purchaser must notify the selling party of any new methods created by the AI (or, in the context of copyright, this may include new techniques which may considerably increase the risk of infringement); iii) and the buyer must not commit infringing activities in bad faith.⁷⁴ The benefit of adopting such an approach is that the buying party (or user) will be actively encouraged to do its part to limit infringement, and the risk of unfairness is mitigated as liability would only be imposed where infringement takes place with the purchaser's knowledge and capability to control. In other words, liability for end users would be reserved for the most egregious cases and only insofar as they act in bad faith (e.g. where they intentionally reprogram the AI to infringe).

Regardless of the approach adopted in the short term, it is clear that this is an increasingly pressing (and difficult) question that will inevitably require the attention of the legislature. It is also evident that failing to hold any party accountable is problematic as this might indirectly encourage the use of AI systems for infringement purposes.⁷⁵ Why should the use of AI be given special treatment, while holding human actors liable in the same or similar set of circumstances? As already mentioned, if some form of protection for AI-generated works is to be granted (whether copyright or a sui generis right), it is also fair that someone (e.g. the person who claims to be the 'author' of the final output) should bear responsibility where the use of the machine constitutes infringement.

CONCLUSION

AI systems consume large amounts of data, often involving creative works such as books, photographs and articles. As AI continues to steadily advance and the amount of protected

⁷³ Watson B, A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence (2017) 58(1) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65.

⁷⁴ Watson B, A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence (2017) 58(1) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65.

⁷⁵ Watson B, A Mind of Its Own-Direct Infringement by Users of Artificial Intelligence (2017) 58(1) IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property 65, p. 70. When addressing this, we must also be cautious not to hinder the advancement of AI technology; see Schaal E, 'Infringing a Fantasy: Future Obstacles Arise for the United States Patent Office and Software Manufacturers Utilizing Artificial Intelligence' (2004) 11 *Jeffrey S. Moorad Sports Law Journal* 173, p. 201.

works consumed by intelligent machines expands dramatically, policymakers will have to consider a range of complex questions regarding copyright infringement, exceptions and allocation of liability which have received surprisingly little attention so far. As our analysis shows, this inevitably raises questions of copyright infringement and liability. When would the use of an AI system infringe? Can AI-induced infringement be exempted and are the existing exceptions satisfactory? There is also no consensus on who should be liable where a machine acts with a considerable degree of autonomy.

While AI-induced infringement concerns both inputs to and outputs of AI, a wide range of (otherwise infringing) uses may be permitted under the existing framework. Under US law, for instance, the processing of protected works for e.g. pattern recognition purposes is likely to be regarded as non-expressive and thus ‘fair use’. This broad fair use doctrine can be clearly contrasted with EU and UK law where the scope of the transient copy and text- and data-mining provisions is considerably more restrictive (e.g. the latter only applies where the activity is carried out by a research/cultural heritage institution). There are thus genuine concerns that this could negatively impact on the advancement of AI and preclude local businesses from carrying out activities that would otherwise be exempt in e.g. the US (and may also force them to relocate to more favourable jurisdictions).

In terms of allocating liability, we note that machines have no legal personality in the US, EU and UK and cannot therefore be held responsible for infringement. While programmers/developers and end users could face liability where the AI infringes on another’s work, the question of who should be liable where the machine carries out tasks with a significant degree of autonomy (and alters its programming to such an extent that neither party can appreciate the risk) has not been fully explored yet and may inevitably require a case-by-case assessment. Regardless of the approach adopted in the future, it is essential that policymakers take into consideration issues of balance between rights and obligations and, crucially, fairness.