

# Generative AI, Education, and Copyright Law: An Empirical Study of Policymaking in UK Universities

**Guido Noto La Diega**  
University Of Stirling

**Christof Koolen**  
KU Leuven

☞ Artificial intelligence; Copyright; Educational policy; Originality; Universities; Virtual reality

## Abstract

*Generative artificial intelligence (GenAI) tools such as ChatGPT, Stable Diffusion, and Gemini are promising to revolutionise higher education. While universities encourage students to embrace GenAI, many are unaware of the risks of crossing the blurred lines between use and misuse. The ongoing uncertainty as to whether copyright law is fit for GenAI highlights the urgency to delineate the policy framework surrounding the use of GenAI in academia. Against this backdrop, the current article explores the intersection between GenAI, education, and copyright by analysing the AI policies of UK higher education institutions (HEIs) obtained through Freedom of Information (FOI) requests. In particular, the article considers the policy-making processes, the stakeholders involved, and the role of copyright in these policies. In addition, it presents two case studies that aim to add nuance to the current discourse around AI authorship*

*and infringement. By addressing these critical issues, this article aspires to shine a light on areas of tension between copyright, education, and GenAI, and to equip HEIs and educators with the tools to use GenAI in more critical and responsible ways.*

*“With proper nourishment and care, there is nothing that will not grow; without proper nourishment and care, there is nothing that will not decay”*

Book of Mencius, 2A:6:7.

## Introduction

ChatGPT, Stable Diffusion, Gemini, and other Generative artificial intelligence (GenAI) tools are likely to be this generation of students' equivalent to the calculator and the smartphone.<sup>1</sup> GenAI bears the alluring promise to spare learners from many repetitive tasks (e.g. proofreading), thus leaving them space for anything that requires originality, critical thinking, intuition, and empathy.<sup>2</sup> Recent data shows that 50% of teachers know at least one student who has faced negative consequences for using ChatGPT in their assignments.<sup>3</sup> This problem can only get worse as AI-generated submissions are becoming as good as human ones, and they are increasingly not caught by anti-plagiarism software, which will incentivise students to use GenAI.<sup>4</sup> For copyright scholars, the growing adoption of GenAI in education poses both immediate and longer-term challenges, as these commercial GenAI tools risk normalising the unauthorised use of protected content, and ultimately may usher in the backdoor privatisation of our education sector. The question of the privatisation of education is closely connected to the increasing reliance on private providers of 'AIED' or more generally 'EdTech'<sup>5</sup> tools and platforms. Developed for commercial purposes, EdTech is then deployed in educational contexts.<sup>6</sup> Whilst the impact of EdTech on education and copyright used to be a niche topic, as GenAI tools become ubiquitous, the question of the nature and contours of AI-enabled education—and the role of copyright law in providing a governance framework—becomes unavoidable.

\* Prof. Dr. Guido Noto La Diega is Chair in Intellectual Property and Technology Law, University of Stirling, and Martin-Flynn Global Law Professor, University of Connecticut. For correspondence: gn12@stir.ac.uk. Dr. Christof Koolen is a postdoctoral researcher at the Centre for IT & IP Law, KU Leuven. The authors are grateful to Lionel Bently KC for his helpful pointers, and the peer reviewers for the supporting comments. The responsibility for this article and any errors herein rests solely with the authors.

<sup>1</sup> Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (UK Parliament, 2023) [36].

<sup>2</sup> There can be no consensus as to what the intrinsically human characteristics are i.e. the ones that cannot be automated and therefore are worth developing in our students. Confucian philosopher Mencius referred to benevolence, justice, propriety, and wisdom as the pillars of human nature and posited that education had to awaken what was innately human. See *Book of Mencius*, 6A:11:4.

<sup>3</sup> Mark Keierleber, 'Survey: ChatGPT Getting Students into Trouble' (GovTech, 13 October 2023), <https://www.govtech.com/education/k-12/survey-chatgpt-getting-students-into-trouble>.

<sup>4</sup> Will Yeadon et al., 'The death of the short-form physics essay in the coming AI revolution' (2023) 58 *Phys Educ* 035027.

<sup>5</sup> EdTech, albeit a contentious term, has proven to be "an increasingly capacious category [designating] a huge variety of actors (human and nonhuman), organisations (public, private or multisector), material and technical forms (hardware, software, supporting documents), modes of practice (of teachers, designers, promoters), and framing discourses, as well as being a highly varied field of research, development and critical inquiry" (Ben Williamson, 'Meta-Edtech' (2021) 46 *Learning, Media and Technology* 1).

<sup>6</sup> This is not a radically new phenomenon; for example, the shift towards distance learning, with the adoption of Teams, Zoom and other services provided by commercial companies, could also be seen in this light, as argued in G. Salza, G. Noto La Diega, G. Priora, and B.J. Jütte, 'Zooming in on Education: An Empirical Study on Digital Platforms and Copyright in the United Kingdom, Italy, and the Netherlands' (2022) 13(2) *European Journal of Law and Technology* 1–24.

We posit that the need to govern GenAI in education through copyright—and to advance the comprehension of this understudied triad<sup>7</sup>—has never been more pressing. While the law itself will require some time to catch up,<sup>8</sup> policymaking by private and public actors ends up playing a prominent regulatory role, especially in the short and medium term. Against this backdrop, this article aims to analyse the policies adopted by UK universities to deal with GenAI, and to critically assess the role of copyright in governing the technology. Currently, there is one openly accessible repository of AI-related university policies.<sup>9</sup> While the listed policies relate exclusively to US colleges, their analysis suggests that copyright is a crucial concern.<sup>10</sup>

This article moves from a reflection on the increasing importance of copyright in a GenAI world, especially in education (*Section I*). *Section II* then presents the methodology used to obtain information on the use of GenAI in education through Freedom of Information (FOI) requests<sup>11</sup> targeted at all UK higher education institutions (HEI). *Section III* presents the results of the computational analysis applied to the FOI responses to shine a light on how HEIs have responded to the advent of GenAI. Particular attention is paid to which actors were involved in policymaking, what processes were followed, how GenAI is used and regulated, and the role of copyright within these policies. *Section IV* concludes with two case studies. Based on some recurring provisions in the FOI data, we investigate the relationship between educational copyright and GenAI by asking:

- (i) Is the AI-generated output the student's own intellectual creation (*Section IV.A*)?
- (ii) Do students expose themselves to liability for copyright infringement by using AI-generated content (*Section IV.B*)?

We put forward that the answers to these questions have wider relevance in terms of how education and copyright in general are changing under the mounting pressure of the latest technological advancements, which are once again rewriting the rules of how knowledge is created, mediated, consumed, and shared.

## 'GenCopyright': Generative AI generates renewed interest for copyright as a tool for AI governance

While copyright law has always shown interest in AI,<sup>12</sup> the regulation of AI has mostly ignored copyright. When the European Commission introduced the draft AI Act,<sup>13</sup> there was little—if any—consideration for copyright.<sup>14</sup> Similarly, in the infamous UK pro-innovation approach to AI,<sup>15</sup> copyright did not feature, and the UK Government made it clear that the framework would not “address all of the wider societal and global challenges that may relate to the development or use of AI [including] access to data”.<sup>16</sup> Education too is overlooked, and it appears only in the form of training for business and awareness campaigns for consumers.

The turning point coincides with ChatGPT going viral in the first half of 2023, which prompted the European Parliament to put forward several amendments to the draft AI Act. Indeed, the initial proposal focused on the risks of specific narrow AI applications (e.g. use of AI to assess students and for enrolment purposes).<sup>17</sup> This architecture is weak if one accepts that GenAI is also, to some extent, general-purpose, and the risk will depend on how it is used.<sup>18</sup> Reacting to the advent of ChatGPT, Gemini, etc., the European Parliament inserted some important copyright provisions, most notably an obligation to be transparent as to the copyright works used to train the relevant models.<sup>19</sup> Similar provisions can be found in the

<sup>7</sup> There is an abundance of studies about generative AI and copyright (e.g., P. Samuelson, “Generative AI meets copyright” *Science* 381, No. 6654 (2023) 158), about copyright and education (G. Noto La Diega, G. Priora, B.J. Jütte, and L. Pascault, “Capturing the Uncapturable: The Relationship between Universities and Copyright through the Lens of the Audio-Visual Lecture Capture Policies” in E. Bonadio and C. Sappa, *The Subjects of Literary and Artistic Copyright* (Edward Elgar 2022) 207), and about generative AI and education (e.g. R. Peres, M. Schreier, D. Schweidel, and A. Sorescu, “On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice” *International Journal of Research in Marketing* (2023)), but most works overlook the intricate relationships between these three domains.

<sup>8</sup> By this, the authors do not mean to espouse the nihilistic view that the advent of new technologies makes existing laws irrelevant or necessarily unfit for purpose. A more nuanced and optimistic view is expressed by Joshua AT Fairfield, *Runaway Technology: Can Law Keep Up?* (Cambridge University Press, 2021).

<sup>9</sup> Lance Eaton, ‘Syllabi Polices for Generative AI’, <https://docs.google.com/spreadsheets/d/1IM6g4yveQMyWeUeWBM6FZVxEWCLfVWDh1aWUErWWbQ/htmlview#>.

<sup>10</sup> While only 2 mention copyright expressly, plagiarism—which presupposes crucial questions of copyright—is a primary concern (63 returns). Of the policies that refer to copyright, rather problematically, one places the burden of ensuring the AI tool does not violate copyright on the student (University of Massachusetts Lowell); the other policy points out that copyright law does not currently protect AI-generated text and images (Berry College).

<sup>11</sup> In the UK, the key statutory references for FOI requests are the Freedom of Information Act 2000 in England, Wales, and Northern Ireland, and the Freedom of Information (Scotland) Act 2002 (FOISA) in Scotland. The former (that applies also to UK-wide public authorities based in Scotland) is administered by the Information Commissioner's Office (ICO), the latter by the Scottish Information Commissioner.

<sup>12</sup> See e.g. Copyright Designs and Patents Act (CDPA), s.29A (text and data analysis defence), introduced a decade ago, and s.9(3) (computer-generated works) dating back to 1988.

<sup>13</sup> Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative act (COM/2021/206 final). As we write, the EU co-legislators have reached a political agreement after the final trilogue on 6–8 December 2023, and are now carrying out further technical work to align recitals with the articles. We have considered this version, as shared by the Belgian Presidency on 24 January 2024: Note 5662/24 from the Presidency to the COREPER: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts—Analysis of the final compromise text with a view to agreement” (Interinstitutional File 2021/0106(COD)).

<sup>14</sup> A minor exception was art.70(1)(a) about the way that authorities are to handle data obtained in carrying out their new statutory tasks.

<sup>15</sup> Secretary of State for Science, Innovation and Technology, ‘A pro-innovation approach to AI regulation’ (29 March 2023, updated 3 August 2023).

<sup>16</sup> Secretary of State for Science, Innovation and Technology, ‘A pro-innovation approach to AI regulation’ (29 March 2023, updated 3 August 2023).

<sup>17</sup> AI Act, Annex III [3]. Some uses of AI in education would be altogether banned e.g. to detect the emotional state of students (recital 26c). In the final version, e-proctoring has been added as another example of a high-risk system that is not banned as such, but is subject to strict rules.

<sup>18</sup> We are not claiming that ChatGPT is a form of Artificial General Intelligence, let alone superintelligence. We are simply saying that GenAI tools can be used for a large variety of downstream tasks and in a multitude of domains, which means that the system may be high risk if applied in a certain way and low risk if applied in others.

<sup>19</sup> AI Act, art.52c(1)(c)(d) and recitals 60f, 60g, 60ih, 60j, 60k, 60ka.

UK's AI (Regulation) Bill.<sup>20</sup> This emphasis on transparency, albeit not in and of itself sufficient, may help copyright perform its social function as guardian of human-centric creativity. It remains to be seen if the shoehorning of GenAI and copyright in this new law will provide an adequate answer to the complex questions that will undoubtedly arise. It also remains unclear how detailed the summaries of the training datasets to be published will be, and how trade secrets will be leveraged to try and retain control over the data.<sup>21</sup>

The tide is turning also in the UK. In August 2023—in parallel with the development of the Code of Practice on Copyright and AI, now seemingly shelved<sup>22</sup>—the House of Commons' Science, Innovation, and Technology Committee published its interim report on the governance of AI.<sup>23</sup> The Committee noted that AI is “already delivering significant benefits in (...) education”,<sup>24</sup> but that there is a “growing imperative to accelerate the development of public policy thinking on AI”,<sup>25</sup> as governance becomes pivotal to safety and innovation. Of the twelve challenges that AI governance needs to address, three relate to copyright. Notably, the Committee observed that, at the centre of AI governance, there must be a commitment to enforcing copyright on the content used to train AI models, to ensure access to data by tackling anticompetitive practices, and to tackle the tension between transparency and safety in open-source models.

A growing number of US lawsuits<sup>26</sup> have highlighted the increasing importance of copyright as a regulatory device in the AI space. This litigation is mostly focused on whether copyright-protected works can be used to train GenAI systems, and under which terms.<sup>27</sup> This is consistent with the shift in copyright from ‘works’ to ‘data’.<sup>28</sup> Litigation has also arisen over whether AI-generated output should be protected. In this respect, the Washington District Court reaffirmed that, in principle,<sup>29</sup> “works generated by new forms of technology operating absent any guiding human hand”<sup>30</sup> should not be protected, thus endorsing the US Copyright Office's view that human authorship is “a bedrock requirement of copyright”.<sup>31</sup> Some recent Chinese<sup>32</sup> and South-Korean<sup>33</sup> cases suggest otherwise and identify the author in the user and the film producer respectively. Upon closer examination, the US and Chinese courts and the Korea Copyright Commission (KCC) have in common an emphasis on human input in the selection, coordination, and arrangement of the works. The latest iteration of the *KCC Copyright Registration Examination Manual*<sup>34</sup> is particularly clear: “[i]f elements of traditional copyright are implemented by artificial intelligence technology, copyright is not recognized, but if humans have modified the artificial intelligence output through selection, arrangement, etc., copyright is recognized to a limited extent”. It has been argued<sup>35</sup> that the Asian decisions should be understood in light of the utilitarian theory that originated in Japan and underpins Chinese and Korean

<sup>20</sup> HL Bill 11, s.5. This is a private Peer's bill, which means that it was not introduced on the UK Government's initiative. Therefore, it does not seem likely that it will become law.

<sup>21</sup> See 60j, AI Act, recital 60, whereby the publication of the summaries about the content used to train the model has to take “into account the need to protect trade secrets”; accordingly, the summaries will be “generally comprehensive [...] instead of technically detailed [...] listing the main data collections or sets”. On the risks of IP provisions in EU data governance laws, see Guido Noto La Diega ‘Ending Smart Data Enclosures: The European Approach to the Regulation of the Internet of Things between Access and Intellectual Property’, in Stacy-Ann Elvy and Nancy Kim (eds), *The Cambridge Handbook on Emerging Issues at the Intersection of Commercial Law and Technology* (Cambridge University Press, 2024).

<sup>22</sup> The UK Intellectual Property Office's *Code of Practice on Copyright and Artificial Intelligence* was set to make licences for data mining more available. In February 2024, it transpired that this initiative would have been shelved as no agreement could be found between the stakeholders involved in drafting it (including creative industries, GenAI companies, etc.). See Daniel Thomas and Cristina Criddle, ‘UK shelves proposed AI copyright code in blow to creative industries’ (*Financial Times*, 5 February 2024), <https://www.ft.com/content/a10866ec-130d-40a3-b62a-978f1202129e>.

<sup>23</sup> Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (UK Parliament 2023).

<sup>24</sup> Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (UK Parliament 2023), p.3.

<sup>25</sup> Science, Innovation and Technology Committee, *The governance of artificial intelligence: interim report* (UK Parliament 2023), p.43.

<sup>26</sup> It is not by accident that most of the current lawsuits are class actions; indeed, the interests at stake go to the heart of the type of society we envision, and what role human learning and creativity will have in it. See *Andersen, McKernan, Ortiz, et al. v Stability AI Ltd* [2023] Case 3:23-cv-00201, and other lawsuits referred in IV.B.

<sup>27</sup> A proposal that deserves to be evaluated is the introduction of a levy system that imposes a general payment obligation on providers of GenAI systems (Martin Senftleben, ‘Generative AI and Author Remuneration’ (2023) 54 IIC 1535). Similar proposals were put forward by Giancarlo Frosio, ‘Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity?’, in Enrico Bonadio and Caterina Sganga (eds), *A Research Agenda for EU Copyright Law* (Edward Elgar 2024) 19, who nonetheless calls for caution as GenAI “requires a nuanced understanding of the roles and liabilities associated with the use of protected content”.

<sup>28</sup> Thomas Margoni and Martin Kretschmer, ‘A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology’ (2022) 71(8) GRUR Int 685.

<sup>29</sup> AI-generated works are not protected under US Copyright Law. However, a work comprising AI-generated works may be protected. In February 2023, the US Copyright Office held that the applicant was the author of a comic book including illustrations generated through Midjourney, because she had written the text and because of the “selection, coordination, and arrangement of the Work's written and visual elements” (*Zarya of the Dawn -Registration # VAu001480196*, 1).

<sup>30</sup> *Thaler v Perlmutter*, No. 22-1564 (D.D.C. Aug. 18, 2023) 8.

<sup>31</sup> *Thaler v Perlmutter*, No. 22-1564 (D.D.C. Aug. 18, 2023) 8.

<sup>32</sup> Judgement No. 11279/2023 *Li Yunkai v Liu Yuanchun* (Beijing Internet Court, 27 November 2023). While this was by many regarded as the first case recognising the copyrightability of AI-generated works, Judgement No. 14010/2019 *Tencent Computer Company v Yingxun Technology* (Nanshan District Court of Shenzhen, 24 December 2019) had already held that an article automatically generated by an intelligent writing software was eligible for copyright protection, thus overruling Judgement No. 239/2018 *Feilin Law Firm v Baidu Technology Company* (Beijing Internet Court, 25 April 2019) where the Court concluded that copyright protects only works created by human authors.

<sup>33</sup> *AI Surobin* (Korea Copyright Commission, Korea Copyright Commission, 29 December 2023).

<sup>34</sup> Korea Copyright Commission, *Copyright Registration Examination Manual* (KCC 2023) [2B6].

<sup>35</sup> Zhe Dai and Jin Banggui, ‘The copyright protection of AI-generated works under Chinese law’ (2023) 13 *Juridical Trib* 241, 254. According to the authors, alongside the civil law *droit d'auteur* and the common law copyright systems, there is a third way (which they call “the work right system”) that characterises Japan, South Korea, North Korea, mainland China, and Taiwan.

copyright, whereby the deciding question is whether the protection of GenAI content promotes the public interest, with less emphasis placed on the human contribution.

Against this backdrop, copyright is set to provide the rocky terrain on which the battle for human creativity in a world of prompt-generated knowledge will be fought. As students increasingly rely on GenAI provided by private corporations through mostly proprietary software,<sup>36</sup> governing the technology becomes a pressing issue if public interest is to remain at the core of education. It is this article's contention that copyright can contribute to this governing function, and that GenAI provides a precious opportunity to elevate copyright to a safety net to ensure that the technology is societally just and oriented towards the common good.

It is no coincidence that UNESCO is calling on governments to regulate GenAI and that copyright is noted as having special importance in that context.<sup>37</sup> UNESCO is asking the fundamental question of how GenAI can be used responsibly in education when issues of copyright infringement and authorship remain unresolved. These issues are explored in more detail in the following Sections, which assess how HEIs are approaching the use of GenAI in education.

## Methodology

As HEIs are under increasing pressure to adopt GenAI,<sup>38</sup> it is vital to understand how they are introducing safeguards to minimise any risks that the technology may pose, while harnessing its potential. To acquire a richer understanding of how UK-based HEIs are approaching GenAI, we submitted FOI requests ( $n=134$ )<sup>39</sup> pursuant to the Freedom of Information Act 2000 and the Freedom of Information (Scotland) Act 2002. HEIs were asked to clarify:

- (a) Their position on the use of GenAI, its degree of formalisation (policy, guidance, etc.), and which stakeholders were involved in this process.
- (b) Whether they permit students to use GenAI for the purposes of learning, revision, formative assessment, and summative assessment.

- (c) Whether the use of GenAI falls within the scope of the HEI's policy on academic misconduct or academic integrity.
- (d) Whether the HEI provides guidance or training on the responsible use of GenAI and, if so, whether copyright issues are addressed.

The purpose of collecting responses to these questions is, first, to shed light on the deployment of GenAI in an academic setting and the considerations, challenges, and concerns that come with it. Second, it serves to identify areas of tension or shortcomings in current policymaking, thereby indicating where improvements might be considered. The main limitation of FOI as a research method is that it tends to capture only documented practices<sup>40</sup>. In our experience, respondents were often more generous and disclosed information about work-in-progress and not fully formalised information (e.g. about copyright training). Additionally, a FOI methodology is useful as not all policies were easily accessible, many were scattered across many documents, and the resulting structured responses make it easy to carry out computational analysis.

A total of  $n=119$  responses to FOI requests were received between 25 August 2023 and 13 November 2023.<sup>41</sup> These responses either fully or partially disclosed the information requested. Responses were partial if, for instance, the HEI relied on one of the exemptions contained in the FOI legislation.<sup>42</sup> Prior to data analysis, the collected responses were consolidated into a spreadsheet. Subsequently, missing values were labelled and free-text responses standardised and converted to categorical data. The compiled dataset was then imported into Python for further analysis.<sup>43</sup>

## Results

### *HEIs' position towards GenAI*

A first finding relates to whether universities have adopted a particular stance on the use of GenAI. For this first exploratory inquiry, "taking a stance" was construed broadly and no particular attention was paid to the degree of formality associated with the HEI's position.

<sup>36</sup> The openness of GenAI is being weaponised in the fight between some of the main organisations in this sector. In July 2023, OpenAI, Microsoft, Google, and Anthropic established the Frontier Model Forum, where the emphasis seems to be on safety, which might mean that these companies are justifying the rejection of the open-source model on the usual (and debatable) grounds that it would not be safe. See, <https://www.frontiermodelforum.org/>. In December 2023, IBM and Meta—in partnership with Stability AI, NASA, Yale University, etc.—responded by launching the AI Alliance, centred on the idea that open source is pivotal to safe and responsible AI. See, <https://thealliance.ai/>.

<sup>37</sup> UNESCO, *Guidance for generative AI in education and research* (UNESCO 2023) esp. 23.

<sup>38</sup> There are financial incentives to embrace AI, or to appear to do so, thus leading to forms of AI determinism. Companies that mention AI in earnings see their stock price increase at double the speed of companies that do not mention AI (Phil Rosen, 'This chart shows how AI hype has taken over the stock market as companies try to cash in on the frenzy' (*Business Insider*, 7 September 2023), <https://markets.businessinsider.com/news/stocks/ai-stock-market-bubble-chart-earnings-investors-hype-apple-nvidia-2023-9>.

<sup>39</sup> While there are 166 universities in the UK (SI-UK 2024), only 134 have easily retrievable FOI contacts. The authors are grateful to Matthew Burgess for compiling a list of the FOI email addresses of UK HEIs at: <https://www.foi.directory/uk-universities-list-foi-emails/>.

<sup>40</sup> Frances Hamilton and Cameron Giles, 'International academic mobility, agency, and LGBTQ+ rights: a review of policy responses to internationally mobile LGBTQ+ staff/students at UK HE institutions with recommendations for a global audience' (2022) 6(1) *Policy Reviews in Higher Education* 46, 52.

<sup>41</sup> One response was received after this date, and could not be included in the dataset.

<sup>42</sup> For example under FOIA, s.21 (the information is reasonably accessible by other means), and s.22 (the information is intended for future publication).

<sup>43</sup> The NumPy and Pandas libraries were used for data handling, while visualisations were created using Matplotlib: X, 'NumPy documentation version 1.26' (NumPy, 2 January 2024), <https://numpy.org/doc/stable/> accessed 14 January 2024; X, 'Pandas documentation version 2.1.4' (Pandas, 8 December 2023), <https://pandas.pydata.org/docs/>; X, 'Matplotlib 3.8.2 documentation' (Matplotlib, 31 October 2023), <https://matplotlib.org/stable/index.html>.

A closer look at the data reveals that  $n=94$  (78.99%) HEIs have either formally or informally stated their position in relation to GenAI. Considering that the public launch of ChatGPT took place on 30 November 2022—almost a year after the FOI request responses were collected—this number suggests that a large number of HEIs are moving quickly to acknowledge the significant implications of GenAI for education. As a corollary, this means that  $n=25$  (21.01%) have not (yet) taken a position on the use of GenAI by students. Among these respondents—and although not explicitly part of the FOI request—it was occasionally mentioned that the institution in question was awaiting further developments before taking a stance, possibly in order to avoid premature action.<sup>44</sup> Several HEIs also indicated that initiatives were underway to clarify the institution's position on GenAI, but that these had not yet been adopted.<sup>45</sup>

For the 94 institutions that indicated that they had adopted a position on GenAI, the responses were further analysed to determine the nature of that position. While answers were provided in free-form text, the following four categories of responses were discerned in order of decreasing degree of formality:

- A *discrete GenAI policy*, understood as a set of rules adopted through a formal deliberative process and that is specifically tailored to GenAI. The formal policy is designed to outline a mandatory framework and is characterised by a high degree of enforceability.
- An *amendment to an existing academic misconduct policy*, understood as an addendum to an HEI's policy that was

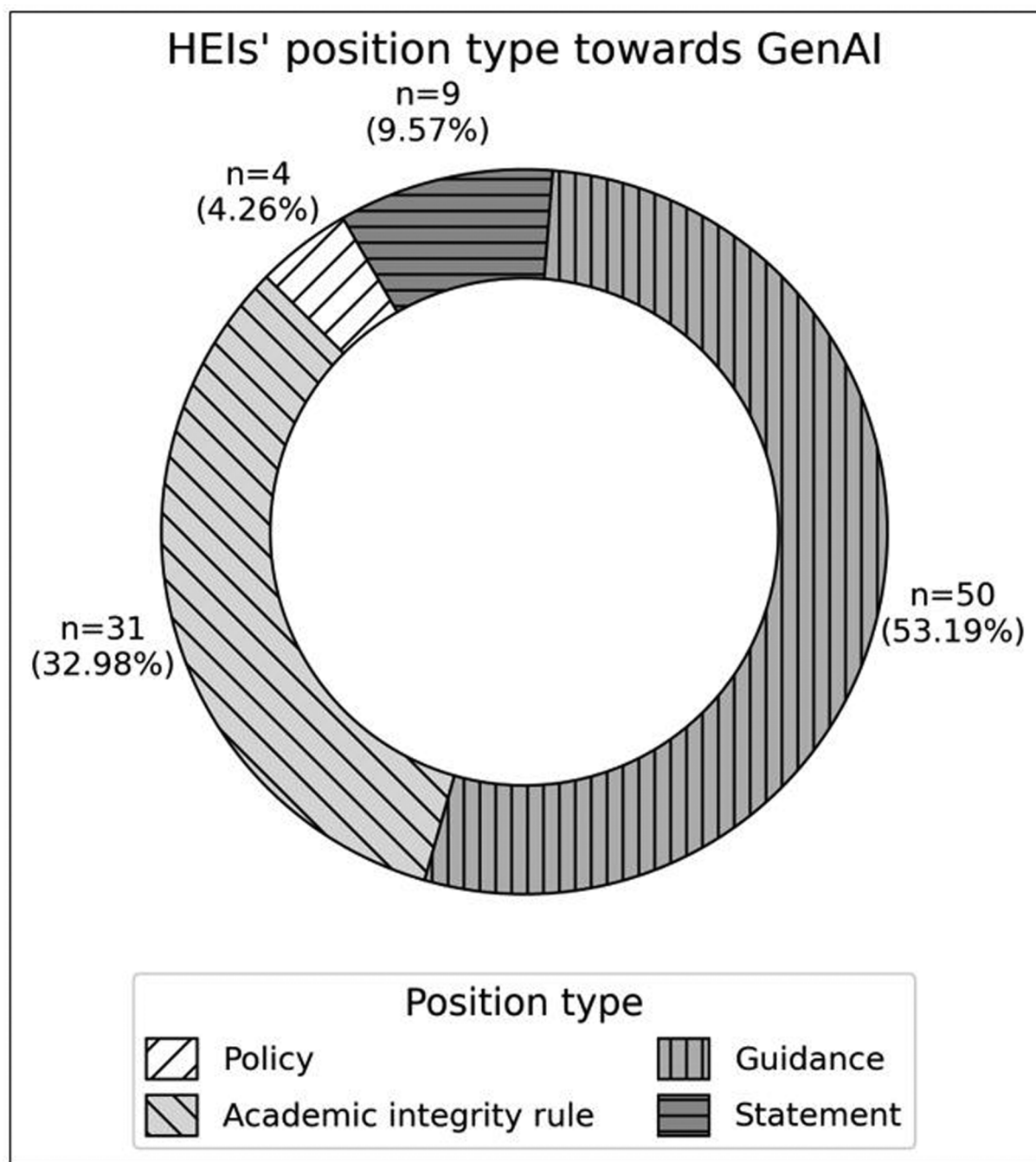
already in place, but that is not specifically tailored to GenAI. Academic misconduct and integrity policies serve as a mandatory framework for students and are characterized by a high degree of enforceability.

- *Guidance on the use of GenAI*, understood as any form of guidelines, best practices, or recommendations concerning GenAI, but that have not been subject to a formal deliberative process. GenAI guidance is intended to steer student behaviour and serves as a point of reference but tends to lack the enforceability of formal HEI policies as well as the transparency of the underlying adoption process.
- A *statement on the use of GenAI*, understood as any informal expression of an HEI's opinion on GenAI. The statement is not subject to a formal deliberative process, does not directly intend to steer student behaviour, and lacks the enforceable character of formal HEI policies.

Figure 1 shows that universities have a strong inclination to adopt guidance regarding the use of GenAI ( $n=50$ ; 53.19%). A second approach was the insertion of a new provision on GenAI in existing academic misconduct and academic integrity policies ( $n=31$ ; 32.98%). Conversely, formal GenAI policy is not widely present in the dataset ( $n=4$ ; 4.26%), and neither is the adoption of an informal statement on the use of GenAI ( $n=9$ ; 9.57%).

<sup>44</sup> For example, the University of Edinburgh, FOI request response of 23 October 2023, notes: "University governance and expert bodies are closely observing developments around the use of AI in education and have established mechanisms to inform future policy and governance to make sure our community benefits from the opportunities AI creates for learning, teaching, research, and innovation while safeguarding our values and the rights and wellbeing of our staff and students."

<sup>45</sup> Some examples include University of Liverpool, FOI request response of 11 September 2023; University of Wales Trinity Saint David, FOI request response of 11 September 2023; University of York, FOI request response of 14 September 2023.



*Figure 1.*<sup>46</sup>

It is hypothesised that the strong preference for guidance could have manifold explanations. First and foremost, the flexibility of adopting guidance compared to more stringent and formally adopted rules should not be overlooked. Indeed, guidance is typically quicker to adopt and easier to change in comparison to policy. At the same time, guidance has a relatively strong signalling value for both students and academic staff. In this sense, the preference for guidance may also reflect the need to simultaneously manage the risks to education posed by GenAI without being overly prescriptive or imposing hard rules.

This reasoning can be extended to the preference for academic misconduct rules. While more limited in scope—mandating a citation requirement or adherence to a requirement of transparency—the adoption of such a rule is done with the intention of preventing instances where GenAI is used for purposes that would negatively impact student learning. Amongst others, the importance of originality as well as the (in)ability of course convenors to verify the accuracy of referenced materials could be mentioned in this respect. Especially since the output of GenAI often blurs the lines of authorship, a clear stance

<sup>46</sup> The visualisation is the authors' own work.

on academic integrity aligns well with HEIs' commitment to address both intentional and unintentional forms of plagiarism in student submissions.

### Permissibility of using GenAI in an academic setting

A second finding pertains to whether students are allowed to use GenAI for certain purposes, namely learning, revision, formative assessment, and summative assessment. For each purpose, the responses received were categorised as 'no', 'conditional no', 'conditional yes', 'yes', 'ambiguous response', and 'no response'. A 'conditional no' is defined as a situation where the given default response is negative, but there are conditions under which the response can become positive.<sup>47</sup> The same

reasoning extends to "conditional yes" responses.<sup>48</sup> The category 'ambiguous response' reflects situations where an answer is given, but it is formulated in such a vague way that it cannot be interpreted as a clear stance on the issue.<sup>49</sup>

Figure 2 confirms what one might expect, i.e. that HEIs exercise the most caution when it comes to summative assessments, with only 30.25% of respondents allowing the use of GenAI ( $n=30$ ; 25.21% for 'conditional yes' and  $n=6$ ; 5.04% for 'yes'). A link between learning and revision purposes emerges, as well as between formative and summative assessment purposes. At the same time, there is a divide between the attitude towards using GenAI as a learning or revision tool and the attitude towards using GenAI as a tool during formative or summative assessments.

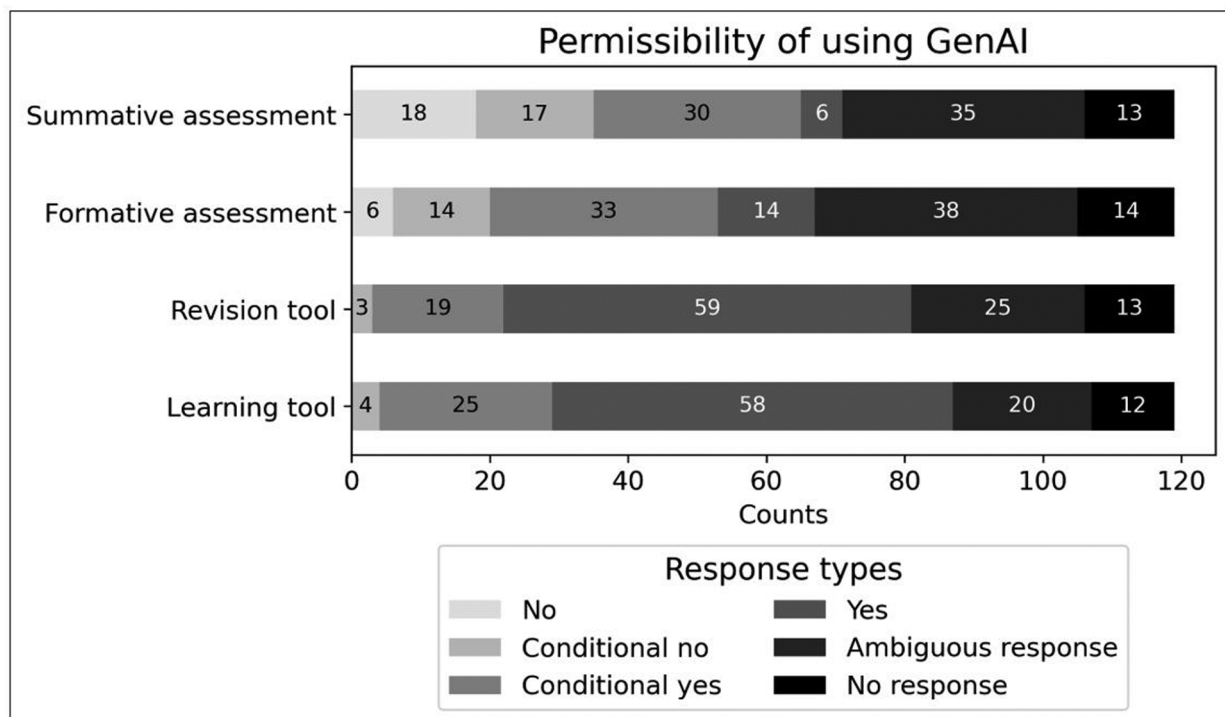


Figure 2.<sup>50</sup>

To formalise these relations, statistical analysis was performed using Fisher's exact test (95% confidence interval).<sup>51</sup> To apply Fisher's exact test, conditional responses were binned with their non-conditional counterparts, while rows with ambiguous responses and no responses were discarded. In doing so, we end up with three 2x2 contingency tables. The results of Fisher's exact test indicate that a significant association exists between an HEI's position in relation to the use of GenAI as a learning tool and as a revision tool ( $p < 0.001$ ). A similar

finding applies between an HEI's position apropos the permissibility of using GenAI in the context of formative assessments and summative assessments ( $p < 0.001$ ). Between "learning tool" and "summative assessment", a significant association was discovered as well ( $p = 0.038$ ). In all three cases, the observed odds ratio was "infinity", which indicates that one of the cells in the contingency table has a count of zero. In effect, this implies that the

<sup>47</sup> For example, University of Brighton, FOI request response of 7 September 2023: "Our policy/guidance states that the use of AI tools is not permitted in assessments unless a module leader has given express permission and incorporated [GenAI] into the assessment design as detailed in the assessment task brief".

<sup>48</sup> For example, Royal Veterinary College, FOI request response of 29 August 2023: "The use of ChatGPT and other AI sites by students in the preparation of their assessments is not prohibited unless specifically stated by Course Directors. However, if students do use AI sites to prepare their assessments they must reference them in the same way as any other source materials."

<sup>49</sup> Consider, for instance, Solent University, FOI request response of 19 September 2023: "[O]ur policy as an institution is to have open and productive conversations with students about their uses of generative AI".

<sup>50</sup> The visualisation is the authors' own work.

<sup>51</sup> Pearson's chi-squared test requires that all response categories have an expected frequency of at least 1 and that more than 80% of the expected frequencies are not less than 5. These criteria were not satisfied in the FOI response dataset.

selected variables are highly dependent and that for some outcomes in the dataset, one variable serves as a perfect predictor of the other variable.

Beyond the statistics, an interesting point can be made in relation to the specific requirements formulated in the context of conditional responses. Especially for the category ‘conditional yes’, it was often noted that the use of GenAI was subject to compliance with the HEI’s academic misconduct policy. More specifically, this meant adhering to specific fairness and transparency rules<sup>52</sup> as well as a requirement to include citations and references.<sup>53</sup> Other conditions included that the use of the tool was not specifically prohibited by the course convenor or that the use of the tool would not defeat the purpose of the learning process.<sup>54</sup> Finally, the high number of ambiguous responses registered in the context of assessment purposes is problematic, especially considering that such use may lead to misconduct proceedings.

### *GenAI as a form of academic misconduct*

Building on the previous paragraph and concerns about academic integrity, one question in the FOI request focused specifically on whether the use of GenAI could fall under the HEI’s definition of academic misconduct. Among the respondents,  $n=102$  (85.71%) answered this question in the affirmative, noting that students are at risk of committing academic misconduct if, for example, they use GenAI outside of permitted use cases.<sup>55</sup> This is in line with research showing how universities tend to frame the issues of GenAI in terms of misconduct.<sup>56</sup> Another frequently expressed view in the FOI request responses is that the HEI should be able to intervene if students are suspected of passing off GenAI work as their own.<sup>57</sup> As one institution states: “Generative AI [...] falls under the same rules and guidelines for academic integrity as any other source of information; where it has been used without attribution and outside of those specific instances where it is permitted, then the student will be subject to the same academic misconduct procedure as any other form of plagiarism.”<sup>58</sup>

A more analytical study of these responses suggests that the concerns expressed are again related to the (ability to bypass) the learning process. Aside from situations where students may submit work that is not their own—thus opening the door to labelling the behaviour

as plagiarism—it is also voiced that an over-reliance on GenAI to complete assignments could fundamentally undermine the sharpening of relevant critical thinking skills. This is particularly well illustrated in a response to the FOI request that explicitly mentions that GenAI can generate artificial references.<sup>59</sup>

Perhaps an easily overlooked follow-up question is what procedures are in place to accurately detect instances of academic misconduct, i.e., how would HEIs go about identifying written coursework as AI-generated? In fact, few responses to the FOI request address this issue. One institution mentions that it relies on plagiarism detection software Turnitin as one source of information to flag non-human authorship.<sup>60</sup> Another response notes that such behaviour is identified through a plagiarism software check and writing style analysis.<sup>61</sup> Essentially, these AI text generation detection tools, such as GPTZero and Turnitin, look for latent features hidden in AI-generated text, such as syntactic and grammatical patterns or patterns in vocabulary usage. However, as the technology currently stands, these detection software suites are not without controversy, and several media outlets have reported on AI detectors falsely labelling texts as being AI-generated.

### *Copyright and intellectual property considerations*

A final element worth highlighting is whether HEIs pay attention to making staff and students aware of the copyright issues associated with the use of GenAI. This was assessed by two questions inquiring (i) whether the HEI provides guidance on how staff and students can use GenAI, and (ii) whether the HEI offers training on how to use GenAI. A follow-up question was asked to determine whether copyright issues were addressed as part of the guidance or training.

A key takeaway is that issues pertaining to copyright are frequently left unaddressed. Figure 3 and Figure 4 indicate that while guidance or training on the proper use of GenAI is often available— $n=86$  (72.27%) and  $n=51$  (42.86%) respectively—only a small minority of HEIs touch on copyright issues. Of the HEIs that did offer guidance or training, only  $n=13$  reported that they had explicitly contextualised copyright issues in relation to GenAI in their guidance and only  $n=8$  in relation to their training on GenAI.

<sup>52</sup> E.g. Aston University, FOI request response of 1 September 2023.

<sup>53</sup> E.g. University of Greenwich, FOI request response of 13 September 2023.

<sup>54</sup> E.g. Royal College of Music, FOI request response of 15 September 2023.

<sup>55</sup> E.g. Queen’s University Belfast, FOI request response of 1 September 2023.

<sup>56</sup> Jiahui Luo (Jess), ‘A critical review of GenAI policies in higher education assessment: a call to reconsider the “originality” of students’ work’ [2024] *Assessment & Evaluation in Higher Education* 1, 12, DOI: 10.1080/02602938.2024.2309963.

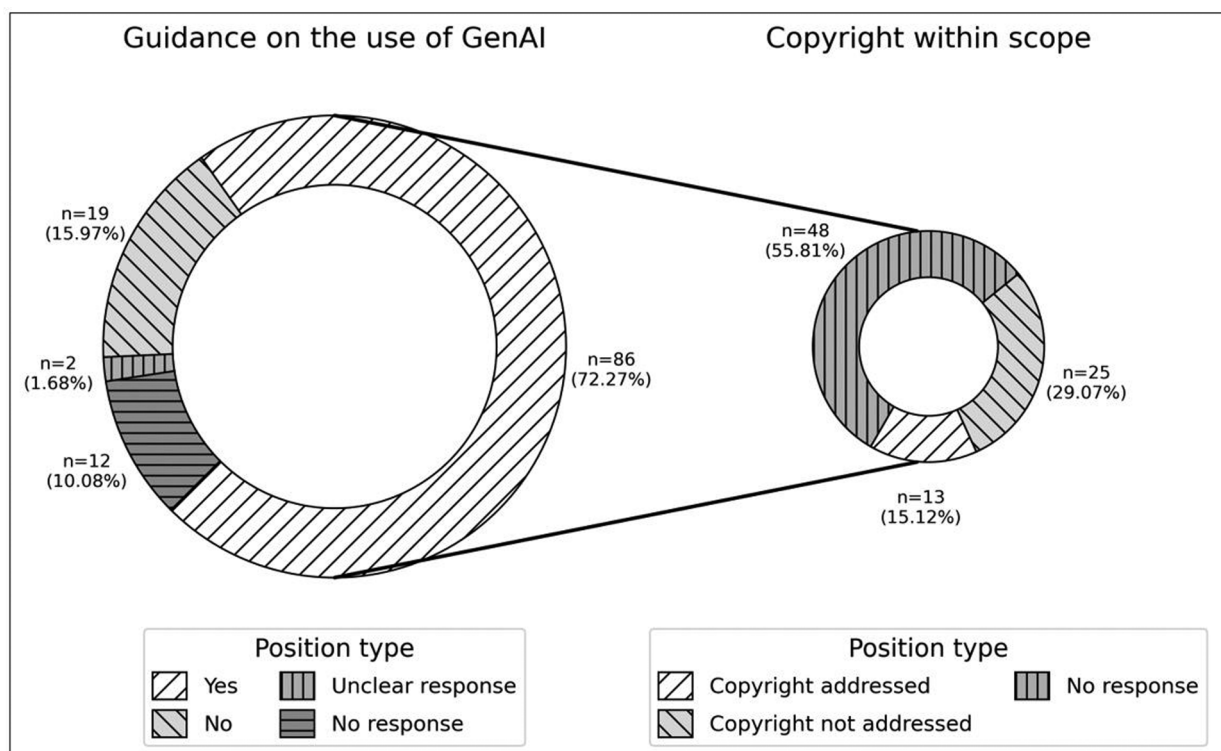
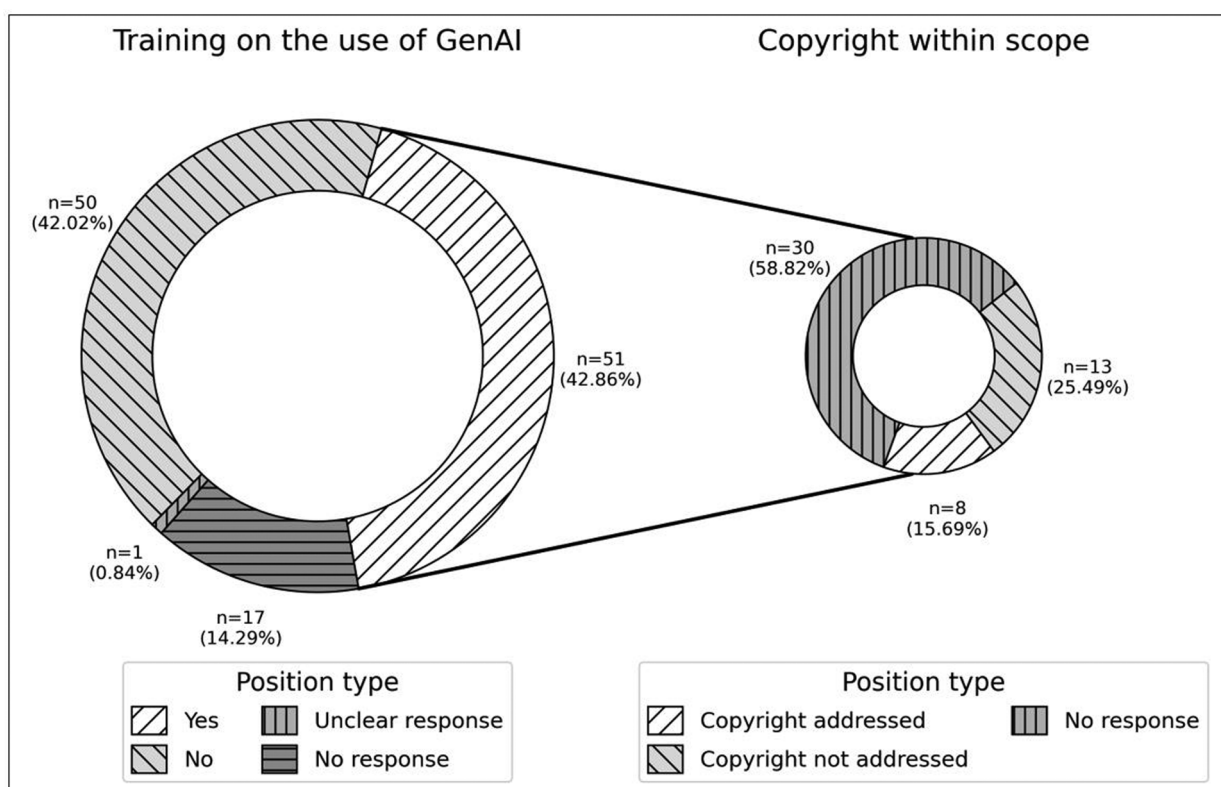
<sup>57</sup> E.g. Newcastle University, FOI request response of 5 September 2023; University of East Anglia, FOI request response of 6 September 2023.

<sup>58</sup> Solent University, FOI request response of 19 September 2023.

<sup>59</sup> E.g. University of the West of England, FOI request response of 5 October 2023.

<sup>60</sup> Birkbeck, University of London, FOI request response of 12 September 2023.

<sup>61</sup> The Open University, FOI request response of 20 September 2023.

Figure 3.<sup>62</sup>Figure 4.<sup>63</sup>

Despite these low numbers, and as a transition to the substantive copyright law analysis, it is worth highlighting some of the copyright concerns that have actually been

<sup>62</sup> The visualisation is the authors' own work.

<sup>63</sup> The visualisation is the authors' own work.

raised. One response is particularly illustrative and brings up the issue of permission to use authorial works to train AI models.<sup>64</sup> More specifically, it is pointed out that training data is often scraped from the internet, which raises questions about the ownership of the underlying data. Questions pertaining to the origins of training data and the ability to subsequently produce new data based on that training data are also noted elsewhere.<sup>65</sup> Interestingly, one institution also mentions the potential copyright implications from a student's perspective: by using one's work as an input prompt for a GenAI tool, this may also entail that the GenAI tool can embed that prompt in its data repository for future training purposes.<sup>66</sup> Aside from these concerns, copyright issues are often only mentioned in passing in most HEIs' guidelines or statements, without truly interacting with them.

Overall, the FOI request responses we received from UK HEIs confirm that the vast majority of respondents have rushed to encourage the use of GenAI by students and educators, while underlining the importance of responsible use. Copyright compliance is instrumental, as reiterated in the *Principles on the Use of Generative AI* launched by the leading research institutions in the UK.<sup>67</sup> They emphasise that GenAI may re-present information developed by others (output issue) and that the content used to train the model may have been included without the appropriate permissions (input issue).<sup>68</sup> With this in mind, they encourage UK universities to embrace the appropriate, ethical use of GenAI while upholding AI literacy, academic rigour, and integrity. The analysis below will shed light on how UK HEIs have changed their practices and policies to push AI adoption while recommending its responsible—including copyright-compliant—use by students and educators.

## A case study analysis of GenAI and copyright in UK universities

As the UNESCO recently pointed out, key barriers to AI becoming an engine for positive change in education are unresolved concerns around authorship and copyright infringement.<sup>69</sup> There can be no responsible AI in education without overcoming these foundational copyright issues. Considering the responses to our FOI requests as well as the current focus of litigation in the

GenAI space, the following sub-sections will answer two questions related to the authorship of content students generated using AI, and to the liability that may arise from the use of AI-generated student submissions. Through these case studies, it will be possible to observe how copyright law is coping with the rise of GenAI, with findings that would be of relevance also beyond educational copyright. We agree with those authors that call for interdisciplinary cross-domain approaches to the regulation of AI,<sup>70</sup> but this can only materialise if we have a sound understanding of how the issues play out in key domains.

### *Is the AI-generated output the student's own intellectual creation?*

As emerged in over 85% of the FOI responses, the principal scenario when the use of GenAI may qualify as misconduct is when the student passes off AI-generated content as their own. For example, the first pillar of the University of Edinburgh's position is not to impose a blanket restriction while emphasising “the expectation that assignments should contain students' own original work”.<sup>71</sup> Accordingly, it states that, unless explicitly authorised in a given assignment, “*content produced for assessments using AI platforms [...] is not representative of a student's own original work* and would therefore be considered a form of academic misconduct”.<sup>72</sup> Similar clauses are common,<sup>73</sup> though not all HEIs are as extreme as to automatically assume that all AI-generated content is not the student's original work. For example, Ulster University leaves it to academic staff to “make a judgement on the originality of material submitted as they would with any plagiarism review”.<sup>74</sup> Regardless of whether most staff would be able to identify the use of AI, it seems positive not to foreclose the possibility that at least in some instances original content can be generated with a GenAI tool. Both educators and students should be invited to reflect critically and openly about the issue at hand. Positively, Durham University warns students that the “law is behind the technology in this area, so there are concerns (and active lawsuits) about the legality of the training data, ownership and copyright”.<sup>75</sup> While making teachers and students part of the decision-making process is a praiseworthy endeavour,

<sup>64</sup> Manchester Metropolitan University, FOI request response of 8 September 2023.

<sup>65</sup> University of Northampton, FOI request response of 11 September 2023.

<sup>66</sup> University of Greenwich, FOI request response of 13 September 2023.

<sup>67</sup> ‘Russell Group principles on the use of generative AI tools in education’ (Russell Group, 4 July 2023), [https://russellgroup.ac.uk/media/6137/rg\\_ai\\_principles-final.pdf](https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf).

<sup>68</sup> ‘Russell Group principles on the use of generative AI tools in education’ (Russell Group, 4 July 2023), [https://russellgroup.ac.uk/media/6137/rg\\_ai\\_principles-final.pdf](https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf) [1.1], which reductively refers to these issues as plagiarism, but they are much wider and go to the core of copyright law.

<sup>69</sup> UNESCO, *Guidance for generative AI in education and research* (UNESCO 2023) 36.

<sup>70</sup> Magali Eben et al., ‘Priorities for generative AI regulation in the UK: CREATE response to the Digital Regulation Cooperation Forum (DRCF)’ (CREATE Working Paper 2023/8) 9, <https://eprints.gla.ac.uk/306163/1/306163.pdf>.

<sup>71</sup> University of Edinburgh, ‘Generative Artificial Intelligence guidance for students’, effective as of March 2023, <https://www.ed.ac.uk/bayes/ai-guidance-for-staff-and-students/ai-guidance-for-students>.

<sup>72</sup> University of Edinburgh, FOI response of 23 October 2023.

<sup>73</sup> For example, “if students use an AI tool to gain an unfair advantage by misrepresenting its work as their own, they risk academic misconduct/academic integrity proceedings” (University of West London, FOI response of 14 September 2023); “[s]ummative assessments must show students' original work and cannot be derived solely from AI generated text” (Anglia Ruskin University, FOI response of 20 September 2023); “[a]ny individual who does not act with academic integrity in their assessed work is considered to have committed academic misconduct. This includes... The use of AI to generate answers which are claimed to be original.” (Harper Adams University, FOI response of 21 September 2023); “[t]he University requires that all work you submit is your own” (Open University, FOI response of 20 September 2023); “a student risks a plagiarism allegation in using generative AI services through submitting work that is not their own” (UWE Bristol, FOI response of 5 October 2023).

<sup>74</sup> Ulster University, FOI response of 25 October 2023.

<sup>75</sup> Durham University, FOI response of 21 September 2023.

it is crucial that the university retains full responsibility for the adoption of GenAI tools whose legality is still unclear.

We would like to put to the test the assertion that content created by students using GenAI is not their own original work. In doing this, we answer the call of those authors that underline that HEI policies fail to understand how GenAI challenges our notion of originality “in a time when knowledge production becomes increasingly distributed, collaborative and mediated by technology”.<sup>76</sup> In copyright terms, this begs the question of whether the relevant works can be regarded as the student’s own intellectual creation. The intuitive notion that GenAI content is not original is challenged by the terms of service of the main providers. Pursuant to OpenAI’s Terms of Use, end-users “own the Output”,<sup>77</sup> and accordingly assign to the user any of the company’s “right, title, and interest, if any, in and to Output”.<sup>78</sup> Similar clauses can be found in the terms of service of Google’s Gemini<sup>79</sup> and Github’s Copilot.<sup>80</sup> These provisions may have a twofold explanation. First, the business model depends on the network effects that allow these companies to monetise through premium subscriptions by consumers (e.g. GPT-4)<sup>81</sup> and by integrating the service into third-party products (e.g. Microsoft Bing reaching unprecedented popularity thanks to Prometheus, which is built on GPT-4).<sup>82</sup> In other words, if the GenAI provider retained ownership over the output, users would be disincentivised from using the tool, as there would be limits to how they could re-use the output. This was also the rationale for the Nanshan District Court of Shenzhen in *Tencent v Yingxun*<sup>83</sup> to recognise the copyrightability of AI-generated works.<sup>84</sup> Second, even though ownership and liability are distinct concepts, the

claim not to own the content generated by their tools could be used by GenAI providers to attempt a liability disclaimer, should the content end up infringing on third party rights (e.g. defamation,<sup>85</sup> GDPR violations,<sup>86</sup> and of course copyright infringement, as we will see in IV.B).

There is a clear divergence between universities, that tend to see AI-generated work as not the student’s own intellectual creation, and the main GenAI providers, who maintain the opposite view. Much of the scholarly debate has considered whether copyright *should* protect AI-generated works, with some convincing arguments being in favour of both public domain<sup>87</sup> and sui generis regimes.<sup>88</sup> Here the question is different, should a policy emerge in favour of copyright over these works, how would this play out in practice in the education domain?

Of the requirements for copyright subsistence in Europe and the UK, only originality appears to be problematic here.<sup>89</sup> The typical assignment (e.g. essay) qualifies as a literary work and as such it is protected by copyright only if it is the student’s own intellectual creation. This means that the submission that the student prepared using GenAI must reflect their “personal touch”,<sup>90</sup> that they must have been able to express their “creative ability”<sup>91</sup> by making “free and creative choices”<sup>92</sup> in such a way that the submission reflects their “personality”.<sup>93</sup>

At a first glance, one could posit that the student can make free and creative choices between what prompts to use, and it would follow that the submission may reflect the student’s personality and as such be regarded as original. The scenario would not be dissimilar to that of photographs, as noted also by the Beijing Internet Court in *Li Yunkai v Liu Yuanchun*.<sup>94</sup> Indeed, the angle of the shot, the field of view, elements which the photographer created (positioning of the object, lighting, focus), and

<sup>76</sup> Jiahui Luo (Jess), ‘A critical review of GenAI policies in higher education assessment: a call to reconsider the “originality” of students’ work’ [2024] *Assessment & Evaluation in Higher Education* 1, DOI: 10.1080/02602938.2024.2309963.

<sup>77</sup> OpenAI Terms of Use, as effective on 31 January 2024, <https://openai.com/policies/terms-of-use>.

<sup>78</sup> OpenAI Terms of Use, as effective on 31 January 2024, <https://openai.com/policies/terms-of-use>.

<sup>79</sup> “Your content remains yours, which means that you retain any intellectual property rights that you have in your content. For example, you have intellectual property rights in the creative content that you make” (Google Terms of Service, effective as of 5 January 2022, <https://policies.google.com/terms>).

<sup>80</sup> “GitHub does not claim any ownership rights in Suggestions. You retain ownership of Your Code” (GitHub Copilot Product Specific Terms [2], effective as of September 2023, <https://github.com/customer-terms/github-copilot-product-specific-terms>). “Suggestions” means the code, functions, and other output returned to the user by GitHub Copilot ([7]).

<sup>81</sup> See Dominik K Kanbach et al. ‘The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective’ (2023) *Rev Manag Sci* 1, 14.

<sup>82</sup> ‘Microsoft shares up 8.3% as AI features give a boost to sales’ (*The Guardian*, 26 April 2023), <https://www.theguardian.com/technology/2023/apr/25/microsoft-shares-up-chatgpt-bing-search-engine>.

<sup>83</sup> Judgement No. 14010/2019 *Tencent Computer Company v Yingxun Technology* (Nanshan District Court of Shenzhen, 24 December 2019).

<sup>84</sup> Indeed, this was seen as an incentive for “people to use AI to carry out creative activities” (one of the judges in the *Tencent* trial, as reported by Zhe Dai and Jin Banggui, ‘The copyright protection of AI-generated works under Chinese law’ (2023) *13 Juridical Trib* 241, 255).

<sup>85</sup> See e.g. *Walters v OpenAI, L.L.C.*, 1:23-cv-03122, (N.D. Ga.), the first lawsuit in which OpenAI was accused of defamation (ChatGPT had claimed the plaintiff had been accused of embezzlement and fraud).

<sup>86</sup> Famously, Italy’s data protection authority was the first to issue an injunction to temporarily suspend ChatGPT’s operation due to privacy concerns (Garante per la Protezione dei Dati Personali, provvedimento 30 March 2023 n.112 [doc. web n.9870832]) and in January 2024 it notified OpenAI its findings of GDPR breaches.

<sup>87</sup> Amongst others, see Patrick Zurth, ‘Artificial Creativity? A Case against Copyright Protection for AI-Generated Works’ (2020-2021) 25(2) *UCLA JL & Tech* i.

<sup>88</sup> See e.g. David Komuves, Jesus Niebla Zatarain, Burkhard Schafer, Laurence Diver, ‘Monkeying around with Copyright—Animals, AIs and Authorship in Law’ (CREATE Working Paper 2015/01, February 2015), <https://www.create.ac.uk/publications/monkeying-around-with-copyright-animals-ais-and-authorship-in-law/>.

<sup>89</sup> The other requirements being that (i) for something to be protected should fall within one of the statutory categories of ‘works’ (e.g. literary works, though in recent years the *numerus clausus* doctrine has been called into question, see e.g. *Case C-310/17 Levola Hengelo v Smilde Foods* EU:C:2018:899 and *Response Clothing Ltd v The Edinburgh Woollen Mill* [2020] EWHC 148); (ii) the work has to be recorded in writing or otherwise (also interpreted extensively, see e.g. *Wright v BTC Core* [2023] EWHC Civ 868); (iii) copyright is territorial both in terms of authorship (so-called qualification) and in terms of infringement (see *Getty Images v Stability AI* [2023] EWHC 3090 (Ch)); (iv) copyright has a limited duration (hence the release in 2023 of the film *Winnie the Pooh: Blood and Honey*, as the book by A.A. Milne lapsed into the public domain in 2022).

<sup>90</sup> *Painer v Standard Verlags GmbH* Case C-145/10 EU:C:2011:798 [92].

<sup>91</sup> *SI, Brompton Bicycle Ltd v Chedech / Get2Get* Case C-833/18 EU:C:2020:461 [34].

<sup>92</sup> *Football Association Premier League v QC Leisure* Cases C-403/08 and C-429/08 EU:C:2011:631 [98]; *Painer v Standard Verlags GmbH* Case C-145/10 EU:C:2011:798 [94]; *Case Land Nordrhein-Westfalen v Dirk Renckhoff* EU:C:2018:634 [14]; *Cofemel - Sociedade de Vestuário SA v G-Star Raw CV* Case C-683/17 EU:C:2019:721 [30]; *SI, Brompton Bicycle Ltd v Chedech / Get2Get* EU:C:2020:461 [34]; *Response Clothing Ltd v The Edinburgh Woollen Mill Ltd* [2020] EWHC 148 (IPEC) [59].

<sup>93</sup> *SI, Brompton Bicycle Ltd v Chedech / Get2Get* EU:C:2020:461 [34].

<sup>94</sup> Judgement No. 11279/2023 *Li Yunkai v Liu Yuanchun* (Beijing Internet Court, 27 November 2023).

elements arising from being at the right place at the right time can contribute to the photograph's originality.<sup>95</sup> It could be objected, however, that the student is further removed from the AI-generated output, compared to the photographer, to the point that the creative link is ultimately broken. The student's originality expresses itself in the prompt and how it is engineered to generate more and more refined answers. While this justifies ownership of the prompt, it is questionable whether this can extend to the ensuing answer, which is entirely dependent on the GenAI tool's automated operation. One of the leading cases in EU copyright law can be of assistance. As held in *Bezpečnostní softwarová asociace v Ministerstvo Kultury* (C-393/09) EU:C:2010:816,<sup>96</sup> where the expression is dictated by technical considerations and the different methods of implementing an idea are so limited that the idea and the expression become indissociable, both become unprotectable. Transformers (advanced neural networks for natural language processing tasks) work by predicting the most likely 'token' (typically a word in a sentence) given the context. Accordingly, it seems unlikely that in entirely AI-generated submissions ideas and expressions become indissociable and therefore unprotectable.<sup>97</sup>

The final question is whether the student submissions could be regarded as a computer-generated work. The UK Copyright, Designs and Patents Act 1988 introduced the legal fiction whereby when a work is generated by a computer without a human author, the person who made the arrangements necessary for the creation of the work is taken to be the author.<sup>98</sup> There is little guidance on what the 'necessary arrangements' may be, but it does seem that, depending on the case, they might cover the person that operates the computer (e.g. the student), the software programmer,<sup>99</sup> the devisers of the algorithm, or the person(s) responsible for the dataset used to train the machine. *Nova Productions v Mazooma* remains the only case to apply the provision on computer-generated works, and it did so with regard to images generated playing a videogame i.e. not AI. The fact that it has never been applied again in court indicates that the provision is not fit for AI.<sup>100</sup> Current AI refers to technologies that are much more complex than the ones the lawmaker had in

mind in 1988, they are increasingly autonomous (especially when reinforcement learning is used),<sup>101</sup> and with incredibly complex supply chains. Finally, because "the European standard [of originality] now applies to all works [...] no computer-generated work can be protected by copyright".<sup>102</sup> After years of uncertainty as to whether this originality standard replaced the old English "skill, labour, or judgement" standard, and whether these standards are in fact different, the recent case *THJ Systems Ltd v Sheridan*<sup>103</sup> answered both questions affirmatively. However, the EU-UK convergence in IP law is under threat as recent legislation made it easier to depart from EU-derived domestic law,<sup>104</sup> and some courts are already diverging from EU IP jurisprudence.<sup>105</sup> Arguably, GenAI outputs might be regarded as the end-user's original creation should the UK revert to the old "skill, labour, and judgement" standard as the generation of good-quality answers does require considerable labour, and potentially skill and judgement. On the one hand, the UK has the aspiration to become a leader in the field of AI regulation, which may provide an incentive to make those necessary changes to allow the copyrightability of GenAI outputs. On the other hand, this policy choice would go against the copyright's *raison d'être*, i.e. rewarding intellectual labour to incentivise human creativity.<sup>106</sup> The onus to prove the need for AI copyright authorship lies with the proponents, and insufficient evidence has been presented so far.<sup>107</sup>

Confusion stems from the divergence between the universities' view on GenAI content ("*AI does not respect copyright*"<sup>108</sup> as Dundee University puts it), the practices of GenAI providers (regarding the content as the students' own), and copyright law (where there is no European or UK case law on AI-generated works, and the scholarly debate has not reached a definitive conclusion). Even though authorship in a copyright sense does not equate authorship for academic integrity purposes, this confusion calls for a lenient approach to integrity. Indeed, it would seem unfair to tarnish the students' record, as they would have a good case to claim that they considered the GenAI content to be their own.

<sup>95</sup> *Temple Island Collections v New English Teas Ltd* [2012] EWPCC 1; *Antiquesportfolio.com plc v Rodney Fitch & Co Ltd* [2001] FSR 345.

<sup>96</sup> *Bezpečnostní Softwarová Asociace - Svaz Softwarové Ochrany v Ministerstvo Kultury* (C-393/09) EU:C:2010:816 [48], [49].

<sup>97</sup> There is little doubt that, if GenAI is used as a mere aid i.e. if the submission is substantially the result of the student's intellectual work, using and acknowledging multiple sources, then the work would be the student's own, unless the university claims ownership over its students' works, which is usually not the case (see e.g. University of Oxford's Statute XVI, Part B).

<sup>98</sup> CDPA, ss.9(3) and 178.

<sup>99</sup> As was the case in *Nova Productions v Mazooma* [2006] RPC 14.

<sup>100</sup> Intellectual Property Office, 'Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation' (*UK Government*, 28 June 2022) [22], <https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation>.

<sup>101</sup> In reinforcement learning models, there is no need for humans to collect the training datasets, as the system learns autonomously as it perceives its environment.

<sup>102</sup> Lionel Bently, Brad Sherman, Dev Gangjee, and Phillip Johnson, *Intellectual Property Law* (6th edn, OUP 2022) [3.7.8].

<sup>103</sup> *THJ Systems Ltd* [2023] EWCA Civ 1354.

<sup>104</sup> For example, retained direct EU legislation will be 'downgraded' so that it could be amended by secondary legislation (The Retained EU Law (Revocation and Reform) Act 2023).

<sup>105</sup> Under the Retained EU Law (Revocation and Reform) Act 2023 s.6, it has become easier to depart from retained EU case law. We have seen a departure in trademark law in *Industrial Cleaning Equipment v Intelligent Cleaning Equipment* [2023] EWCA Civ 1451, where the Court of Appeal departed from a decision of the CJEU (Case C-482/09 *Budejovický Budvar np v Anheuser-Busch Inc* [2011] ECR I-8701).

<sup>106</sup> See e.g. Anthoula Papadopoulou, 'Creativity in crisis: are the creations of artificial intelligence worth protecting?' (2021) 12(3) JIPITEC 408.

<sup>107</sup> In this sense, see Martin Kretschmer, Bartolomeo Meletti, Luis H Porangaba, 'Artificial intelligence and intellectual property: copyright and patents—a response by the CREATe Centre to the UK Intellectual Property Office's open consultation' (2022) 17(3) JIPLP 321.

<sup>108</sup> Dundee University, FOI response of 19 September 2023.

Finally, the determination of authorship (and ownership) of GenAI output has wider relevance. Indeed, the UK Department for Education is calling for education institutions not to “allow or cause [IP], *including pupils’ work*, to be used to train”<sup>109</sup> GenAI models. GenAI tools tend to use the inputs submitted by users, including students, to further train and refine their models.<sup>110</sup> If students are not found to own the AI content of their submissions their interactions with these systems risk being used as training data. This makes it important for HEIs and schools to help students opt out under art.4(3) of the CDSM Directive.<sup>111</sup>

### ***Do students expose themselves to liability for copyright infringement by using AI-generated content?***

Authorship and infringement are two distinct issues in copyright law. One could be infringing third party copyright (e.g. by using the substantial part of someone else’s work) and yet be the author of the new work (unless there is no original content in the latter, not even in the way that the non-original components were selected and assembled). Therefore, the fact that a student may be regarded as the author of GenAI content leaves open the question of whether they could be held liable for copyright infringement.

To answer the question, one needs to distinguish between input and output.<sup>112</sup> The former refers to the unauthorised use of copyright works to train GenAI models. This has been identified as one of the key concerns warranting regulatory intervention,<sup>113</sup> as most of the litigation and scholarly debate focus on the input issue.<sup>114</sup> The output issue refers to GenAI answers reproducing substantial parts of copyright works, as

contended in *New York Times v Microsoft*.<sup>115</sup> This will be the focus of this section, as students could only be held liable for the illegality of the GenAI output.

The problem of whether students can be held liable for copyright infringement because they use AI-generated content in their assignments is pressing for two reasons. First, the reproduction of protected material will often result from the so-called memorisation, whereby GenAI output is a replica of the input i.e. if prompted appropriately, they will emit the memorised training data potentially word by word. This often depends on multiple copies of the same content being used to train a model.<sup>116</sup> For example, in *New York Times v Microsoft* many passages of the newspaper’s articles that ChatGPT reproduces are particularly popular articles that have been republished several times. Second, universities are trying to ensure academic integrity by asking students to submit an ‘AI statement’ alongside their submissions. This may include the exact reproduction of the responses provided by the GenAI tools to the student’s prompts. For example, the Open University recommends the inclusion of a copy of the GenAI output as an appendix to the student’s submission.<sup>117</sup> This appendix shall include a list of the prompts and the verbatim reproduction of the output.<sup>118</sup> While we do not contest to pedagogical value of such AI statements, one should consider whether they may have the unintended consequence of exposing our students to copyright infringement.

For copyright infringement to arise, three requirements are to be made out. First, the student would have to carry out a ‘restricted act’ without the rightsholder’s permission.<sup>119</sup> Let us imagine that they have copied and pasted the GenAI output into their AI statement. This will undoubtedly qualify as a reproduction of that content.<sup>120</sup> The student may put forward that this reproduction is not unauthorised as we have seen that the terms and

<sup>109</sup> Department of Education, ‘Generative artificial intelligence (AI) in education’ (UK Gov, 26 October 2023), <https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education>.

<sup>110</sup> Department of Education, ‘Generative artificial intelligence (AI) in education’ (UK Gov, 26 October 2023), <https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education>.

<sup>111</sup> The general-purpose TDM exception applies only “on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their right-holders in an appropriate manner” (CDSM Directive art.4(3)). It is becoming apparent that Big Tech are already trying to impose their proprietary standards as defaults for the opt-outs, and that this should be resisted (Paul Keller, ‘Generative AI and copyright: Convergence of opt-outs?’ (*Kluwer Copyright Blog*, 23 November 2023), <https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/>).

<sup>112</sup> Andres Guadamuz, A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs, GRUR International, 2024, ikad140, <https://doi.org/10.1093/grurint/ikad140>. While we do not necessarily agree that most ‘input’ cases will be decided in favour of the defendant (the text and data mining exceptions in Europe and the UK are quite narrow, unlike the US fair use), we do concur that the case about the legality of the output will have a higher rate of success.

<sup>113</sup> Andres Guadamuz, A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs, GRUR International, 2024, ikad 140, <https://doi.org/10.1093/grurint/ikad140>.

<sup>114</sup> *J. Doe 1 and J. Does 2 v Github*, Case 3:22-cv-06823 (US District Court for the Northern District of California); *Sarah Andersen v Stability AI*, Case 3:23-cv-00201-WHO (US District Court for the Northern District of California); *Paul Tremblay v OpenAI*, Case 3:23-cv-03223 (US District Court for the Northern District of California); *Sarah Silverman v OpenAI*, Case 3:23-cv-03416 (US District Court for the Northern District of California); *Michael Chabon v OpenAI*, Case 3:23-cv-04625-PHK (US District Court for the Northern District of California); *Concord Music Group Inc v. Anthropic PBC*, No. 3:23-cv-01092 (US District Court for the Middle District of Tennessee); *Authors Guild v OpenAI*, Case 1:23-cv-08292 (US District Court for the Southern District of New York); *The New York Times v Microsoft, OpenAI and others*, Case 1:23-cv-11195 (US District Court for the Southern District of New York).

<sup>115</sup> *The New York Times v Microsoft, OpenAI and others*, Case 1:23-cv-11195 (US District Court for the Southern District of New York).

<sup>116</sup> Nicholas Carini et al., ‘Quantifying Memorization Across Neural Language Models’ [2022] arXiv:2202.07646. For an initial consideration of memorisation and copyright, see Andres Guadamuz, ‘A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs’ (2024) GRUR Int ikad140.

<sup>117</sup> ‘Generative AI for students’ (Open University), <https://about.open.ac.uk/policies-and-reports/policies-and-statements/gen-ai/generative-ai-students>.

<sup>118</sup> ‘Generative AI for students’ (Open University), <https://about.open.ac.uk/policies-and-reports/policies-and-statements/gen-ai/generative-ai-students>.

<sup>119</sup> CDPA s.16(2).

<sup>120</sup> Reproduction is only one of the restricted acts that, if carried out without the rightsholder’s permission, may constitute copyright infringement. Copying and pasting GenAI answers is likely to be regarded as a form of reproduction. Of the other restricted acts, adaptation is likely to play an important role (CDPA ss.16 and 21). Adaptation includes translation (CDPA, s.21(2)(a)(i)) and it is becoming increasingly clear that one of main risks in terms of academic integrity is the use of GenAI as a translation tool, e.g. when students who speak foreign languages use these tools to translate into English in a way that would be incredibly difficult to detect (see e.g. Lancaster University, ‘Using AI in your learning and assessment’ [2024], <https://portal.lancaster.ac.uk/ask/study/developing-academic-skills/using-ai-in-your-learning-and-assessment/>). Communication to the public would be less relevant because it requires the communication to a new public i.e. a public not considered at the time of initial publication. As assignments are shared only with the marking team, the examiner, and few colleagues in professional services, this is not a new public as they all have free access to the main GenAI tools (See *Svensson v Retriever Sverige AB* Case C-466/12 EU:C:2014:76 [27]). The outcome would be different in the event of use of premium paid-for tools.

conditions of the most popular GenAI tools go as far as considering the students (and users more generally) to be the owners of the output. However, if the GenAI providers do not own the output—as they would not in the event of verbatim reproduction due to memorisation—then one could question whether the student could consider themselves to be authorised as no one can transfer a right that they themselves do not have.<sup>121</sup>

Second, the rightsholder would have to prove that the student reproduced a substantial part of a protected work.<sup>122</sup> It is immaterial whether the GenAI content constitutes the substantial part of the student's submission; indeed, substantiality is assessed with reference to the claimant's work<sup>123</sup>. As held in *Designers Guild v Russell Williams* [2000] 1 W.L.R. 2416, “whether the part taken is substantial must be determined by its quality rather than its quantity”, which means asking whether the student appropriated “that which conferred originality on the claimant's copyright work (or a substantial part of it)”,<sup>124</sup> regardless of how many words the student copied.<sup>125</sup> If the GenAI output is not original as seen above, then its reproduction would not be problematic. The issue would arise in the event of verbatim reproduction of third party works due to memorisation.

The third requirement is the causal link: did the student derive their submission or AI statement from the rightsholder's work? While direct evidence of copying, is rare,<sup>126</sup> the similarities between the student's AI statement and an article of the New York Times would be part of the indirect evidence necessary to prove derivation, but it would not suffice on its own.<sup>127</sup> The crux of the matter is the question of access, which in our scenario would equate to asking whether the student had seen the original work. The reasonable conviction not to be infringing or the mere ignorance are immaterial as primary infringement does not require intention or

knowledge to infringe, to the point that subconscious copying may amount to an infringement of copyright.<sup>128</sup> If the student has a record of having plagiarised content in the past, this may have a bearing in copyright infringement proceedings, as courts consider the claimant's propensity to copy, as seen in *Sheeran v Chokri*.<sup>129</sup> Often the success of the claimant's work is used as evidence that it is likely that the defendant had access to the former's work.<sup>130</sup> As the issue of memorisation occurs especially with particularly popular sources, this could point in the direction of access. However, when there are obvious similarities between the works and undoubted access to the claimant's work, the burden of proof shifts on the defendant to prove that the two works are not causally connected.<sup>131</sup> In our scenario, the student could in all likelihood use the screenshots and the metadata of the GenAI tool to prove that they sourced the material from said tool as opposed to the original source. A final defence for the student could be grounded in the claim by many scholars<sup>132</sup> that copyright should protect only works that are used for their expressive or communicative content. If a student includes content in the AI statement precisely to evidence the use of GenAI content, the argument could be made this use should not be caught by copyright infringement.<sup>133</sup>

Courts may interpret the concept of ‘subconscious copying’ broadly, as encompassing the unwitting copy from a copy.<sup>134</sup> In that case, it will become crucial to have access to the training datasets. A positive innovation on this front is being introduced by the AI Act, which will oblige the providers of general-purpose AI models to “draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model”.<sup>135</sup> Although much will depend on how detailed the summary will be (e.g. what roles will trade secrets play),<sup>136</sup> and how proactive the AI

<sup>121</sup> *Nemo plus iuris ad alium transferre potest quam ipse haberet* (Ulpian, *Book 46 ad edictum*, D 50, 17, 54) The principle applies also to IP transfers, where the attempt to transfer third party copyrights would qualify as failure to perform, rather than illegality of contract or frustration, according to Andreas Rahmatian, ‘Contracts infringing intellectual property rights’ (2003) 4 IPQ 411, 429.

<sup>122</sup> CDPA s.16(3).

<sup>123</sup> *Designers Guild v Russell Williams* [2000] 1 WLR 2416, 2426.

<sup>124</sup> *Mitchell v BBC* [2011] EWPCC 42 [28] per HHJ Birss QC.

<sup>125</sup> Quantity matters only to some extent. We know that there is no copyright in single words (*de minimis* rule, *Exxon Corp v Exxon Consultants* [1982] Ch. 119) and that, in principle, 11 words may constitute the protectable original expression of an idea (*Infopaq International A/S v Danske Dagblades Forening* Case C-5/08 [2009] ECR I-6569 [48]) but it cannot be predicted with certainty what is the point where ‘quantity’ becomes ‘quality’.

<sup>126</sup> *IBCOS Computers v Barclays Bank* [1994] F.S.R. 275.

<sup>127</sup> This notwithstanding, the more similar two works are, the more cogent evidence the defendant will need to present for the rebuttal (*Mitchell v BBC* [2011] EWPCC 42 [25]).

<sup>128</sup> *Francis Day & Hunter v Bron* [1963] 1 Ch. 587, 614. This is the key difference between primary and secondary infringement, with the latter arising when commercially dealing in articles which the defendant knows or has reason to believe are infringing copies (CDPA s.23).

<sup>129</sup> *Sheeran v Chokri* [2022] EWHC 827 (Ch) [148], following *Mood Music Publishing Co Ltd v De Wolfe Ltd* [1976] Ch. 119, 127.

<sup>130</sup> *Sheeran v Chokri* [2022] EWHC 827 (Ch) [78], [83], [146].

<sup>131</sup> *Temple Island Collections Ltd v New English Teas Ltd* [2012] EWPCC 1 [55].

<sup>132</sup> See, more prominently, Abraham Drassinower, *What's Wrong with Copying?* (HUP, 2015) esp Ch.3 and Alain Strowel, ‘Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals’, in P. Bernt Hugenholtz (ed), *Copyright Reconstructed: Rethinking Copyright's Economic Rights* (Wolters Kluwer, 2018) 203.

<sup>133</sup> By way of example, Lionel Bently, Brad Sherman, Dev Gangjee, and Phillip Johnson, *Intellectual Property Law* (6th edn, OUP 2022) [2] refer to a scenario where “a person might digitize a work in order to conduct a search (e.g. to detect plagiarism), or to allow others to search the text (by creating indices for search engines)”.

<sup>134</sup> Abbe Brown, Smita Kheria, Jane Cornwell, and Marta Iljadica, *Contemporary Intellectual Property: Law and Policy* (6th edn, OUP 2023) [4.29] put forward that it would be “irrelevant that the defendant did not know or was unaware of the existence of the original work (as might be the case, for example, where he was copying from a copy of it) or of the fact that the original work had copyright”. This may be one of the areas where GenAI may lead to a rethinking of existing copyright law.

<sup>135</sup> AI Act art.52c(1)(d).

<sup>136</sup> There is the risk that the summary will remain quite high-level, as might follow from the requirement to duly consider “the need to protect trade secrets and confidential business information” (recital 60k). On the dangers of IP balancing in data governance see Guido Noto La Diega ‘Ending Smart Data Enclosures: The European Approach to the Regulation of the Internet of Things between Access and Intellectual Property’, in Stacy-Ann Elvy and Nancy Kim (eds), *The Cambridge Handbook on Emerging Issues at the Intersection of Commercial Law and Technology* (Cambridge University Press, 2024).

Office will be in monitoring compliance,<sup>137</sup> the UK and other countries internationally should consider transplanting this provision, as it is become clear from the current US litigation that much will depend on whether claimants can have access to the GenAI training datasets.<sup>138</sup>

From an educational perspective, UK universities may have been too quick in pushing students to embrace GenAI, when there is the possibility that the use of these tools will expose them to potential liability for copyright infringement. From a copyright perspective, it can be questioned whether existing rules around access and subconscious copying are fit for GenAI. More fundamentally, we may be witnessing a shift of copyright from property rules to liability rules.<sup>139</sup> The evidence is threefold. First, under the general-purpose text and data mining exception introduced by the CDSM Directive, copyright content can be used to train GenAI models unless the rightsholders opt out. This means that the default is that copyright content can be used without the author's permission. Second, the pragmatic calls for the introduction of a levy system<sup>140</sup> point in the direction that what matters is no longer the proprietary expectation that one's rights are in principle exclusive: what matters is the circulation of the knowledge commodity insofar as monetisation mechanisms are in place. Third, in November 2023 the CEO of OpenAI announced their new copyright strategy: rather than removing copyright material from its datasets, the company will indemnify customers should they be accused of copyright infringement ("copyright shield"<sup>141</sup>).

## Conclusion

With GenAI, we are witnessing an unprecedented shift of emphasis from privacy to copyright. The manifold explanations for this shift relate to broader changes, from the rise of data as an IP subject (database rights, text and data mining, etc.) through to the fact that both GenAI and copyright go to the heart of what it means to be human. Copyright is taking centre stage in AI-related lawmaking—most notably with the copyright provisions in the AI Act, which were absent in the pre-ChatGPT version—and litigation—with lawsuits such as *New York Times v Microsoft* making many commentators fear or hope that copyright will spell the end of GenAI.

The analysis of our FOI request data suggests a disconnect. On the one hand, universities are, taking a positive stance with regard to GenAI in education and issuing policies, guidelines, etc. to ensure the responsible use of these tools. On the other hand, copyright is still often overlooked and misunderstood. We urge HEIs to reconsider, as there can be no responsible GenAI without adequate copyright governance, guidance, and training. This is especially true since the law on the matter is far from settled, as our case study analysis of authorship and infringement shows.

Determining authorship of GenAI content is important for two reasons. First, the divergence between the universities' view on GenAI content, the practices of GenAI providers, and copyright law justifies a lenient stance on academic integrity, as penalising students could be deemed unjust given their plausible belief that they do own the GenAI content. Second, the increasing difficulty in determining originality, and consequently authorship, may be an indication that we need to re-envision what it means to create knowledge in these increasingly technology-mediate times. Third, GenAI tools tend to use the inputs submitted by users, including students, to refine their models. If students do not own the GenAI content, their interactions with these systems are at a higher risk of being used as training data. We recommend that HEIs support students in opting out under art.4(3) of the CDSM Directive to address this concern.

As to infringement, UK universities rushed to embrace GenAI, but there is the possibility that the use of these tools will accidentally expose their students to liability for copyright infringement. While the scenarios that we presented regarded educational copyright, our findings have wider relevance. Indeed, GenAI provides an opportunity to rethink existing rules around access and subconscious copying. More fundamentally, we may be witnessing a shift of copyright from property rules to liability rules, where creators have no longer a say as to whether their content can be used to train GenAI models: private-copy-levy-like compensation is set to become the default. Thus, GenAI may bring to completion the long process—resulting from the joint operation of high-speed technology, non-tech-savvy lawmaking and powerful private stakeholders—of loss of control over, and ultimately privatisation of knowledge.

<sup>137</sup> This is still unclear, as the AI Act provides that the "AI Office should monitor whether the provider has fulfilled those obligations without verifying or proceeding to a work-by-work assessment of the training data in terms of copyright compliance" (Recital 60ka).

<sup>138</sup> See *Andersen v. Stability AI Ltd*, Case 3:23-cv-00201-WHO, Order on motion to dismiss by Judge William H. Orrick. (District Court, N.D. California, 30 October 2023).

<sup>139</sup> The reference is to the distinction elucidated by Guido Calabresi and A. Douglas Melamed, 'Property rules, liability rules, and inalienability: one view of the cathedral' (1972) 85(6) *Harvard Law Review* 139. To put it simply, a property rule is in place if an entitlement (typically a good) can only be transferred through a voluntary transaction where the price is agreed by the seller; a liability rule is in place when an entitlement can be removed without the initial holder's consent and the state tolerates it as long as the person who removed it pays an objectively determined sum for it (typically through compensation). Traditionally, property rules are a sign that the State recognises a higher importance to the relevant interests involved.

<sup>140</sup> Martin Senfleben, 'Generative AI and Author Remuneration' (2023) 54 IIC 1535; Giancarlo Frosio, 'Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity?', in Enrico Bonadio and Caterina Sganga (eds), *A Research Agenda for EU Copyright Law* (Edward Elgar 2024).

<sup>141</sup> Blake Montgomery, 'OpenAI offers to pay for ChatGPT customers' copyright lawsuits' (*The Guardian*, 6 November 2023), <https://www.theguardian.com/technology/2023/nov/06/openai-chatgpt-customers-copyright-lawsuits>.