Published on 10 February 2022

DATA ANALYTICS IN FINTECH AND THE NEW COPYRIGHT ACT

[2022] SAL Prac 4

David **TAN**[1]
PhD (Melbourne), LLM (Harvard), LLB (Hons), BCom (Melbourne); Professor, Faculty of Law, National University of Singapore.

I. Fintech and data analytics

1 It has been well chronicled that regulatory and technological developments are changing the nature of financial markets, services and institutions in completely unexpected ways.[2] Fintech – a portmanteau of "financial" and "technology" – is a widely recognised term that refers to the confluence of financial services and information technology whose interlinkage is mutually reinforcing.

2 Douglas Arner *et al* have noted that with the development of new technology such as artificial intelligence ("AI") and deep learning, there seems to be great potential for more to be done in terms of automating market supervision, consumer protection and prudential regulation.[3]

3 While the issues in fintech today tend to centre on financial inclusion, data protection, risk management and other

1 I would like to thank Dr Chuanman You, Post-Doctoral Fellow at EW Barker Centre for Law & Business, for his invaluable comments.

2 For example, Douglas W Arner, Jànos Barberis & Ross P Buckley, "FinTech, RegTech, and the Reconceptualization of Financial Regulation" (2017) 37 *Northwestern Journal of International Law & Business* 371; Douglas W Arner, Jànos Barberis & Ross P Buckley, 'The Evolution of FinTech: A New Post-Crisis Paradigm?' (2016) 47 Georgetown J Int'l L 1271.

3 Douglas W Arner, Jànos Barberis & Ross P Buckley, "FinTech, RegTech, and the Reconceptualization of Financial Regulation" (2017) 37 *Northwestern Journal of International Law & Business* 371 at 397. See also Matthew U Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) 29 Harvard Journal of Law & Technology 354 at 359.

regulatory compliance concerns, the vast amount of data mining and analytics performed in the fintech industry may just be falling afoul of copyright law.[4]

4 Generally, data analysis can provide significantly valuable insights into consumers' profiles and allow financial institutions, as well as non-bank entities such as AliPay and P2P lending platforms, to better develop and customise their offerings.[5] There are different types of analytics. For instance, predictive analytics as a form of advanced analytics seeks to use data and information to answer the question "What is likely to happen?". It often involves techniques such as regression analysis, pattern matching, predictive modelling and forecasting.[6] Predictive analytics can bring significant value to a company: if a company can accurately identify which action caused a certain outcome, it can then more reliably predict which course of action would achieve a desired result. Another example is prescriptive analytics, which is a method of analytics that analyses information to answer the question "What should be done?". This type of analytics is characterised by techniques that include graph analysis, simulation, complex event processing, neural networks, recommendation engines and machine learning.[7]

5 Through the use of technology to monitor and analyse markets, trends and consumer behaviour, algorithms and bots will inevitably be searching on the Internet for relevant information contained in reports, commentaries and articles – all of which

4 For some examples of data mining and analytics in fintech, see Shian-Chang Huang *et al*, "Intelligent FinTech Data Mining by Advanced Deep Learning Approaches" (2021) *Computational Economics* (17 April 2021).

5 Through big data analysis, AliPay was reported to be able to process a loan application from a qualified individual or small and medium enterprises within three minutes. See also Chuanman You, "Recent Development of Fintech Regulation in China: A Focus on the New Regulatory Regime for the P2P Lending (Loan-Based Crowdfunding) Market" (2018) 13 *Capital Markets Law Journal* 85 at 92.

6 Dennis D Hirsch, "From Individual Control to Social Protection: New Paradigms for Privacy Law in the Age of Predictive Analytics" (2020) 79 Maryland Law Review 439 at 453–459.

7 Katerina Lepenioti *et al*, "Prescriptive Analytics: Literature Review and Research Challenges" (2020) 50 *International Journal of Information Management* 57 at 70.

are literary and artistic works protected by copyright.[8] While one would often argue that what is in fact accessed is "data" or "facts" for the purpose of performing analysis, as opposed to the creative expression as contained in these authorial works, it is nonetheless "copying" as understood in copyright law. The user of these works will have to argue that such infringing conduct is nonetheless fair dealing or fair use.

II. The new Copyright Act 2021[9]

6 The Singapore Copyright Act was first enacted in 1987 and was largely based on the copyright regimes of the United Kingdom and Australia at that time. Major revisions to the Copyright Act were made in 1998, 1999 and 2004, which ensured that Singapore's copyright regime was aligned to international norms and bilateral treaties, and was relevant to content that was being created, distributed and consumed digitally. A significant public consultation exercise was carried out, which culminated in the introduction of a future-ready Copyright Bill 2021 in July 2021, which sought to replace the Copyright Act 1987 in its entirety. The new Act was passed by Parliament after its second reading on 13 September 2021, and came into force on 21 November 2021. Some key amendments certainly position the legislation to be future-ready as they strike an appropriate balance between protecting rights owners and enabling the public and other users to have access to these works in order to create new ones.

7 This ambitious revamp with wide-ranging reforms included the introduction of the moral right of attribution (or right to be identified), the recognition of an open-ended fair use provision (modelled after the one in the United States), the inclusion of a computational data analysis exception, and a new class licensing scheme to regulate collective management organisations in Singapore.

8 Section 13(1)(*a*) of the Copyright Act 2021 (2020 Rev Ed) states that a "literary work" includes a compilation in any form, while s 20(1)(*a*) states that an artistic work includes a photograph or a drawing.

9 2020 Rev Ed.

8 This commentary will focus on changes to fair use and the introduction of a text and data mining ("TDM") exception known as the computational data analysis exception.[10]

III. Computational data analysis exception and fair use

9 Data analytics is an inevitable technological development of the 21st century, and more specifically predictive analytics can find patterns contained within data in order to detect risks and opportunities. It is applicable to a panoply of activities not just in the finance industry, but also in other industries such as healthcare, retail, pharmaceuticals, automotive, aerospace and manufacturing. The introduction of a computational data analysis exception (new ss 243–244) for TDM appears to be sensible and timely, but may in practice be superfluous in the light of the open-ended fair use provision. Computational data analysis is defined under s 243(*a*) as the use of a computer program to "identify, extract and analyse information or data from the work" – which is synonymous with TDM – and crucially, miners must prove that they have lawful access to online works or data in order to qualify for protection under this provision. The UK has already in place a text and data mining exception – albeit narrower than the Singapore version – that "a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose".[11]

10 There are a number of conditions to be satisfied for the new s 244(2). They include:

> (a) the user proving that the copy is made for the purpose of computational data analysis and not for any other purpose;

10 This commentary is adapted from a longer article: David Tan & Thomas Lee Chee Seng, "Copying Right in Copyright Law: Fair Use, Computational Data Analysis and the Personal Data Protection Act" (2021) 33 SAcLJ 1032.

11 Copyright, Designs and Patents Act 1988 (c 48) (UK) s 29A(1)(a).

(b) the user not supplying the copy to any person other than for the purpose of verifying the results of the computational data analysis carried out by the user;

(c) the user having lawful access to the material (the first copy) from which the copy is made; and

(d) the first copy not being an infringing copy.

11 A "literary work" is defined in the new s 13 as a compilation in any form, and can contain only data that is "an intellectual creation because of the selection or arrangement of its contents". Financial reports that analyse numbers or selections of data presented in a table format can all constitute compilations in which copyright subsists. Hence the scanning, downloading and storage of these works for the purpose of performing analysis would be *prima facie* infringing.

12 The clarification in the new s 183 that permitted uses are not rights infringement is an important one. In addition, the explicit recognition of an open-ended fair use approach in the new s 190, akin to that in the United States, allows the courts to better assess whether the panoply of technological and artistic uses, such as sharing on social media or the commercial production of satire and pastiche, are permitted uses. In *Global Yellow Pages Pte Ltd v Promedia Directories Pte Ltd*, the Court of Appeal had noted that the scope of the former s 35 fair dealing provision may be interpreted to mean that a fair dealing for "any purpose" (as opposed to merely for "research or private study") might be held not to amount to an infringement of copyright, and that "[t]his also made Singapore's fair dealing provisions more similar to its American counterpart, which is more open-textured".[12]

13 Sundaresh Menon CJ, in authoring the unanimous judgment, also hinted at the willingness of the local courts to take greater cognisance of US and Australian decisions in this area. The persuasiveness and relevance of US fair use decisions

12 *Global Yellow Pages Pte Ltd v Promedia Directories Pte Ltd* [2017] 2 SLR 185 at [76] (citing *Authors Guild, Inc v Google, Inc* 804 F 3d 202 at 207 (2nd Cir, 2015)).

were similarly argued in earlier academic articles.[13] In the Court of Appeal's reference to *Authors Guild, Inc v Google, Inc*[14] ("*Authors Guild v Google*"), decided by the US Second Circuit Court of Appeals, it noted that "it was a transformative use (and therefore fair dealing) for Google, a search engine operator, to digitise books and make them searchable (albeit limiting the portion of the search result that users could see) because this 'augment[ed] public knowledge by making available information *about* [the plaintiffs'] books without providing the public with a substantial substitute for matter protected by [the plaintiffs'] copyright interests in the original works or derivatives of them'".[15] The Court of Appeal also commented that "we do not go as far as those cases which suggest that a commercial nature or purpose of the dealing will presumptively be regarded as unfair",[16] perhaps demonstrating a greater tolerance for commercial data mining activities.

14 No TDM case has been litigated in Singapore to date under the former s 35(2) open-ended fair dealing exception (now in ss 190–191 of the Copyright Act 2021). Although the new Copyright Act explicitly allows the fair use defence to be pleaded together with the computational data analysis exception, in practice, litigants may still inevitably find themselves having to choose only one due to the frequent incompatibility of both defences. The miner who seeks to justify that the TDM activity was "transformative" under the open-ended fair use defence, thus obviating the need to seek an *ex ante* licence, could potentially undermine its pleadings (in the alternative under the computational data analysis exception) that the miner had "lawful access" to the original material. Take, for example, the creation of a search engine: works are likely to have been copied without lawful access in order to create a full-text searchable

13 For example, David Tan & Benjamin Foo, 'The Unbearable Lightness of Fair Dealing: Towards an Autochthonous Approach in Singapore' (2016) 28 SAcLJ 124.

14 804 F 3d 202 at 207 (2nd Cir, 2015).

15 *Global Yellow Pages Pte Ltd v Promedia Directories Pte Ltd* [2017] 2 SLR 185 at [81].

16 *Global Yellow Pages Pte Ltd v Promedia Directories Pte Ltd* [2017] 2 SLR 185 at [81].

database, but this would nonetheless have been a highly transformative purpose.

15 In two complementary decisions, the US Second Circuit Court of Appeals found fair use, notwithstanding that the libraries had downloaded and stored complete digital copies of entire books, because such copying was essential to permit searchers to identify and locate the books in which words or phrases of interest to them appeared.[17] If similar facts were encountered in Singapore, as this is a TDM use, the litigant in Singapore may find it difficult to successfully plead the computational data analysis exception, as there may not have been lawful access to *all* of the copyrighted works; the litigant will have a better chance of succeeding on pleading fair use.

16 But a critical consideration in these American precedents, including the Supreme Court's decision in *Google LLC v Oracle America, Inc*[18] in 2021 where fair use was found, was that the public may nonetheless benefit from the unauthorised copying in line with the objective that the Copyright Clause in the US Constitution was designed to promote. Breyer J, delivering the majority opinion, took into account the significant public benefit that such copying by Google in the Android mobile phones would likely produce and held that "[g]iven the costs and difficulties of producing alternative APIs with similar appeal to programmers, allowing enforcement here would make of the Sun Java API's declaring code a lock limiting the future creativity of new programs".[19]

17 In *Authors Guild v Google*, the Second Circuit Court of Appeals found that Google's making of a digital copy to provide a search function was a transformative use, because such a use "augments public knowledge by making available information *about* Plaintiffs' books without providing the public with

17 *Authors Guild, Inc v Google, Inc,* 804 F 3d 202 at 216–217 (2nd Cir, 2015); *Authors Guild, Inc v HathiTrust*, 755 F 3d 87 at 97–105 (2nd Cir. 2014).
18 141 S Ct 1183 (2021)
19 *Google LLC v Oracle America, Inc* 141 S Ct 1183 at 1208 (2021).

a substantial substitute for matter protected by the Plaintiffs' copyright interests in the original works or derivatives of them".[20]

18 In these paradigmatic cases, there is unequivocal benefit to the public at large. The purpose of Google using parts of the Sun Java API was to create a new platform that could be readily used by programmers to expand the use and usefulness of Android-based smartphones, thus enabling innovation that creates new opportunities for the whole market to grow.[21] Google's making of a digital copy of books was for the purpose of enabling a search for identification of books containing a term of interest to the searcher which further advances the pursuit of knowledge.[22]

19 Generally, the fintech players tend to be profit-oriented, and it is more difficult to argue that the copying of works to generate reports that help them better service existing customers or acquire new customers is analogous to the kind of "public benefit" that a Google search engine or API brings to the wider community. In *Authors Guild v Google*, the Second Circuit found no reason why Google's overall profit motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose epitomised by public access and the concomitant significant public benefit, together with the absence of significant substitutive competition, as reasons for granting fair use. When fintech players engage in TDM for the purpose of analytics, the collected information is unlikely to be publicly searchable or freely accessible. Consequently, it is more challenging to argue that such TDM activities are clearly serving copyright's overall objective of contributing to public knowledge.

IV. A (fairly) rocky road ahead for text and data mining in fintech

20 At first blush, the computational data analysis exception seems to be a boon to fintech players, as it allows for commercial

20 *Authors Guild, Inc v Google, Inc*, 804 F 3d 202 at 207 (2nd Cir, 2015).
21 *Google LLC v Oracle America, Inc* 141 S Ct 1183 at 1206 (2021) ("we must take into account the public benefits the copying will likely produce").
22 *Authors Guild, Inc v Google, Inc*, 804 F 3d 202 at 216 (2nd Cir, 2015).

use, but the conditions to be satisfied under s 244(2) seem particularly onerous. It should be noted that He Tianxiang, who compared the approaches in East Asian jurisdictions to TDM exceptions, proposed a less onerous exception for China phrased as "reproduction… in the course of data analysis and mining, in order to uncover new knowledge or insights".[23] The two illustrations in s 244 also offer limited understanding of what "lawful access" entails; they classify acts of bypassing terms of service agreements and circumventing paywalls of a database as unlawful. However, these illustrations are inadequate and there still exists significant ambiguity as to what "lawful access" means in a number of situations. Illustrations, if any, should only reflect clear situations where the use of such agreements is easily accessible such that the mining of related works would most likely constitute unlawful access. In addition to the two illustrations in s 244, perhaps more illustrations could be included, *eg*, "X does not have lawful access to the original if access to the original has been restricted by Y through machine-readable means in the case of content made publicly available online, such as metadata".

21 In summary, it is doubtful that in Singapore, a hybrid legal defence strategy, even if theoretically plausible, would likely be employed in practice as it is time-consuming, expensive and may not reflect favourably on the defendant. A miner in the fintech business is more likely to plead fair use under the new ss 190 and 191 with a preponderance of US decisions in its favour, but a commercially motivated purpose and a weak showing of public benefit akin to the *Google* decisions could nonetheless render such unauthorised copying unfair. The technology for AI and data analytics may be developing in leaps and bounds, but the legal road ahead may yet be a rocky one.

23 Tianxiang He, "Copyright Exceptions Reform and AI Data Analysis in China: A Modest Proposal" in *Artificial Intelligence and Intellectual Property* (Jyh-An Lee, Reto M Hilty and Kung-Chung Liu eds) (Oxford University Press, 2021) at p 218.