ALPSP Annual Conference and Awards
14-16 September 2022

#alpsp2022
www.alpsp.org/Conference
Session 9
A New Copyright Exception for Text and Data Mining: What Next for Government and Publishers?
Will Crook, PLS (Chair)
Leslie Lansman, Springer Nature

#alpsp2022
www.alpsp.org/Conference
A New Copyright Exception for Text and Data Mining: What Next for Government and Publishers?

Will Crook
Leslie Lansman
Sept 15, 2022
Introductions

Will Crook
Policy and Communications Manager
Publishers’ Licensing Services

Leslie Lansman
Chair/Directo ALPSP copyright/policy committee
Global Permissions Manager Springer Nature
All round copyright enthusiast

[LinkedIn profile link] Leslie Lansman
Artificial Intelligence/Text and Data Mining

Image courtesy of PLS.org
What is TDM/AI all about?

DALL-E (an example) which may currently be used non-commercially.

- Trained on 650 million images and text extracts.
- Creates wide range of images: ‘Dystopian Great Wave off Kanagawa as Godzilla eating Tokyo’ or ‘Teddy bears working on new AI research on the moon in the 1980s.’
- Can create in the style of Dali, or based on a popular program such as Unreal Engine

BUT, huge problems with bias (amongst other issues)
Dystopian Great Wave off Kanagawa as Godzilla eating Tokyo: images from OpenAI's DALL·E 2 (dalle dalle dall-e 2 open ai)
Teddy bears working on new AI research on the moon in the 1980s images from OpenAI's DALL·E 2 (dalle dall-e dall-e 2 open ai)
So how does TDM work?

Articles/Images/Corpus

LDMA works retrieved

Content indexed/refined for scope

LDMA works stored/copied. Turned into machine readable format

Long term storage of TDM input and output

Extracted information/usable output generated

Semantic query/user query

Dataset in some sort of deliverable form created

‘TDM’ analysis occurs: lexicons; identification patterns; etc

LDMA: literary, dramatic, musical, artistic
There’s no such thing as a free lunch! Literary/artistic (or other) works are the ‘food’ that TDM needs to create AI. These works may be protected, under a CC/OA licence, or in the public domain.

What AI/TDM creates may be independently be protected by:
- Copyright law (software is protected as a literary work)
- Copyright protection for computer generated works
- Database law
- Other (ie confidentiality, patent, trade secret, contract etc)
Where is AI/TDM going?

Everywhere but as it relates to publishing...

- Books ‘in the style’ of an author dead or alive?
- Books ‘tailored’ to the reader? Ie a summary of all the current info on lithium batteries
- Artwork for covers, images, ads etc
- Social media posts
- Nefarious purposes: fake authored articles; fake take-down notices;
- Artificially created data/images?
- AI tailored webpages?
- AI replacing jobs? ‘Robot Lawyers’?

First ‘AI’ written book: Springer Nature
Current UK (c) exception

- Introduced in 2014 after the Hargreaves Review
- The 2011 Review looked at how IP could be reformed to help increase innovation.

Section 29A - Copyright, Designs and Patents Act 1988

1) The computational analysis must be for the purpose of non-commercial research

2) The copy is accompanied by sufficient acknowledgment (unless this is practically impossible)

Contractual terms which purport to restrict or prevent the doing of the acts permitted under the exception are unenforceable.

Although text and data analysis is mainly concerned with mining literary works, the exception covers all categories of copyright works, and a parallel exception applies to recordings of performances.
Global Approach

**European Union**

- DSM Directive
  - Article 3 (Research organisations and cultural heritage institutions / scientific research)
  - Article 4 (Broad group / opt-out provision for rightholders)

**US**

- Fair Use
- Case by case basis

**Singapore**

- Broad exception for TDM
- No requirement for ‘legal access’
Why make a change?

**UK government focus on improving innovation.**

- The [Taskforce on Innovation, Growth and Regulatory Reform (TIGRR) report](#) urged UK leadership on AI. Unleashing the power of AI is a top priority in the plan to be the most pro-tech Government ever.
- The [UK Innovation Strategy](#) - long-term plan for delivering innovation-led growth.
- The [National AI Strategy](#) - How the UK will remain global AI superpowers over the next decade.

   >>> ‘Text and data mining will become easier, quicker and cheaper....’

**UK government searches for ‘Brexit opportunities’**

- Former minister now Secretary of State for BEIS
UK consultation on AI/TDM 2021

**Text and Data Mining (TDM)**

Option 0: Make no legal change

Option 1: Improve licensing environment for the purposes of TDM

Option 2: Extend the existing TDM exception to cover commercial research and databases

Option 3: Adopt a TDM exception for any use, with a rights holder opt-out

Option 4: Adopt a TDM exception for any use, which does not allow rights holders to opt out
Proposed New Copyright Exception for TDM

- A copyright and database exception that allows TDM for any purpose.
- The new exception would retain a requirement for lawful access.
- Rights holders will not be able to contract or opt-out of the exception.
Responses

Industry

- Publishers Content Forum
- Trade association/PLS data collection
- Wider creative industry opposition
- All Party Parliamentary Group for IP

Government

- Promise to ‘look again’
- IPO engagement
- New administration
What has ALPSP done?

ALPSP response to initial consultation:

- Highlighted the host of successful licensing strategies and policies publishers already have adopted
- Emphasised commercial TDM does not live in a bubble, but is part of oftentimes complex partnerships (worth estimated £335M to publishers)
- Stressed publisher importance as ‘gatekeepers’ of sensitive information/information not fit for purpose
- Underlined the complexity of this landscape: commercial versus non-commercial AI development; variations in AI output etc

Reiterated our concerns in follow up letter after government proposal announced (Aug 2022)
What’s next?

Legal considerations:
- Does the proposed exception satisfy international treaty obligations (the three step test)?
- Liability determinations

Implementation problems;
- Will there be any kind of ‘opt out’?
- Different effects on different markets? STM versus humanities?
- What is ‘lawful access’ and how can anyone enforce it?

Social/Cultural issues:
- If publishers are not part of the AI/TDM industry, who will be the counterweight to the tech industry/safeguarding author’s/publisher’s concerns against misuse?
- Concerns over misinformation?
Thank you! Questions?