AI, machine learning and the role of copyright in protecting non personal "data"

AI and Audiovisual Archives
Learning on Screen

Dr. Thomas Margoni
Senior Lecturer in Intellectual Property and Internet Law
Director of the IP LLM Programme
Co-director, CREATe
School of Law – CREATe
University of Glasgow
www.create.ac.uk

thomas.margoni@glasgow.ac.uk

CREATe

OpenAIRE

AI and IP law

3 main levels of interplay

1) AI helps IP law: e.g. network analysis of patent databases to find new areas of exploitation; legal research automation and prediction, smart contracts, etc.

2) IP law helps (?) AI: machines improve, enhance and assist human creativity and inventorship; Can/should AI produced creations and inventions be protected?

3) Could AI be "better" and what role IP should/could play?: AI, especially data driven approaches such as machine learning, neural networks and the like need to be trained on "data" in order to learn certain tasks. It follows that the availability and quality of "data" plays a central role in how good (precise but also fair, transparent, accountable, ethical) AI will be.

CREATe

thomas.margoni@glasgow.ac.uk

3) Is data owned?

Copyright theory (and sometimes copyright law) says: no
E.g.: principles, facts, data as such, etc are not protected by copyright law (Arts. 2 WCT, 9(2) TRIPs, but also generally in Berne and most legal traditions).

Other areas of law may say yes, but usually in specifically identified situations, or with limited remedies e.g.:
1) Trade secrets, confidentiality (only if secret and limited remedies)
2) Contracts (enforceability and remedies)
3) Data protection (only qualifying information and scope is protection of private life not investment)

CREATe

thomas.margoni@glasgow.ac.uk

3) Is data owned?

Some jurisdictions however offer some sort of protection to data: e.g. EU offers protection against substantial extractions of certain non original databases therefore effectively protecting data (but not created data). This is called sui generis database right.

Other jurisdictions (actually the same jurisdiction) have considered creating a new right protecting **data producers** (not just databases)

CREATe

thomas.margoni@glasgow.ac.uk

3) Is data owned and why does it matter for AI?

But if we take a closer look, EU Copyright law (doesn't say anything explicit but) often means: YES.

Interplay between broad rights (e.g. reproduction) and limited fair dealing/exceptions requires to obtain an authorisation for data capturing/extraction. Otherwise we would not need a TDM exception!

Modern data analysitics (e.g. TDM, machine learning etc. normally extracts principles, facts, data, correlations, etc,) which copyright theory says are not protected, thus the extraction of those unprotected elements from protected works should not need an exception if copyright framework was properly designed.

Main problem with EU/UK copyright design is that it is not properly designed: it harmonises rights broadly (reproduction, redistribution, communication to the public, etc), but does not do the same with exceptions (exhaustive but not mandatory list, narrow interpretation, etc). Main problem with this design is that you can only allow what you are aware of. This is an implicit (or perhaps quite explicit) presumption in favour of protection of investment vs. open, transparent and distributed innovation.

CREATe

thomas.margoni@glasgow.ac.uk

Example: TDM exceptions

1) Text and Data mining: computational analysis of anything recorded in the work (sec. 29A CDPA) or any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations (Art. 3 CDSM);

2) **Scope**: exception to the right of <u>reproduction</u> (both);

3) **Beneficiaries**: <u>Non commercial research</u> (29A CDPA), <u>research organisations</u> with lawful access <u>for research purposes (Art. 3 CDSM), anyone for any purpose but can be opted-out (Art. 4 CDSM).</u>

4) **Relationship to contracts**: <u>Cannot</u> be limited by contract (except for Art. 4 CDSM).

5) **Relationship to technology**: <u>Can</u> be limited by technological measures (integrity measures and TPM)

CREATe

thomas.margoni@glasgow.ac.uk

TDM exceptions

1) Text and Data mining: computational analysis of anything recorded in the work (sec. 29A CDPA) or any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations (Art. 3 CDSM);

Comment: definition is broad enough to cover current TDM practices.

CREATe

TDM exceptions

2) **Scope**: exception to the right of <u>reproduction</u>;

Comment: Problematic. It does not cover rights of redistribution/communication to the public and adaptation (derivative works). It means that all the times that the results of TDM are a copy in part of a protected work (CJEU in Infopaq says that even 11 consecutive words can infringe) or when the results can be an adaption (derivative) of the original (thumbnails?) the exception is not available.

thomas.margoni@glasgow.ac.uk

TDM exceptions

3) **Beneficiaries**: <u>Non commercial research</u> (29A CDPA), <u>research organisations</u> with lawful access <u>for research purposes (Art. 3 CDSM), anyone for any purpose but can be opted-out (Art. 4 CDSM).</u>

Comment: Problematic. Individuals, micro and SMEs, industry, etc cannot benefit even if acting non commercially. Purposes other than research (e.g. journalism, criticisms, review, etc) are not covered. Why? Potential contrast with fundamental rights?

CREATe

TDM exceptions

4) **Relationship to contracts**: <u>Cannot</u> be limited by contract (except for Art. 4 CDSM).

5) Relationship to technology: <u>Can</u> be limited by technological measures (integrity measures and TPM).

Comment: 4) is good. But 5) is contradictory. It creates imbalance and uncertainty with regards to the medium through which a prohibition is expressed. If "exception not available" is expressed in human/legal language (contract) this is not enforceable, but if the same condition is expressed in computer language (DRM or TPM) then it is allowed. 5) basically circumvents 4) in a way that is unreasonable, not proportionate and harmful for researchers and other users.

CREATe

thomas.margoni@glasgow.ac.uk

Extra EU/UK examples

Fair use approaches (e.g. US) do not suffer of all these limitations and TDM can be performed as it is usually a transformative use.

Fair use (US) or **flexible fair dealing** (Canada) models have been implemented in other similar jurisdictions (Singapore)

Other innovation-friendly countries (e.g. Japan) have implemented **specific but broad TDM exceptions.**

CREATe

thomas.margoni@glasgow.ac.uk

Useful sources:

- UK IPO Guides on copyright exceptions
  https://www.gov.uk/government/publications/changes-to-copyright-law

- Copyrightuser.org – TDM exception
  https://www.copyrightuser.org/understand/exceptions/text-data-mining/

- CREATe website – Open Science
  https://www.create.ac.uk/open-science/

- OpenAIRE guides on data reuse
  https://www.openaire.eu/can-i-reuse-someone-else-research-data

- A Legal Perspective on Training Models for Natural Language Processing
  http://eprints.gla.ac.uk/159231/

- Artificial Intelligence, Machine learning and EU copyright law: Who owns AI?
  https://zenodo.org/record/2001763