

AI and the Copyright Amendment Bill

Dr. Andrew Rens and Hanani Hlomani

AI: the next general purpose technology

- Protein structure prediction for drug discovery
- Medical image analysis
- Automated translation
- Fraud detection
- Automated image recognition of defects in manufacturing
- Predictive maintenance
- Supply chain demand forecasting

Copyright and Outputs: Machine Learning

- INPUT DATA



- MODEL ← PROMPT



- OUTPUT

DALL-E2 An expert explains AI copyright to parliament



Large language models

One kind of machine learning is large language models such as ChatGPT that trained using millions of inputs e.g. most of the Internet in English - when the resulting algorithm is fed new data - a text prompt that might be only a few words - it predicts the next word, and the next and the next....

Producing a static ML Model

Corpus Compilation	Reproduction or temporary copy?
Corpus Technical Pre-Processing	Reproduction
Corpus Annotation	Adaptation or Derivative Work?
Corpus Redaction	Delete reproduction
Train Model	New software work?
Static Model	New software work?

3 copyright questions for AI

- 1 Does use in training an algorithm infringe copyright in the individual items in a training corpus?
- 2 Do the outputs infringe copyright in individual items in the training corpus?
- 3 Does the output have copyright itself?

3 copyright questions for AI

1 Does use in training an algorithm infringe copyright in the individual items in a training corpus?

Maybe. But (most) use for computational analysis should be clearly permitted by an appropriate provision.

2 Do the outputs infringe copyright in individual items in the training corpus?

Very very rarely, and when it does existing law deals with it.

3 Does the output have copyright itself?

Probably not but its important to make it clear that it will not.

No copyright for AI outputs

- Copyright is intended as an incentive to humans to create, to produce text, images, video and software.
- AI does not need copyright as an incentive to produce. In July 2019 Microsoft invested 1 billion USD in Open AI (producer of ChatGPT) in January 2023 Microsoft invested an additional 10 billion USD.
- AI processes do not resemble human creative processes - they are probabilistic.
- Most AI systems which are used to generate creative output were trained on very large numbers of creative inputs.
- AI systems that produce outputs that mimic their creative inputs risk undercutting the ability of artists to generate earnings.

No Copyright in AI outputs

- Amend the 1978 Act to clarify that there is no copyright in AI outputs.
- But this technology is developing very rapidly.
- Specifying which processes don't result in copyright creates a trap: legislation that will become rapidly outdated because its technology specific.
- Instead make it clear that copyright only attaches to human creativity - the results of a human authors' skill, effort and creativity.
- And for text, images, music, software and audiovisual works make it clear that authors must be human authors.

Enable AI Research

- AI is a strategic technology. If South Africa does not develop our own AI we will be reliant on countries that do.
- AI is the next big computing technology - if it is not clearly enabled in the Cape there will be no Silicon Cape.
- Big tech companies don't need a copyright exception – they have millions of our images and texts they can use under their terms and conditions.
- There is a right to research under Section 16(1)(d) of the Bill of Rights.
- A flexible provision can balance competing interests.

Enable AI Research

- Insert a clear exception for use in computational analysis - text and data mining and training AI.
- But how?
- The European Union and the UK have narrow, detailed exceptions but
 - they don't work, and
 - the UK is already debating how to change their exception.
- Meanwhile the US has fair use that permits most computational analysis and leads the world in AI.
- Keep s12A in the CAB but include “computational analysis” as part of research.

Thank you

Research made possible by

