

A glowing lightbulb is positioned in the center-left of the image, set against a dense background of vibrant green leaves. The lightbulb is illuminated from within, casting a soft glow. The background consists of many small, rounded green leaves, some with tiny yellow flowers.

# AI & Copyright Law – What is it and what are the issues?

**Benjamin White**

**KR21 / Centre for  
Intellectual Property Policy  
& Management**

**Bournemouth University**



# KR21

Arcadia Foundation funded IFLA programme working in Europe to create a more education and research friendly legal regime.

We work mainly on copyright, open access and competition law.

Management Committee = LIBER / IFLA / SPARC Europe

# KR21

Policy lead is in Brussels

National coordinators in Ireland, Sweden, Finland, UK, Germany, Netherlands, France, Spain, Italy, Portugal, Poland, Slovenia, Bulgaria.

# What is AI?

- Software that makes predictions and classifications.
- It produces interesting result but is not “intelligent” as there is no autonomous or independent activity.
- Combination of data scientists, hardware, software and very often large datasets. e.g. The internet.
- Many different types of AI software – it’s not all neural networks by any stretches of the imagination. (NN not even top three most used algorithm)
- Not all models train on data.

# What is AI? Jargontastic!

- Algorithms
- Artificial intelligence, incl. generative AI
- Machine learning
- Text and Data mining
- Content mining
- Data analytics

# Applications

- Robotics
- Spam filters
- Routing on your phone
- Internet search
- Chat GPT
- Weather predictions
- New antibiotic discoveries
- Profiling

# Inputs

- **Data.** Whether as i) the object of the analysis (all models) or ii) for some models (“eager learners”) training the models.
- Creation of the **software** which is usually derived from open source AI toolkits and may involve downloading a whole model, or combining different software programs.
- If requires training, the data is analysed trying different software settings (“hyperparameters”) to create the best result.
- A model shouldn’t “overfit” the data as it will reduce its predictive capabilities and from a copyright perspective could lead to the output being substantially similar to the input.

# Outputs

- Movement of a robot.
- A percentage chance.
- A route.
- A picture or a sentence.
- A weather prediction.
- A new antibiotic.



# Copyright and AI

Professor Joseph Stiglitz:

Information technologies have a tendency to create natural monopolies / oligopolies creating markets with a few very large players and which are marked by a lack of competition.

No reason to think that Artificial Intelligence markets will be any different – in fact there are reasons to think that they will be exactly, not least data and reliance on intangible assets like IPR.

# Copyright and AI

Professor Pamela Samuelson:

*“Copyright has the potential to bring AI to its knees”*

- Training data frequently includes third party copyright and database rights. e.g. Internet
- Subject of analysis frequently involves third party copyright.

# Copyright Over Extension

Copyright used to regulate the printing of books - now **it regulates everything we do in a digital world and therefore entire market places** because of the simple fact that information technologies copy.

How governments conceive and craft copyright matters because it dictates the speed of innovation and scientific advancement.

US, Japan, Singapore, Taiwan etc. have more flexible copyright regimes and actively recognise that copyright should not be used to stifle research and digital innovation.

European policy makers still act as if copyright is a cultural policy and that it only really affects the entertainment industries.

# Different © Approaches

## UK & EUROPE (RIGID & PRESCRIPTIVE)

- No commercial use (UK)
- EU allows commercial AI but allows opt outs
- Don't allow contract override of TDM exceptions
- EU doesn't allow technical protection measures to prevent TDM (72 hours in Bulgaria / Slovenia)
- No contract override of copyright law

## JAPAN & FAIR USE COUNTRIES

- Japan – introduced an exception for non-enjoyment purposes
- US, Israel, Singapore, Nigeria, Taiwan, South Korea etc allow data analytics / AI under fair use (flexible, principles based approach rather than prescriptive and rigid)
- Singapore has a prescriptive TDM exception and fair use.

# Entertainment Industry





# Entertainment Industry – Misinformation!

“would allow any ... businesses of any size, located anywhere in the world, **to access all my members’ data for free** for the purposes of text and data mining. There is no differentiation between a large US tech firm in the US and an AI micro start-up in the north of England.”

UK Publishers Association. House of Lords.

**Fact Check:** Libraries spend circa \$10 billion a year on STM databases alone.

# Regulating Copying

How is regulating certain forms of AI different to the way copyright regulates internet search?

Both involve processing data to update software resulting in an output.

Copyright has developed and search engines are not required to pay each website owner some money.

Certain protocols, like robot.txt, have developed and the newspaper industry in Europe has been given some special treatment in Europe.

# Policy Issues for Research and Innovation

## Example Europe:

- To what extent can you share the training data or the outputs with others?
- In the EU, no cross-border effect created so working in the European Research Area under the exception difficult.
- Push back from the entertainment industry saying the UK / EU copyright framework doesn't allow AI, just TDM.
- European research exceptions only allow non-commercial research.

# Killer Robot Syndrome



# Rights Reservations Cause Bias – STM Publishers and Authors

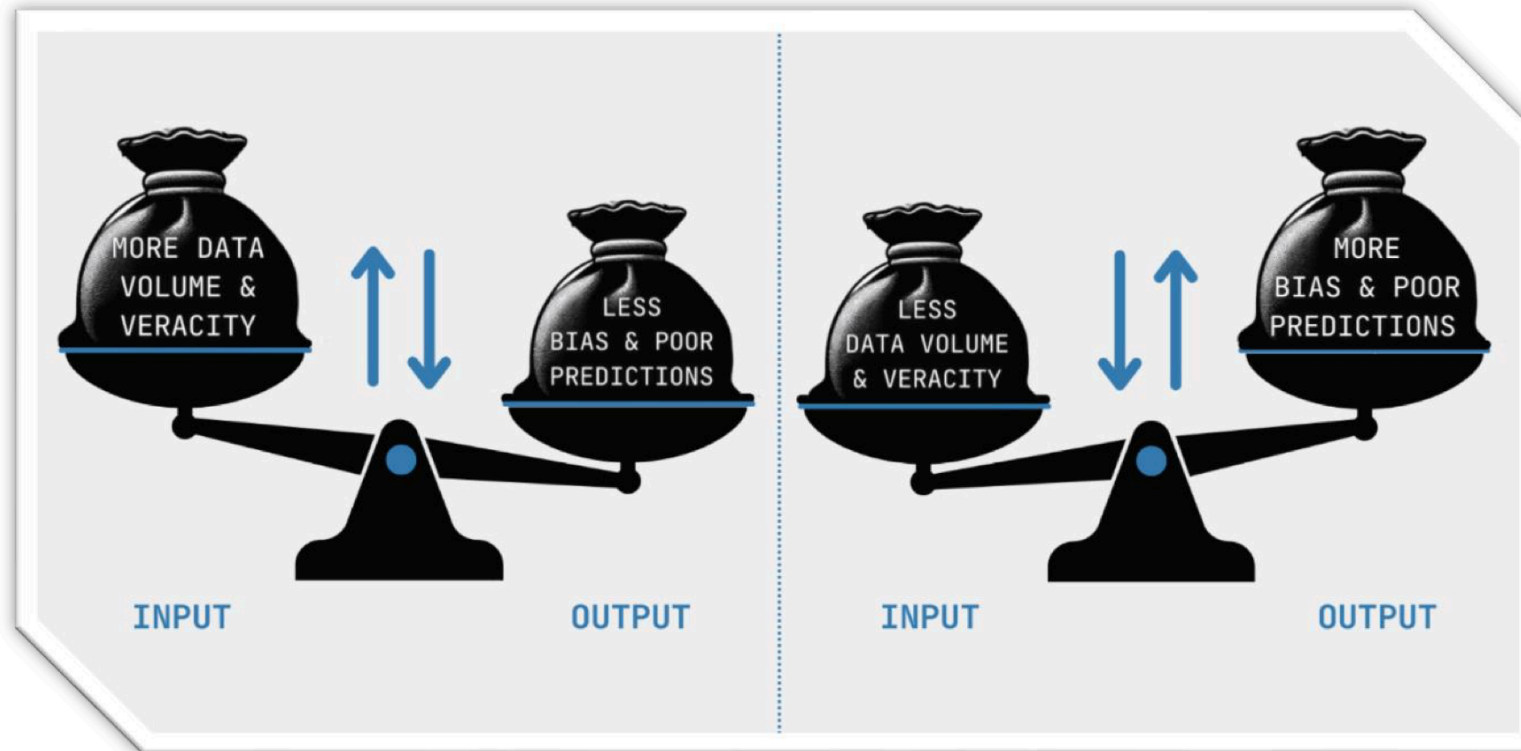


Figure 1: Illustration of the impact data volume and veracity on the input side have on the accuracy of AI's output



# From a © Perspective is AI any different from Internet Search?

- Based on access to masses of data.
- Involves training software / algorithms
- Produces a result

# Libraries Giving Access to Collections

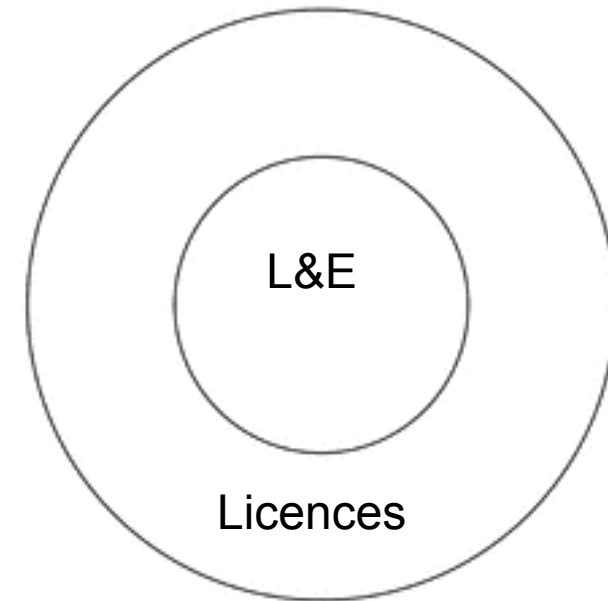
- Nothing new here in many senses – Public-Private Partnerships are nothing new.
- Libraries need to think what they need from such relationships.

# STM Publishers

- Placing contractual restrictions around AI – making a distinction between TDM and AI
- 38% of articles were Gold OA in 2023
- Licensed databases are a mishmash of proprietary and non-proprietary, so asserting control over open materials by licence?
- Underlines importance of public infrastructures

# Library Senior Leadership – Libraries are about access = © & L

- Library Directors need to invest in copyright and licensing as inter alia publishers licensing *around AI* exceptions.
- Government affairs engagement to shape exceptions.
- Risk management profile around © exceptions



# Librarianship and Knowledge Equity

- Freedom of Information and Expression
- Knowledge Equity
- Bias?
- How does that sit with choosing “good actors” and “bad actors”...
- Transparency v Monitoring



# KR21 Principles on Artificial Intelligence, Science and Research

- The needs of science and research should underpin AI policy formulation
- Volume and veracity of data matter
- Opt-outs increase bias and poor AI in certain sectors
- Opt-outs reinforce market concentration and a lack of competition
- We need transparency, not monitoring
- Uphold academic freedoms and wider human rights
- Promote a pro-research and modern information access regime
- Take an evidence-based and domain-specific approach to regulation

# KR21 Further Reading

- Why are legacy publishers trying to prevent universities from undertaking AI? KR21 blog.
- KR21 Principles on AI science and research
- KR21 AI Primer
- Plus further Recommended Reading
- Available at [www.knowledgerights21.org/news-story/undermining-scientific-research/](http://www.knowledgerights21.org/news-story/undermining-scientific-research/)



## Undermining Scientific Research

? Why are legacy publishers trying to prevent universities from undertaking AI ?

"...licences, the European AI Act and copyright framework discourage safe and unbiased AI by overly regulating and restricting access ("inputs") to training materials."  
"In the interests of science and society more broadly

**Europe should avoid a "garbage in garbage out" scenario,**  
but on the contrary appears to be facilitating it."



**"KR21 ARE VERY CONCERNED ABOUT RESTRICTIONS BEING PLACED ON UNIVERSITIES BY PUBLISHERS."**

[www.knowledgerights21.org](http://www.knowledgerights21.org)

[ben.white@ifla.org](mailto:ben.white@ifla.org)

