Copyright & AI Training Data

Roundtable meeting of rights holders and AI developers

9th July 2024, 56 Greek Street, London, W1D 3DU

The Generative AI (GenAI) business sector depends on access to human-originating text, image, music, video and voice data at scale. Such data is routinely scraped from the internet by many AI developers for use in training foundation and large language models, without consent or remuneration for the creators of this data.

Objective: To establish agreement on a pathway towards the licensing of writers' and creative artists' copyrighted datasets to AI developers for use in training AI models.

This meeting responds to two aligned interests:

– The wish of an important section of the AI developer community to build an alternative, ethical business model based on the licensing of creative artists' datasets;

– The wish of copyright holders for a system of consent and compensation for the use of their work by the GenAI industry.

Hosts

The Synthetic Media Research Network is an independent group of researchers, stakeholders, experts and businesses working in the field of GenAI. Members of this group have provided consultancy for policy organisations, industry, trades unions, the UK Parliament and UNESCO. The roundtable meeting will be chaired by Dr Dominic Lees and Dr Mathilde Pavis.

Dr Dominic Lees

Dr Mathilde Pavis

Objectives

To understand the appetite of a sector of the AI industry to operate an ethical business model when accessing training data.

To hear the specific business requirements of AI developers when licensing datasets.

To express the specific concerns of copyright holders in relation to the use of their work in training AI models.

For the copyright holder representatives:

1. To share internal discussions around AI and copyright, establishing mutual agreement on fundamental principles of consent and remuneration.

2. To outline the technical and legal challenges involved in establishing a central resource of writers' and creative artists' datasets.

3. To establish a consensus on the tasks needed to establish a system for licensing creative datasets to AI developers.

Agenda

10.00	**Arrivals, Coffee – The Miquita Room**
10.30	**Move to the Elton Room** Introduction to the Roundtable meeting; self-introduction by attendees; Chair's overview of objectives.
10.45	**Copyright Holders' Representatives** Each organisation to share their current discussions on AI and the challenges related to the use of members' work by developers.
12.00	**AI developers** will outline their business needs related to licensing training data (includes remote contributions by: Vered Horesh of BRIA; Alex Serdiuk, CEO of Respeecher; Jillian Bommartino of KL3M).
13.00	**Lunch and networking**
13.30	**Discussion:** common themes, concerns and ambitions.
14.00	**Technical and legal challenges** in establishing a UK centre for licensing creative datasets.
14.30	**The pathway forward** Commitment to consensus points of agreement.
15.00	**Close**