March 15, 2025

Faisal D'Souza
Networking and Information Technology Research and Development (NITRD), National
Coordination Office (NCO), National Science Foundation.
Via Email: ostp-ai-rfi@nitrd.gov

Re: Request for Information on the Development of an Artificial Intelligence (AI) Action
Plan

Introduction

Knowledge Ecology International (KEI) asks that two recent KEI documents be include
by reference.

- KEI's February 2025 submission to the UK IP office consultation on AI and
  Copyright (available here: https://www.keionline.org/40558),
- KEI's February 3, 2025 comments to USTR regarding its Special 301 list
  (Available here: https://www.regulations.gov/comment/USTR-2024-0023-0051)

KEI's comments today focus on text and data mining exceptions, metadata and
formalities, competition, and data spaces.

Text and data mining exceptions

When considering policies regarding exceptions for text and data mining (TDM), for training AI services, pay attention to global norms, and make distinctions between different types of data. From KEI's USTR Comments:

> Any country that provides robust exceptions for using copyrighted material for AI will have a significant advantage in terms of training such services. But a lack of harmonization may create a situation where services developed in one country, such as in the United States, will not be legal in another, because of non-authorized use of copyrighted works to train the service. This makes it a significant trade issue.
>
> The most emotive and politically important voices to make it an infringement of copyright to train AI services are cultural industries and journalism. KEI has been concerned that restrictive policies on the uses of copyrighted works to train AI will be extended to much broader classes of works. We have singled out science, drug development and legal issues, to illustrate cases where society is best off if AI services have access to everything possible in terms of data, and where omissions may have considerable downsides. (2023. James Love. We Need Smart Intellectual Property Laws for Artificial Intelligence: "One-size-fits-all" regulation will sideline medical and research benefits promised by the advent of artificial intelligence, Scientific American, August 7, 2023.)
>
> In regard to scientific and medical information, note that authors are rarely paid for their works, and a handful of companies control a large number of journals, many of them foreign-owned. It would be a very bad outcome if the Holtzbrinck Publishing Group, the private equity firm BC Partners, Wiley and the Relx Group, are able to significantly limit which companies can use the leading medical journals to train AI services, or for that matter, if any publisher can opt-out of the science being used to train programs that are used for drug discovery or to treat and protect patients.

Knowledge Ecology International (KEI) encourages governments to extend TDM exceptions to both commercial and noncommercial works.

KEI notes that there may be cases where opt-in or opt-out options are appropriate, particularly for some uses involving cultural industries. However, more robust TDM exceptions should be the general rule. Opt-in or opt-out options should be only considered for a narrow set of cases where they are appropriate.

KEI's submission to the UK IP office included comments on the importance of being inclusive when it comes to the text and data used to train AI services dealing with the following topics:

- Medical Products and Services
- Legal Services
- Weather Forecasting
- Engineering and Manufacturing
- Financial Services
- Education and Academic Research
- Patent Examinations

Metadata and Formalities

For a variety of reasons, it will be useful to have accurate metadata attached to published articles, photographics, designs and more generally, for data of all sorts.

Some WIPO administered treaties on copyright and related rights place practical restrictions on requirements for formalities, which make it more difficult to identify works used to train AI, and to provide remuneration, if that is a policy objective.

Metadata adds crucial context, explaining the meaning, origin, and characteristics of the data, and allows machine learning models to interpret the information more accurately.

Standards for metadata are important, challenging to develop and typically are a cross border concern. KEI has been asking the WIPO Standing Committee on Copyright and Related Rights to have as an agenda item discussions on metadata, beginning with recorded music and photographs.

Competition Concerns

Economies of scale are important and present risks that there will be excessive concentration of control of training data.

KEI is concerned that opt-in or opt-out rules for text and data mining will lead to high entry barriers for some types of services. Exclusionary or discriminatory licensing also presents risks.

It may be important to treat some types of data or databases as essential facilities, and mandate some types of sharing of access.

Data Spaces

For a variety of reasons, governments should consider the benefits of data spaces when it comes to the management of data to train AI services.

Data spaces is a term used to describe a framework where organizations exchange or provide controlled access to data, under agreed-upon rules, standards and infrastructure.

In decentralized systems, data remains held by various entities, but can be accessed by authorized parties for specific uses. For example, medical records data around the world is held by different government and private entities, and is subject to rules to protect the privacy of patients. Queries can be run on decentralized datasets, with the results benefiting researchers, while avoiding excessive concentration of the control over records, and allowing each organization holding the records to protect patient privacy according to relevant ethical and national legal obligations.

Sincerely,

James Love
Executive Director
Knowledge Ecology International (KEI)
https://keionline.org