



AI AND THE LAW: SAFEGUARDING AND PROMOTING INNOVATION, CREATIVITY AND COMPETITION

Juha Vesala, Postdoctoral Researcher
Faculty of Law, University of Helsinki

AI Day 2019, Finnish Center for Artificial Intelligence
26 November 2019



DATA ISSUES IN AI DEVELOPMENT

- While AI contributes to productivity and enables improved and new products, some factors associated with AI may threaten competition, innovation and creativity in markets where AI is used
- Issues relating to training data used in AI development are one potential source of concern
- I'll examine concerns related to access and use of data and approaches that are emerging in European intellectual property and competition policies

Disclaimer: I'm a member of a research project on AI generated content (prof. Taina Pihlajarinne) that has received funding from the Helsingin Sanomat Foundation



TRAINING DATA: POTENTIAL OBSTACLES

- While many types of data are available and others can be independently produced, some may only be available to a few organizations unwilling to grant access to others
- Even when data is available, it may be unlawful to use it in AI development
 - Technical restrictions: unlawful to circumvent
 - Agreements: restrictions on use
 - IP rights: infringement



INNOVATION AND COMPETITION CONCERNS

- Difficulties in gaining access and permission to use data may hinder the development of AI technologies and applications
- Moreover, access to data limited only to some firms can threaten competition in AI development and AI applications as well as facilitate acquisition of market power
- Data access issues can be detrimental to innovation and competition as well as EU competitiveness



COPYRIGHT OBSTACLES AS AN EXAMPLE

- Using materials protected by copyright and related rights (e.g. songs, photos, databases) to train neural networks potentially involves infringing reproduction of the materials
 - Collection and preparation
 - Processing in networks
 - Outputs
- Exceptions in EU copyright law apply to commercial AI development only narrowly and subject to considerable uncertainty – in comparison, for instance, U.S. fair use may permit activities that do not harm rightsholders
- Obtaining authorization from possibly thousands of rightsholders is often unfeasible



“We created a new dataset which emphasizes diversity of content, by scraping content from the Internet. In order to preserve document quality, we used only pages which have been curated/filtered by humans — specifically, we used outbound links from Reddit which received at least 3 karma.”

<https://openai.com/blog/better-language-models/>

“This site may contain copyrighted material the use of which has not always been specifically authorized by the copyright owner. We are making such material available non-commercially in an effort to educate and advance research in machine learning, generative music, music information retrieval, computational creativity, etc. *We believe this constitutes a ‘fair use’ of any such copyrighted material as provided for in section 107 of the US Copyright Law.*”

<https://dadabots.bandcamp.com/album/bot-prownies>

“One of the hard limitations faced by MIR researchers is the lack of publicly available datasets due to copyright issues. Here at Deezer, we have access to a fairly large catalog that we’ve been leveraging to build Spleeter.”

<https://deezer.io/releasing-spleeter-deezer-r-d-source-separation-engine>



NEW EXCEPTIONS ON TEXT AND DATA MINING (TDM) IN THE EU

- Directive on copyright in a Digital Single Market requires EU Member States to adopt exceptions for TDM by 7 June 2021
 - Copying of lawfully available materials which is necessary for TDM does not infringe copyright or certain other rights, unless rightsholder has expressly reserved that right
 - For solely non-commercial research purposes, rightsholders cannot even through express reservations prevent the exception from applying or enforce agreements that contravene the exception
- To the extent that AI development constitutes TDM, it may become possible to use lawfully available materials without authorization from copyright holders (in commercial context provided no rightsholder reservation)

TDM=“any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations”



COMPETITION POLICY: TOWARDS EXPANDED DUTIES TO PROVIDE DATA?

- Current EU case law requires dominant firms to grant access to an input (e.g. data) where the firm reserves a downstream market to itself by refusing to provide that indispensable input
- European Commission and Member States are considering whether dominant or particularly “strategic” firms should be required to provide data to others also in other circumstances
- So far, there has not been any major antitrust decision relating to data access but Germany is considering a proposal for stricter duties related to data access and is planning to pursue similar changes to competition rules also at the EU level



MISAPPROPRIATION OF DATA AS A CONCERN

- AI could be used to freeride on investments made by others in data (e.g. production, collection) in ways that undermine incentives to invest and compete
- While AI applications are not exempt from copyright or other laws, there could be gaps in how the law protects valuable data
- For example, imagine a service which scrapes content from the web and generates short summaries using AI and thereby diverts consumers away from the original sources (e.g. nonfiction text)
 - Some types of data and manners of use escape IP infringement: e.g. summaries not containing original expression
 - Firms may be able to force others to tolerate or allow use of data: e.g. condition for operating on a platform



IP AND COMPETITION LAW REACTIONS

- A new press publishers' exclusive right in the DSM Directive targets use in online services beyond conventional copyright law
- Competition law may condemn the use of data by dominant firms when it harms competition on the market
 - Google News investigations: use of materials from news websites without consent or under unfair conditions?
 - Amazon investigations: requiring access to and using merchant data to compete unfairly (e.g. private label imitations)?
- A platform favoring its own services may engage in abuse (e.g. Google shopping comparison)



CONCLUSIONS AND OUTLOOK

- Data issues relating to AI are examined by authorities and policymakers in competition, IP and other fields of law
- Legislative action is expected as AI and data are top priorities for the new European Commission
- Complex tradeoffs between protecting data and allowing its use



THANK YOU!

juha.vesala@helsinki.fi