Expanding Business Thanks to Data Economy: IP as a Tool

Encounter 14

FINAL REPORT

Date: July 21st, 2021

Dominique Guellec, Scientific Advisor at The Observatory of Science and Technology, Paris.

Prof. Guido Noto La Diega, Associate Professor of Intellectual Property and Privacy Law at the University of Stirling.

Prof. Aurelio López-Tarruella, Associate Professor of Private International Law at the University of Alicante.

Introduction

Objective of the session: In a post-crisis environment, re-imagining IP may mean better understanding how IP can and shall be a tool in the new data economy, in particular in science, business and education connected to data economy: how can patents (and scientific publications) be used as a source of data for monitoring developments in technology, using AI-based methods (e.g. natural language processing), how decisions can be at the crossroads of intellectual property, data protection, and freedom of information, how the text and data mining exception might or might not be a lost opportunity to boost European companies' competitiveness in AI?

Moderator: Prof. Aurelio López-Tarruella laid the foundation by speaking about the need for discussion and conversation on **data and their access for innovation** in EU and not just be confined or limited to personal data rights of subjects. The session took the form of Q&A (question and answer) session with Prof. López-Tarruella raising the below questions:

1. **Do public data such as patent data provide new opportunities for availability of data for patent analytics?**

Dr. Dominique Guellec responded stating: Searching for relevant information in over 3 million patent applications filed every year is akin to searching for a needle in a haystack. Patent analytics or exploitation of patent data in the past was primarily restricted patent metadata, that is, information related to inventors, geography, technological category etc. However, coming of AI has made possible the analysis of text, giving access to the actual content of the patent, that the machine transforms into a vector using methods called "Natural language processing". Vectors have enabled contextualizing text, which was not the case in the earlier key-word search process. Word associations, meanings and significance can be extracted due to vectorization, and this will enable more precise categorization of patent data, searching for closest technical neighbour and clusters within patent data. This is going to ease freedom to operate assessments, state of art assessments etc, thereby easing landscaping of technological opportunities and would enable enterprises to be for more strategic. Trends of technologies, forecasting and strategic landscaping of future technologies can be done using AI and patent data. Millions of patents have been vectorized. Natural language processing, though a powerful tool is still limited in deciphering correctly the meaning of the text, the drawings and figures. There is rapid progress being made in these fields, though limitations do exist presently.

2. **Prof. Tarruella opined that the data mining exception introduced in European Union (articles 3 and 4 in Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market) envisaged to facilitate research, however it has turned out to be restrictive as it permits data mining solely in research, cultural and heritage institutions thereby not furthering innovation cause in EU. Prof López-Tarruella requested Prof. Guido Noto La Diega to share his views on the issue and on UK's legislation in this regard.**

Prof. Guido Noto La Diega responded stating: Text and Data mining exception was introduced in UK in 2014. There is data that shows libraries have been using this exception to analyze large amount of data. The UK approach has been far more restrictive than in European Union (EU) approach as it is restricted to non-commercial research only. In EU they do have text and data bifurcated thus providing some maneuverability. Prof. Noto La Diega opined that we have been moving on a regressive path since the time of InfoSec directive where non-commercial activity was defined based on the activity, the present exception defines it based on organizational structure.

Further, transferring a copy of data and sharing it with a team member might be deemed as a copyright violation. Seconding with Thomas Margoni & Martin

Kretschmer's thesis that activity of text and data mining should not even be viewed as an exception to copyright protection as it should not be regarded as an infringement to begin with.

Dr. Guellec agreeing with the need for broader exception and further opined that though patent data is public, there are technical barriers to access databases, as well as many scientific publications and other research materials cannot be accessed due to technical or contractual restrictions on use.

3. **In comparison with US where data mining is covered by fair use doctrine and there is no discussion on data mining exception. Is Europe lagging behind due to higher legal standard or barriers in comparison to other countries?**

Prof. Noto La Diega: Opined that the three-step test in Europe (Article 13 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) 1. in special cases, 2.which do not result in a conflict with the normal exploitation of a work and 3. which do not unreasonably prejudice the legitimate interests of the author (or another right-holder) has been narrowly interpreted thus limiting its scope and Europe should have a fair use doctrine similar to US.

4. **Personal data, Interaction between IP and GDPR. Many times, we can find obstacles to comply with transparency obligation in the GDPR without infringing on IP.**

Prof. Noto La Diega: Opined that this is an unresolved issue. We have companies basing their business models on data (personal data included). Does control of data by companies constitute intellectual property? This is a grey area because partly it is controlled by trade secrets, partly by database sui -generis rights, partly by technical control. Under GDPR rights, as data subjects in Europe citizens / users have right to access and have information of this data. There is no definitive answer.

The starting point could be Recital 63 of GDPR , states "right to access personal data should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software".

This provision is very opaque as it concerns only with right to access (rather, one of the rights included in the concept of access, as seen below) and doesn't concern with other data subject rights such as the right to be informed as in the previous data protection regulation. A company under the previous legislation had a duty to inform about data processing and data appropriation but IP could prevail on this duty. Not it is no longer the case.

The other point to be deliberated upon is right to access is not a single right, it constitutes of 4 different rights viz:

Right to obtain information about whether personal data is processed
The right to obtain information about the purposes of processing.
Right to access the data.
Right to obtain free of charge a copy of the data.

Companies' IP rights should prevail solely on the 4th right, i.e. "Right to obtain free of charge a copy of the data". Companies should enable users to access appropriated data under the other three rights of Article 15. It is a delicate balance to be achieved and needs further discussion.

Prof. Lopez-Tarruella (Moderator) added to the debate the topic of the black-box problem. How to inform and provide a reply to data subjects to the access requests when the processor does not really know, or cannot really explain in an affordable way, how are these data being processed via a certain software or automated means.

Dr. Guellec opined restricting to the case of patient's data; European firms are going to China to access patient data. European Commission and lawyers should be aware of this. There is clearly excessive restriction on access to data in Europe which is not the best path for AI (artificial intelligence) development in Europe.

5. **The new AI regulation was a missed opportunity to introduced access to data in the context of developing trustworthy AI. AI data can have inherent bias which needs to evaluate or checked by public authorities. The AI regulation with no discussion on this, seems to be a missed opportunity, what are your views?**

Prof. Noto La Diega: AI regulation mimics Ethical guidelines on Trustworthy AI and there is confusion between ethics and law. Although he would generally agree that the draft AI Act is a missed opportunity, on the specific point of this question an answer may be provided by Article 64 whereby "the market surveillance authorities shall be granted full access to the training, validation and testing datasets used by the provider, including through application programming interfaces ('API') or other appropriate technical means and tools enabling remote access."

This is a positive step but is limited due to granting access solely to market surveillance authorities overlooking other stakeholders who would be interested in accessing through this right. It should be a distributed type of access and we should also reflect on the role of IP in all this as it can grant rights to enterprises to restrict access of such data.

6. **Access goes beyond data, should open data be interpreted to open software and internal data models and variables used by AI?**

Prof. Noto La Diega: Art 64 of the AI Act also offers possibility to access the source code of the AI system, if it is a high-risk AI system and there is a reasonable request. When we talk about open data it should also incorporate open software. However, open software is not a free ride for all, there are various forms of software which are mixed or in hybrid forms. To conclude, yes open data should mean access to the software used by the AI system.

Dr. Guellec opined that opening software nearly always is a good thing. But in AI, software which is written is just a learning mechanism, and the actual operating software is a huge number of parameters learnt from the data;in the case of GPT-3 there are over 170 billion parameters, making it impossible for anyone to open-up and look at that to check for confirmability. Results from Machine learning systems may differ from on training session to the next, making regulation hard to design and monitor. There are biases in humans much more than in machines, and the biases of machines come from the biases from humans. In machines one can identify and correct biases, but not in humans. There are many cases in which machines are less biased such as the case of appointing board directors and other top positions.

Sufficient patent disclosure is increasingly insufficient!

7. **The DABUS case saying that the AI generated inventions are not patentable. People might hide the behind the AI and claim to be inventors. Can an AI system read and analyze patent application to know or discover if it has a human inventor or an AI invention (invented by Artificial Intelligence system)?**

Dr. Guellec opined that the EPO negatived the notion of AI being an inventor. However, if a flood of inventions come from machines, Europe and EPO might have to adapt and relook at the possibility of machines being inventors. Considering also the new technological means under current development, such as quantum computation.

With regards to the notion of inventive step the bar might be elevated due to AI. The notion of Person skilled in the art might have to be changed to machine skilled in the art. The assessment of inventive step might get higher and difficult for human comprehension.

8. **Would this mean that we should open the possibility for machine inventors, that is granting patent rights to AI machines?**

Dr Guellec: Granting right to a machine is related to the legal status of machines. It is a larger legal question than just patents. It concerns with legal personality to be accorded to machines.

Prof. Noto La Diega: I am against granting rights to AI machines. Granting patents to AI inventors can leads to excessive monopolization thus leading to overprotection. During the current crises we have seen how in the case of vaccines private enterprises are prevailing on public interest. If we place this in the context of AI generated vaccines can result in further skewing the balance between private rights and public interests.

Humans will have to surrender and concede some advantage to the machines.

9. **Can false inventors (AI/ machine inventors) be identified by the requirement of sufficiency of disclosure?**

Prof. Noto La Diega: As an applicant one is supposed to sufficiently disclose the invention, however in highly complex fields it is largely seen that sufficiency is increasingly insufficient. Barriers to training data, clinic trials all make things complex to garner information out of patent description.

Dr. Guellec: Sufficiency is complex concept in patent law. In chemistry if there is a demonstration done by AI, humans cannot check it. There is tendency that machines will get further complex, and it will get harder for humans to comprehend the reasons and the methods followed by the machine to arrive at a certain result. At some point humans might have to surrender and concede some advantage to the machines. How to concede is an open question. We will have to trust the results of machines that we do not understand. The notion of discovery and transparency become fragilized in this context and humans have to cede some ground to the machines.

Francisco Javier LÓPEZ GUZMÁN
Girish Somawarpet