

The text and data mining exception: a lost opportunity
to boost European companies' competitiveness in AI

**文本和数据挖掘的例外规定：欧洲公司在加强人工
智能领域竞争力方面如何错失良机**

Aurelio Lopez-Tarruella

University of Alicante 阿利坎特大学

- Access to **vast amount of data** is essential for research in AI
- 可获取**海量数据**是研究人工智能(AI)的基本条件
- Part of that research consists on **text and data mining (TDM)**
- **文本和数据挖掘(TDM)**构成人工智能研究的一部分
 - “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**”
 - 任何旨在分析数字形式的文本和数据的**自动分析技术**，以生成包括但不限于**模型、趋势、相关性**等信息
- The TDM exception was introduced because:
- 引入文本和数据挖掘例外规定的原因：
 - “Mining” of a work is considered a “**reproduction**” of the work in the sense of copyright law
 - 从版权法的角度看，“**挖掘**”一项作品是对该作品的“**复制**”
 - “Right to read DOES NOT include the **right to mine**”
 - “**读取权并不包括挖掘权**”
 - Impossibility or great difficulty to apply the **existing exceptions** to TDM activities.
 - 将现有例外规则应用于文本和数据挖掘几乎不可能，或者存在极大困难
- **The TDM exception** is introduced in arts. 2.2, 3, 4 and 7 Directive 2019/790 (DCDSM)
- 欧盟第2019/790号指令的第2.2条、第3条、第4条、第7条（《单一数字市场版权指令》）引入了**文本和数据挖掘例外规定**。

- The exception covers **reproductions or extractions in order to carry out TDM** of works or other subject matters to which they have lawful access.
- 例外规定涵盖**以文本和数据挖掘为目的**，对合法获取的作品或其他内容进行**复制与提取**的行为。
 - Works, data bases and press publications.
 - 作品、数据库和新闻出版物
 - **Art. 3:** research organisations and cultural heritage institutions, research purposes
 - **第3条:** 研究机构和文化遗产机构，研究目的
 - **Art. 4:** anyone, for any purpose (including commercial purposes).
 - **第4条:** 任何人，出于任何目的（包括商业目的）

Common requirement: Lawful Access – Recital 14. Two situations

共同要求：合法获取——引言第14条（两种情况）

- Content available under **open content licenses**
- 基于**开放内容许可证**而可用的内容
 - the rightholder voluntarily so decides (e.g. Wikipedia),
 - 权利人自愿决定（比如，维基百科）
 - or he/she is under an obligation established by an Open Access / Open Data Policy (Directive 2019/1024 PSI, Horizon 2020, Commission Recommendation 2018/290)
 - 权利人需履行开放访问/开放数据政策（欧盟第2019/1024号指令，PSI指令，地平线2020计划，欧委会第2018/290号推荐）所规定的义务
- Content available under a “**contractual arrangements** ... such as subscriptions, or through other lawful means”.
- 基于“**契约性安排**.....比如订阅，或者其他合法途径”而可用的内容
 - Subscription services (e. g. Access to a data base, Netflix, Spotify)
 - 订阅服务（比如，访问数据库、Netflix、Spotify）
 - Services provided for free (e. g. YouTube, Facebook)
 - 免费服务（比如，YouTube、Facebook）

Research organisations vs. Private entities

研究机构与民间主体

- In the case of **research organisations** carrying out TDM activities “for the purpose of **scientific research**”:
- **当研究机构 “为了科学研究” 而开展文本和数据挖掘活动时：**
 - Any clause prohibiting or restricting TDM are **null and void** (art. 7.1)
 - 所有禁止或限制文本和数据挖掘活动的条款均**无效**（第7.1条）
 - E. g.: “Share Alike” clause.
 - 例如：“相同方式共享” 条款
- **Private entities or TDM for other purposes:**
- **民间主体或者出于其他目的开展文本和数据挖掘：**
 - **Opt-out mechanism** (Art. 4.3): “The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph **has not been expressly reserved by their rightholders in an appropriate manner**, such as machine-readable means in the case of content made publicly available online”
 - **选择退出机制**（第4.3条）：“第1款规定的例外或者限制应有条件适用，即**权利人没有以适当方式明确保留**对上述作品或其他内容的使用，例如针对网上公开提供的内容采取机器可读的方式。”

YouTube Terms of Service

YouTube 的服务条款

“Permissions and Restrictions 准许与限制

You may access and use the Service as made available to you, as long as you comply with this Agreement and the law. You may view or listen to Content for your personal, non-commercial use. You may also show YouTube videos through the embeddable YouTube player.

只要你遵守本协议和法律的规定，就可以访问和使用提供给你的服务。你可以出于个人、非商业目的来观看或收听相关内容。你还可以通过嵌入式YouTube播放器展示你的YouTube视频。

The following restrictions apply to your use of the Service. **You are not allowed to:**

以下限制条件适用于你对服务的使用。 **你不得：**

1. **access, reproduce,** download, distribute, transmit, broadcast, display, sell, license, alter, modify or otherwise use any part of the Service or any Content

部分 **获取、复制、** 下载、分发、传输、广播、展示、出售、许可、更改、修改或以其他方式使用服务或任何内容的任何一

...

10. ...

Reservation 保留条款

Any right not expressly granted to you in this Agreement remains the right of YouTube or the respective rights holders”.

凡本协议中未明确授予你的权利都归属YouTube或对应权利人。”

Research organisations vs. Private entities

研究机构与民间主体

- Obstacle for private entities (in particular **SMEs**) to research on AI, and to **access the market**.
- 民间主体（尤其是中小型企业）研究人工智能、**进入市场**的障碍
- The opt-out mechanism **does not seem justified** in the need to protect rightholders' interests: lawful access is required for TDM.
- 在保护权利人权益方面，选择退出机制**似乎并不合理**：开展文本和数字挖掘活动需要对数据的合法获取
- The discrimination does not seem **coherent with the EU's approach to public-private partnership** in research (Recital 11)
- 这种区别对待似乎与**欧盟就政府与民间开展研究合作所采取的方法不一致**（引言第11条）
- Does not seem **coherent with the EU's Open Access Policy** to facilitate that research reaches society (and private entities).
- 似乎与**欧盟为促进研究活动惠及公众（和民间主体）而制定的开放访问政策也不统一**
- EU regulation of the TDM exception is **less attractive** than the regulation in third countries where private entities are not discriminated (e. g. US, China, Japan).
- 与第三国（如美国、中国、日本）对民间主体的无区别对待相比，欧盟在文本和数字挖掘领域所作的例外规定显得**不尽如人意**

Thank You 谢谢

Aurelio.Lopez@ua.es

@lucentinus

[Http://www.lucentinus.es](http://www.lucentinus.es)