



Copyright and the training of
generative AI:

An EU perspective

版权与生成式人工智能(AI)
的训练： 欧盟视角

Eleonora Rosati

13 December 2023

2023年12月13日

AI “creativity”: how do we get ther? AI “创意”：我们如何做到？

- Overview of AI applications of machine learning process

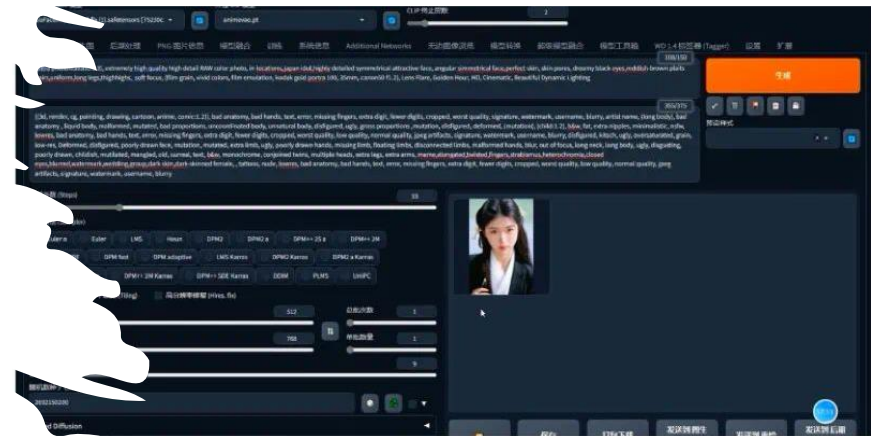
机器学习过程的AI 应用概览

- Text and data mining (TDM) at the basis of machine learning: The copyright dimension

机器学习的基础上文本和数据挖掘（TDM）：
版权角度

- Regulating TDM: Overview of approaches and the EU experience

监管TDM：方法及欧盟经验概览



(图 2)



(图 3)

6. 在上述参数不变的情况下，将“Additional-Networks”中的模型 lord-hanfugirl-v1-5. safetensors”的权重修改为 0.75。生成结果如图 4 所示。

Overview of AI applications of machine learning processes

机器学习过程的AI应用概览

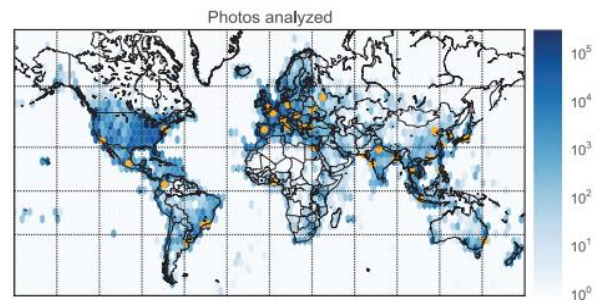
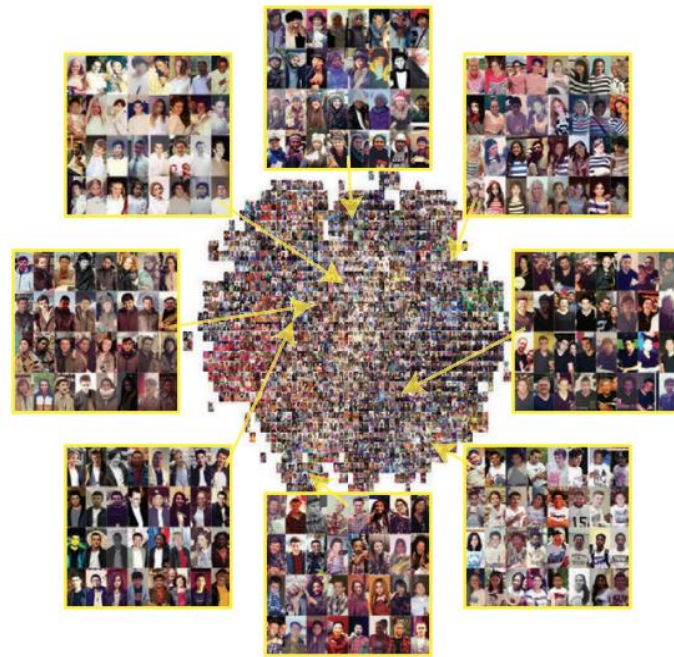
StreetStyle: Exploring world-wide clothing styles from millions of photos

Kevin Matzen *

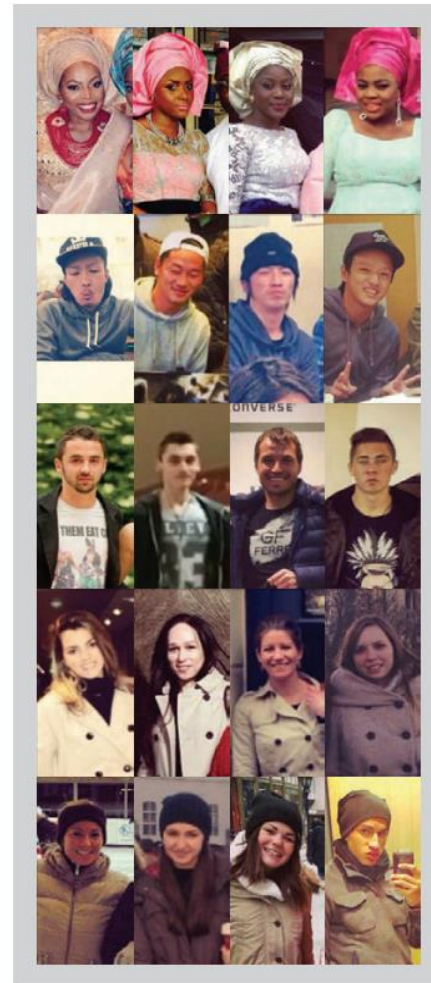
Kavita Bala †

Noah Snavely ‡

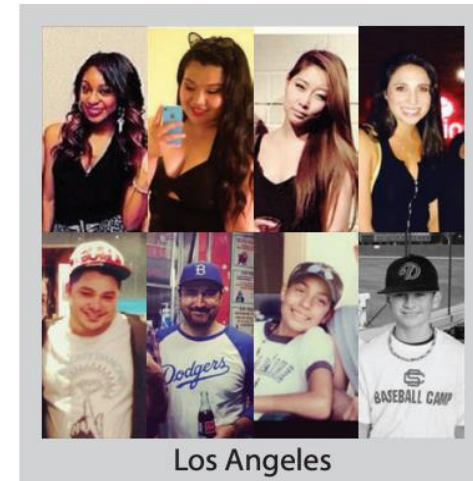
Cornell University



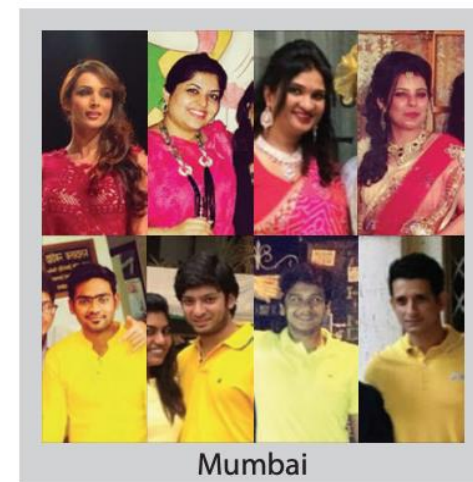
(a) Massive dataset of people



(b) Global clusters



Los Angeles



Mumbai

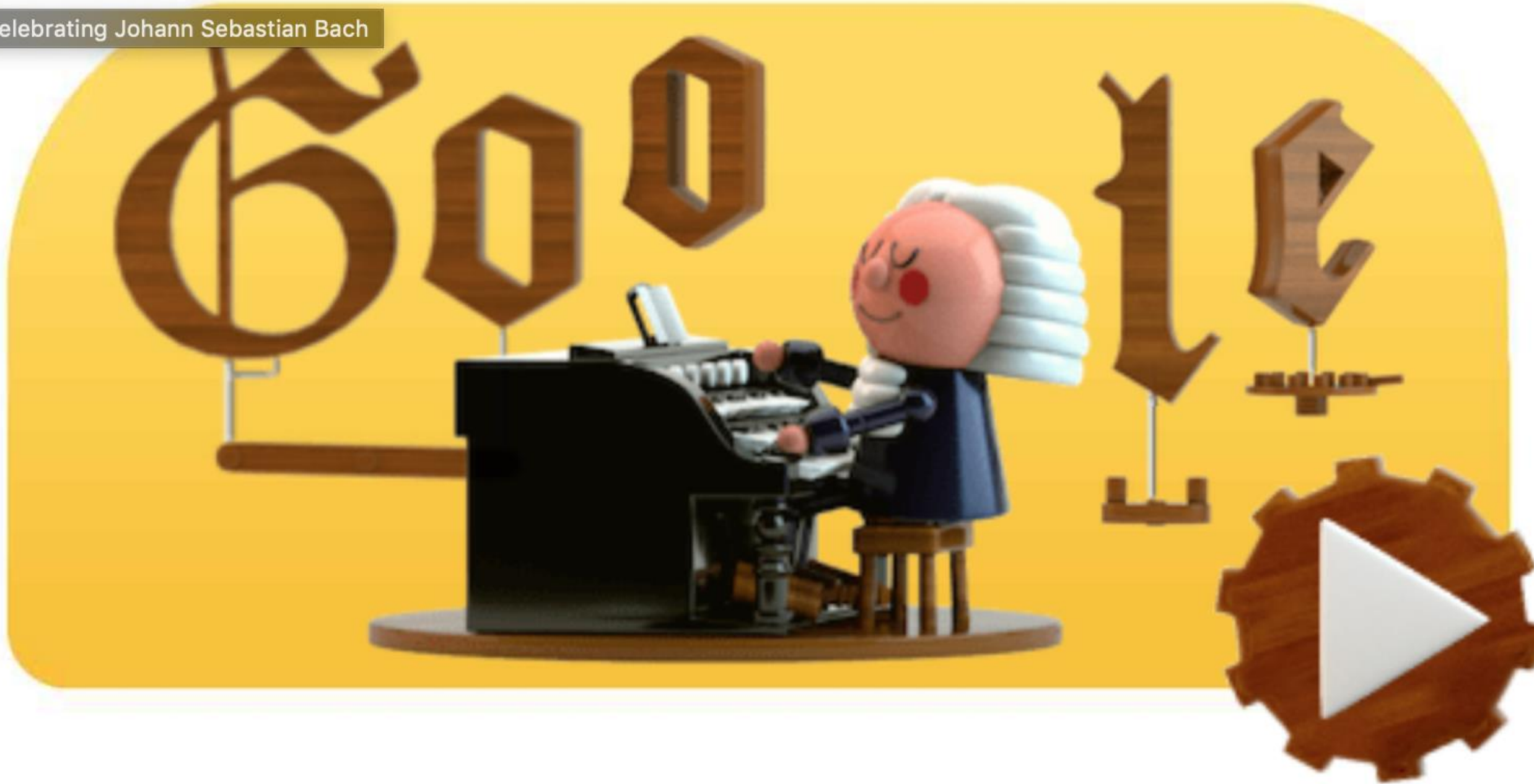
(c) Representative clusters per-city

By

DECI

r end

Celebrating Johann Sebastian Bach



<https://www.google.com/doodles/celebrating-johann-sebastian-bach>



<https://uncrate.com/ai-generated-portrait-of-edmond-belamy/>



A Recent Entrance to Paradise ([Thaler v Perlmutter](#))
近访仙境之门 (([Thaler](#) 诉 [Perlmutter](#)))

Photographer admits prize-winning image was AI-generated

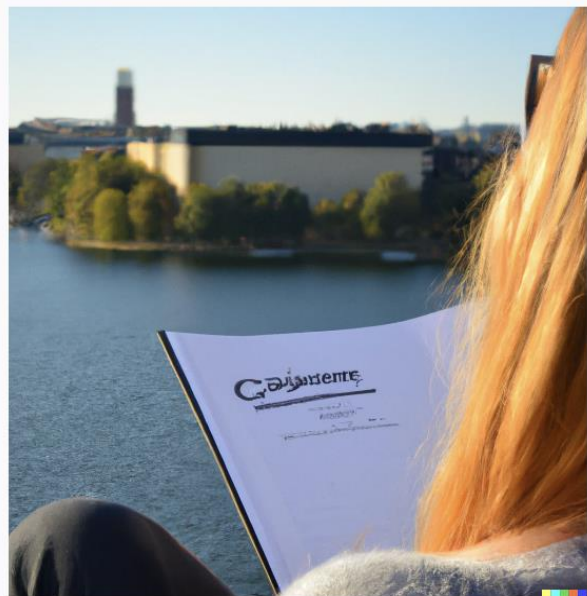
German artist Boris Eldagsen says entry to Sony world photography awards was designed to provoke debate



DALL·E History Collections

Edit the detailed description

Student in Stockholm studying copyright law



Surprise me

Upload



Generate

+ New chat

Today
New chat

New chat

Previous 7 Days

Copyright Definition

January

New chat

Decline Party Invitation

Upgrade to Plus **NEW**

Eleonora Rosati



write a university essay on the notion of authorship in copyright law
请写一篇有关版权法对作者归属权问题的大学文章

简介:

著作权是版权法的核心。版权法通过把专有权给到作品的作者或创作者来对原创作品的著作权提供法律保障, 比如: 文学作品、艺术作品和音乐作品。该文章会论述版权法当中著作权的问题, 并讨论与之相关的法律及哲学问题。

定义:

著作权指的是创造一件原创艺术品、文学作品或音乐的行为。版权法对作者的定义是创造了一件原创作品的人...



Introduction:

The concept of authorship is central to copyright law. Copyright law provides legal protection to original works of authorship, such as literary, artistic, and musical works, by giving exclusive rights to the authors or creators of these works. This essay will explore the notion of authorship in copyright law and examine the legal and philosophical issues that arise from this concept.

Defining Authorship:

Authorship refers to the act of creating an original work of art, literature, or music. The Copyright Act defines an author as the person who creates an original work of authorship that is fixed in a

Stop generating

Send a message.



Drake的声音、喜好、
对自己形象的控制以
及讲述自己的故事，
这些都被偷走了。

Patrick Rogers
CEO
加拿大音乐

“

[Drake's] voice,
his likeness,
his right to control his image,
and tell his own stories
were **stolen** from him.

Patrick Rogers
CEO
Music Canada

Text and data mining (TDM) at the basis of machine learning: The copyright dimension

机器学习基础上文本和数据挖掘 (TDM) : 版权角度

Text and data mining (TDM): How do you get to Step 3? 文本和数据挖掘 (TDM) : 如何到达第三步?

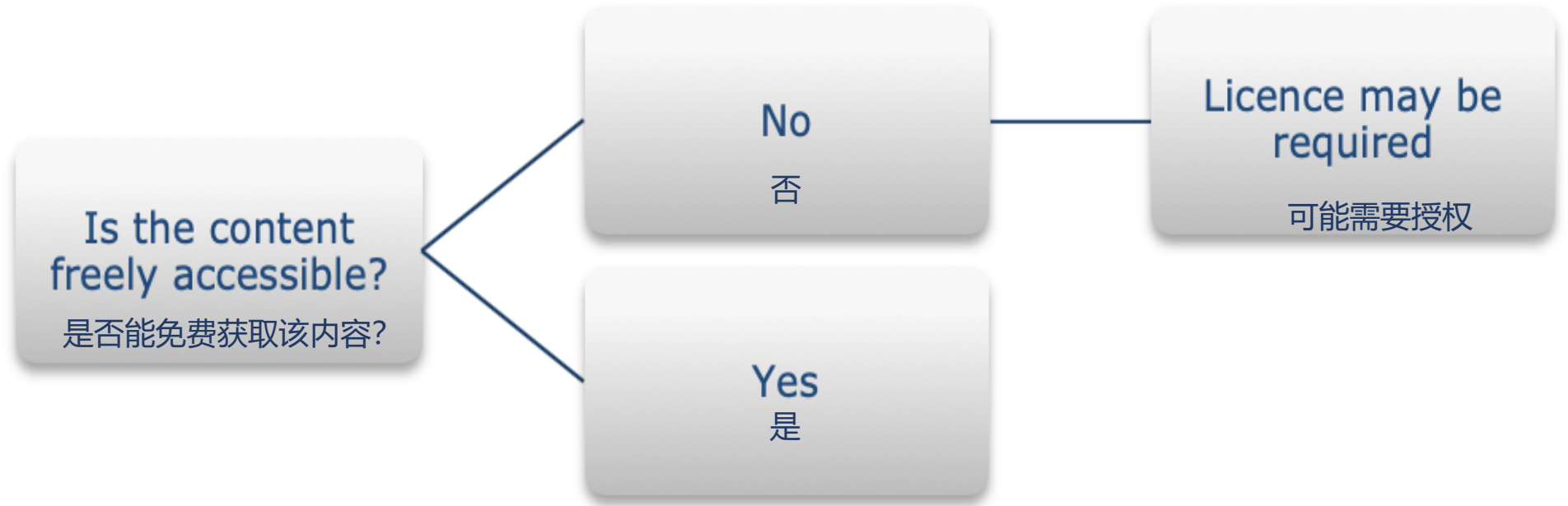
Step 3 - Mining of text and/or data and knowledge discovery

第3步：文本挖掘和/或数据及返现知识



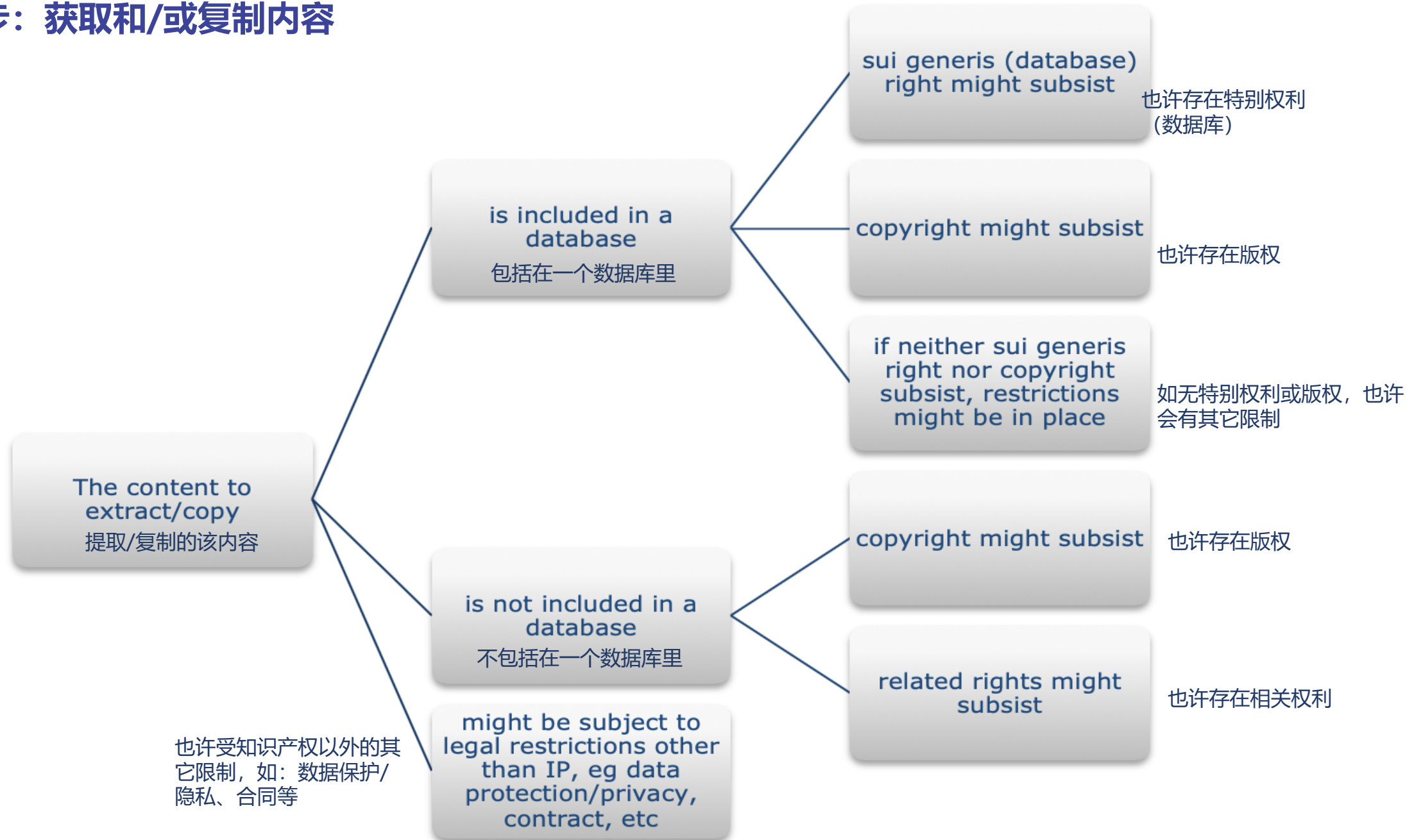
Step 1 – Access to content

第1步：获取内容



Step 2 – Extraction and/or copying of content

第2步：获取和/或复制内容



HOME > DIGITAL > NEWS

Nov 1, 2023 11:57am PT

BOT COURT

New ChatGPT Laws AI's Legal Sh-tstorn App That Ripped Off Her Likeness in Advertisement

OpenAI faces allegations of privacy inv but this may be just the tip of the icebe

BY MILES KLEE

By Ethan Shanfeld



Business Politics Culture Tech

s OpenAI and



Regulating TDM: Overview of approaches and the EU
experience

监管TDM：方法及欧盟经验概览

Approaches to TDM

TDM方法

USA: "to furnish statistical information to Internet users about the frequency of word and phrase usage over centuries" Authors Guild v Google, Inc, No. 13-4829 (2d Cir. 2015)

Fair use doctrine: Section 107 US Copyright Act

美国: "统计互联网用户在几个世纪以来单词和短语使用的频率情况", 作者工会诉谷歌公司第13-4829 (2d Cir. 2015)

合理使用原则: 《美国版权法》第 107 条

Japan (non-enjoyment), Singapore, but not Hong Kong

日本 (非娱乐)、新加坡, 但不包括香港

Certain EU Member States (first was UK – Article 5(3)(a) InfoSoc Directive)

某些欧盟成员国 (首先是英国——《信息社会指令》第 5(3)(a)条)

Now 2 mandatory exceptions at the EU level (Articles 3 and 4 DSM Directive 2019/790)

现在, 欧盟层面有 2 项强制性例外规定 (DSM 指令 2019/790 第 3 和第 4 条)。

But what is the
scope?
The input / output
divide
但范围是什么?
输入/输出的区分

Articles 3 and 4 DSM Directive 2019/790 DSM 指令 2019/790 第 3 条和第 4 条

	Art 3 第3条	Art 4 第4条
Beneficiaries 受益人	Research organizations and cultural heritage institutions 研究机构和文化遗产机构	No restrictions 无限制
Public-private partnerships 公私合作	👍 (but not if decisive influence) (但是如果具有决定性影响力则不符合)	N/A 不适用
Scope 范围	<ul style="list-style-type: none"> • Rights of reproduction under Directive 96/9 and Directive 2001/29 按照96/9指令和2001/29指令下的复制权 • Press publishers' right 媒体出版权利 • Database (sui generis) right in Directive 96/9 按照96/9指令的数据库(特别)权利 	Also rights in computer programs 也包含计算机变成权利
Storage of copies 复制件存储	v (with appropriate security level) (按照合适的安全等级)	Not mentioned 未提及
Retention of copies 复制件保留	👍 (for scientific research purposes) (科研目的)	👍 (but as long as necessary for TDM purposes) (只要TDM目的是有必要的)
Reservation 保留	⊖	"in an appropriate manner" (以合适的方式)
Compliance with three-step test 符合三步测试要求	👍	👍

Key issue is reservation under Article 4

第 4 条下面的主要问题是保留

English ▾ Newsletters Podcasts Events Jobs f X in v s @ d o B

Login/Register



EURACTIV

The Capitals

The Brief

Ukraine

Intelligence

Agrifood

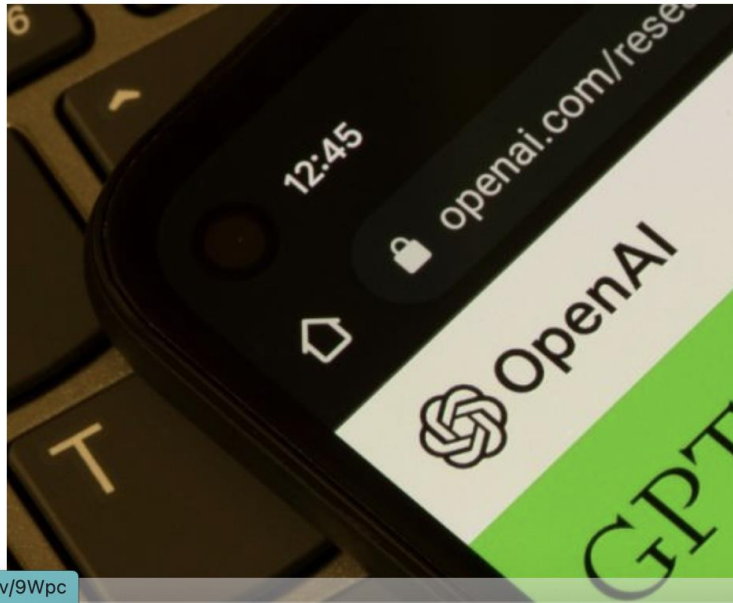
Economy

Energy & Environment

Spanish presidency pitches 0 AI law

By Luca Bertuzzi | Euractiv.com ⌚ Est. 6min

Content-Type: News



Foundation models


The Foundation model is defined as “a large AI model that is trained on a large amount of data, which is capable to competently perform a wide range of distinctive tasks, including, for example generating video, text, images, conversing in lateral language, computing or generating computer code”.

These models must comply with transparency obligations, namely, provide AI system providers with up-to-date technical documentation explaining the capacities and limitations of foundation models and a set of elements to be detailed in an annex.



Additionally, foundation model providers will have to demonstrate that they have taken adequate measures to ensure the training of the model has taken place in compliance with EU law related to copyright, in particular, the disposition that requires data mining activities of content made publicly available online to require the consent of rightsholders, including by machine-readable means, if rightsholders have opted out of the copyright exception for text and data mining.

Thus, the foundation model developers will have to put in place a system to respect the opt-out decisions of content creators.



Conclusion

结论

- Litigation ongoing in several countries around the world
全球多个国家正在进行诉讼
- Key point: Input phase more important than output protection
关键点：输入阶段比输出阶段保护更重要
- Global race to the bottom or race to the top?
全球竞相逐低还是竞相逐高？



Thanks for your attention!
感谢聆听!

eleonora@elawnora.com

<https://www.linkedin.com/in/eleonorarosati/>

Further reading:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4629528