大模型之司法篇

在这个教程中，我们将探讨法律对大型语言模型的开发和部署有何规定。我们将会按照以下的步骤进行讨论：

1. **新技术与现有法律的关系**

   与我们之前的讲座一样，比如关于社会偏见的讲座，我们将要讨论的很多内容并不一定特指大型语言模型（并没有特别的大型语言模型法律条例）。然而，每当有新的强大的技术出现，它都会引发很多关于现有法律是否仍然适用或有意义的问题。例如，随着互联网的重要性日益提高，互联网法律（或称为网络法）应运而生。它从现有的领域中汲取知识，如知识产权法，隐私法，和合同法等。

2. **互联网的独特挑战**

   法律通常有明确的管辖范围（例如，州，联邦），但互联网并不受地理限制。在互联网上可以保持匿名，任何人都可以发布一段内容，理论上可以被任何人查看。

3. **法律与道德的区别**

   法律可以由政府强制执行，而道德无法强制执行，可以由任何组织创立。例如医师的希波克拉底誓言、ACM的道德与职业行为准则、NeurIPS的行为准则和斯坦福的荣誉准则等。

4. **法律的管辖权问题**

   根据你所在的地方（哪个国家，哪个州等），适用的法律会有所不同。例如，欧盟的数据隐私法GDPR比美国的法律更全面。法律可以在联邦、州或地方级别存在。

5. **法律的类型**

   常见的法律类型包括普通法（司法）、成文法（立法）和监管法（行政）。

6. **大型语言模型**

   我们将会把注意力转向大型语言模型。回忆一下大型语言模型的生命周期：收集训练数据，训练大型语言模型，将其适应到下游任务，向用户部署语言模型。

   在大型语言模型的生命周期中，有两个主要领域与法律交叉：数据和应用。

7. **数据**

   所有的机器学习都依赖于数据。语言模型依赖于大量的数据，尤其是为其他

目的制作的他人的数据，这些数据往往在未经许可的情况下被抓取。知识产权法保护数据的创作者，那么在这些数据上训练语言模型是否构成侵犯版权？隐私法保护个人隐私权，那么在公开或私密数据上训练语言模型是否可能侵犯隐私？对于私密数据，何时可以收集和汇总这些数据?

8. **应用**

   语言模型可以被用于广泛的下游任务（例如，问答，聊天机器人）。技术可能被有意用于伤害（例如，垃圾邮件，网络钓鱼攻击，骚扰，假新闻）。现有的互联网欺诈和滥用法律可能覆盖其中的一部分。他们可以被部署在各种高风险的环境（例如，医疗，贷款，教育）。现有的在相关领域的规定（例如，医疗）可能覆盖其中的一部分。

   大型语言模型的扩展能力（例如，真实文本生成，聊天机器人）将带来新的挑战。

版权法

大型语言模型或任何机器学习模型，都是基于数据进行训练的，而这些数据是人类劳动的结果（例如，作者，程序员，摄影师等）。除了创作者外，其他人可以对这些创作（例如，书籍，代码，照片等）进行何种使用，属于知识产权法的范畴。

知识产权法

其动机是鼓励创建各种类型的知识产品。如果任何人都可以利用你的辛勤劳动并从中获利，人们就会对创造或分享失去动力。知识产权包括：版权，专利，商标，商业秘密。

在美国，决定版权的关键法规是1976年的《版权法》。版权保护适用于“已经以某种可以感知、复制或以其他方式直接或通过机器或设备传达的有形媒介中固定下来的原创性作者作品”。1976年版权法扩大了版权保护范围，从“已发布”（1909年）扩大到“已固定”。虽然不需要登记就可以获得版权保护，但在起诉他人侵犯版权之前，创作者必须注册版权。版权保护期限为75年，然后版权到期，作品成为公有领域的一部分（如莎士比亚、贝多芬的作品等）。

使用版权作品有两种方式：获取许可或依赖公平使用条款。

许可

许可（来自合同法）是由许可人授予许可使用者的。实际上，“许可就是承诺不起诉”。创作共享许可，允许免费分发版权作品。例如，维基百科、开放课程、可汗学院、免费音乐档案、来自Flickr的307百万图像、来自MusicBrainz的39百万图像、来自YouTube的1000万视频等。

公平使用（第107条）

自1840年代以来，公平使用一直是普通法。决定是否适用公平使用的四个因素是：

1. 使用的目的和性质（教育用途优于商业用途，转型用途优于复制）；
2. 版权作品的性质（虚构作品优于事实作品，创新性的程度）；
3. 使用的原作部分的数量和实质性；和
4. 使用对原作市场（或潜在市场）的影响。

服务条款可能会

增加额外的限制。例如，YouTube的服务条款禁止下载视频，即使视频在创作共享下许可。

注意：事实和想法不受版权保护。如果策划/安排被视为表达，事实数据库可以受版权保护。复制数据（训练的第一步）就已经是侵权，即使你不做任何事情。法定损害赔偿可以高达每件作品150,000美元（版权法第504条）。

案例研究

接下来，我们将回顾一些已经裁定公平使用或反对公平使用的案件。

作家协会诉Google

Google Book Search扫描了印刷书籍并使其在线可搜索（显示片段），始于2002年。作家协会抱怨Google没有寻求他们对仍受版权保护的书籍的许可。2013年，地区法院判定Google公平使用。

Google诉Oracle

Google在Android操作系统中复制了Oracle（原Sun Microsystems）所有的37个Java API。Oracle以版权侵权起诉Google。2021年4月，最高法院裁定Google的使用Java API属于公平使用。

Fox News诉TVEyes

TVEyes记录电视节目，创建了一项服务，使人们可以搜索（通过文本）并观看10秒片段。Fox News起诉TVEyes。2018年，第二区裁定赞成Fox News，不是公平使用。

Kelly诉Arriba

Arriba创建了一个显示缩略图的搜索引擎。Kelly（个人）起诉Arriba。2003年，第九巡回法庭裁定赞成Arriba，认为其属于公平使用。

Sega诉Accolade

1989年，Sega Genesis游戏主机发布。Accolade想要在Genesis上发布游戏，但Sega收取额外费用，希望成为独家发行商。Accolade反向工程Sega的代码，制作新版本，绕过安全锁。Sega在1991年起诉Accolade。1992年，第九巡回法庭裁定赞成Accolade，认为其属于公平使用。

公平学习与机器学习

公平学习主张机器学习属于公平使用。机器学习系统的数据使用是变革性的，它不会改变作品，但会改变目的。机器学习系统对想法感兴趣，而不是具体的表达。

对于将机器学习视为公平使用的论据：训练数据的广泛访问会为社会创造更好的系统。如果不允许使用，那么大部分作品无法用来产生新

的价值。使用版权数据可能更公平。

反对将机器学习视为公平使用的论据：认为机器学习系统不会产生创意的“最终产品”，而只是赚钱。生成模型（例如，语言模型）可以与创意专业人士竞争。机器学习系统的问题（传播假信息，实现监控等），因此不应该给予机器学习系统利益的怀疑。

在版权法下，很难分离可保护的（例如，表达）和不可保护的（例如，想法）。虽然构建机器学习系统可能有很多原因不妥，但版权是阻止它的正确工具吗？对于训练大型语言模型是否属于公平使用的问题正在迅速发展。

阶段性结论

查看信息技术的历史，我们可以看到三个阶段：

1. 第一阶段：文本数据挖掘（搜索引擎），基于简单的模式匹配。
2. 第二阶段：分类（例如，分类停止标志或情感分析），推荐系统。
3. 第三阶段：学习模仿表达的生成模型。

上次，我们看到从GPT-2中提取训练数据可能会出现隐私问题。如果语言模型直接复制哈利·波特，那么这对公平使用来说是有问题的。然而，即使语言模型不直接生成以前的作品，版权仍然相关，因为以前的受版权保护的作品被用来训练语言模型。

事实上，语言模型可以与作家竞争。例如，作家写了3本书，语言模型在这3本书上进行训练，并自动生成第4本。

因此，面对大型语言模型，版权和机器学习的未来还未知。

隐私法律教程

在本教程中，我们将简要讨论一些隐私法律的例子，包括Clearview AI、加利福尼亚消费者隐私法案（2018）、加利福尼亚隐私权法案（2020）以及欧盟的一般数据保护条例（GDPR）。

Clearview AI

Clearview AI是一家成立于2017年的公司。2019年，纽约时报曝光了它。到2021年10月，该公司已经从Facebook、Twitter、Google、YouTube、Venmo等网站抓取了100亿张人脸图片。该公司将数据销售给执法机构（例如，FBI）和商业组织。该公司辩称有权使用公开的信息。由于侵犯隐私，该公司已被起诉。

伊利诺伊州生物识别信息隐私法（2008）

这项法律通过私人实体对生物识别标识符进行监管（不包括政府实体）。Clearview删除了伊利诺伊州的数据。欧盟汉堡数据保护机构（DPA）认为该行为违法。

加利福尼亚消费者隐私法案（2018）

这项法案赋予加利福尼亚居民以下权利：

- 了解收集他们的哪些个人数据。
- 了解他们的个人数据是否被出售或公开，以及给了谁。
- 拒绝个人数据的销售。
- 访问他们的个人数据。
- 请求业务删除从消费者处收集的任何个人信息。
- 不因行使他们的隐私权利而被歧视。

个人数据包括：真实姓名、别名、邮寄地址、唯一个人标识符、在线标识符、IP地址、电子邮件地址、账户名称、社会保障号码、驾驶执照号码、车牌号码、护照号码等。

该法适用于在加利福尼亚经营且年收入至少为2500万美元的企业。美国联邦尚无相应法律。与GDPR不同，这项法律不允许用户更正数据。

加利福尼亚隐私权法案（2020）

这项法案创立了加利福尼亚隐私保护机构，将于2023年1月1日生效，适用于2022年1月1日之后收集的数据。

意图：

- 了解谁在收集他们及其孩子的个人信息，如何使用，以及向谁公开。
- 控制他们个人信息的使用，包括

限制他们敏感个人信息的使用。

- 访问他们的个人信息并有能力纠正、删除和转移他们的个人信息。
- 通过易于获取的自助工具行使他们的隐私权利。

- 行使他们的隐私权利而不受罚款。
- 将未采取合理信息安全预防措施的企业追究责任。
- 从企业使用他们的个人信息中受益。
- 作为员工和独立承包商也能保护他们的隐私利益。

GDPR（欧盟一般数据保护条例）

该规定是欧盟法律关于数据隐私的一部分，于2016年通过，2018年可执行。其范围比CCPA更广泛。不适用于处理个人数据的国家安全活动或执法行为。数据主体可以同意处理个人数据，并可以随时撤回。人们应有权访问自己的个人数据。因为在Android手机设置过程中未获得广告个性化的同意，Google被罚款5700万美元。

其他法律

加利福尼亚的机器人披露法案：

如果使用机器人与人进行通信，而不披露它是一个机器人，这是违法的。限制：只适用于激励销售或影响选举投票的情况。限制：只适用于每月在美国有1000万访问者的公开网站。

总结

在我们训练大型语言模型时，我们必须面对版权和公平使用的问题。由于网络爬取的未筛选性质，你必须诉诸公平使用（从每个人那里获得许可证将非常困难）。模型的生成性可能会对争论公平使用提出挑战（可以与人类竞争）。在什么水平上进行调控（语言模型还是下游应用）是有意义的？这个领域正在迅速发展，需要深入的法律和人工智能专业知识才能做出明智的决定！