海外智库观点要览
北京大学国际战略研究院
2024年5月13日 第99期

欧盟《人工智能法案》：未雨绸缪，抑或为时过早？

伯廷·马滕斯（Bertin Martens）[1]

编者按：欧洲议会于2024年3月13日投票通过了全球首部人工智能领域的全面监管法案——欧盟《人工智能法案》。欧洲议会表示，该法案旨在保护公民基本权利、民主、法治和环境可持续性免受高风险人工智能的危害，同时促进人工智能领域的创新，确立欧洲在该领域的领先地位。本期推荐的文章认为，目前尚不清楚该法案是会有效规范人工智能的应用，还是会扼杀人工智能领域的创新。

世界各国政府正在制定法规，以应对人工智能带来的风险。美国发布了一项关于监管人工智能的行政令。英国发布了一份不具约束力的原则宣言。中国实施了一项不过分干预、对商业友好的人工智能法规，主要是为了释放加快技术进步的信号。2024年3月13日，欧洲议会通过了欧盟委员会于2021年4月提出的、全球首部人工智能领域的全面监管法案——《人工智能法案》。

一、欧盟《人工智能法案》的宗旨

欧盟《人工智能法案》本质上是一项产品安全法规，旨在降低使用人工智能系统给人类带来的风险。最新的通用大型语言模型和生成式人工智能系统可以被塑造成用途近乎无限的模型，很难评估所有风险，并为所有可能的用途制定法规。该法案试图通过一般性的义务，即避免损害人的基本权利，来绕过这个难题。根

[1] 伯廷·马滕斯（Bertin Martens）是比利时布鲁盖尔研究所（Bruegel）的高级研究员。本文英文原文登载于布鲁盖尔研究所官方网站：https://www.bruegel.org/analysis/european-union-ai-act-premature-or-precocious-regulation。此为中文摘译版。

据欧洲议会一位该法案的联合设计者的说法，这种产品安全与基本权利标准的监管组合并不适合人工智能模型。

《人工智能法案》根据风险水平对在欧盟境内使用的人工智能系统进行了分类。大多数人工智能应用程序都被认为风险极小，不必受到监管。存在有限风险的系统仅受到透明度和用户告知义务的约束。被认为构成不可接受风险的系统则被禁止，包括远程生物识别及分类系统、面部识别数据库和社会信用评分系统，除非有医疗和安全方面的理由。《人工智能法案》侧重于监管介于风险有限和风险不可接受之间的高风险人工智能系统。这些系统用途单一或用途有限，在教育、就业、公共服务等领域与人类互动。法案包含一套复杂的规则和要求，以评估高风险系统是否以及在何种条件下可以被使用。

除了高风险人工智能系统，还有通用人工智能（GPAI）模型，因为它们的用途十分广泛。GPAI 提供商必须提供技术文档和使用说明，除非它们是开放许可模型，用户可以根据自己的目的进行修改。用于训练的数据必须记录在案，且必须符合《欧盟版权指令》（EU Copyright Directive）。

二、人工智能模型的基本人权保障和风险

人工智能界投入了大量精力，以使人工智能以人类为本，使人工智能模型的回答与人类价值观保持一致，避免歧视和有害的回答。这与当代关于多元、公平和包容的辩论相呼应。但是，还有很多其他经常被使用的歧视标准。例如，在被用于提升经济服务和补贴的针对性时，价格和收入歧视既可能增加福利，也可能是剥削性的。无论允许还是禁止这种行为，肯定会在其中一个方向上犯错误。那么问题来了：我们应该与谁的价值观、损害和收益保持一致？人工智能已经被用于国防和战争，这是一种以人为本的应用吗？

GPAI 开发者试图通过在模型中构建“护栏”来确保对人类价值观的尊重。然而，有很多方法可以绕过这些“护栏”。一些开发者还试图建立“宪法式护栏”，让模型自我检查其回应是否遵守了义务。

开放式 GPAI 模型更容易出现可以规避“护栏”的漏洞。但其开放性可能会刺激创新型应用程序和新商业模式的产生，这对小规模人工智能企业尤其重要。《人工智能法案》使开放式模型的开发者得以脱身，法案免除了其测试义务，并

将责任交给了可以修改这些模型行为的再开发者和部署者。当多个层次的配套应用程序交互时，追踪其合规性会更加困难。

规模较小的模型也避开了《人工智能法案》规定的严格义务，因为它们未达到算力门槛。除了降低训练成本外，这种豁免还降低了合规成本。虽然较小的模型可以和大模型一样用途广泛，但它们给出的回答通常不太准确，除非获得额外的提示和用户数据。因此，模型的规模小并不一定意味着风险低。

《人工智能法案》要求大模型开发者向下游部署者和服务供应商解释模型如何与不属于该模型的硬件或软件交互。这是一项非常笼统的规定，需要通过法案的实施进一步加以澄清。这引出了有趣的问题：GPAI 模型与配套服务之间的纵向和横向整合，各方的责任如何分配？当在线旅游服务平台引入人工智能应用程序来改善服务时，谁是部署者？是应用程序供应商还是平台？可能有多层部署者——这一问题目前未被《人工智能法案》涵盖，但在实施指南中可能会有解释。

《人工智能法案》并不完整，未能为人工智能开发者和部署者提供法律确定性。此外，它还产生了高昂的合规成本，尤其是对于中小企业和初创企业来说，它们可能会认为欧盟的监管环境成本太高、风险太大。然而，该法案为欧盟委员会及其新成立的人工智能办公室的进一步监管工作奠定了基础。该办公室将对人工智能开发者发送的通知进行登记和验证。然而，该办公室资源有限，需要一段时间才能进入状态。它必须制定十几项实施细则和指南。此外，在必要时，它可以澄清被禁止的人工智能行为、对高风险系统的要求以及透明度义务。这可能会扩大或收紧对在欧盟运营的人工智能模型开发者进行监管的空间。

三、《人工智能法案》与竞争

人工智能开发者之间的竞争非常激烈，目前还没有出现垄断的迹象，除了大型科技公司在人工智能计算基础设施层面明显占据主导地位。

欧盟境内目前还没有特大型人工智能模型。监管机构可能会依赖《人工智能法案》的“布鲁塞尔效应”：如果其他国家采用类似的法规，就会创造公平的竞争环境，削弱开发者为规避合规成本而转移到其他地方的动机。此外，得到欧盟的认可可能会使人工智能模型更具吸引力和竞争力。然而，对大模型严格而昂贵的监管可能会进一步固化小型人工智能开发者的市场定位，即只开发小型模型，

使其能够规避对系统性风险的监管，但也远离技术前沿。

现阶段，人工智能技术竞争政策的影响尚不清楚。前沿模型的开发和训练耗资数亿美元，往往超出了人工智能初创企业的能力范围。初创企业通常由前大型科技公司员工创建，更具创新性，也更接近人工智能技术前沿。然而，它们需要与大型科技公司达成密切合作的协议，给予大型科技公司对模型的访问权限，以换取对方提供昂贵的计算基础设施和数据。在各类任务中表现足够出色的小型人工智能模型在多大程度上能够与大型公司和模型竞争，仍然是一个悬而未决的问题。

《人工智能法案》只关注对独立大型人工智能模型的监管，但去中心化人工智能生态系统迅速崛起，人工智能模型正越来越多地通过应用程序和插件与互补性和竞争性平台及软件交互，这意味着无法通过关注单个模型来评估系统性风险，而是需要审视整个系统。风险在开发者、部署者和用户之间转移。应该在多大程度上允许模型与其他系统组件之间进行纵向和横向的互补和集成才能避免扭曲市场？

四、人工智能与版权

训练模型需要大量的数据输入，包括从网页上收集的文本、扫描的文档和书籍、从网络上收集的图像、从电影档案中收集的视频，以及从音乐收藏中收集的声音。其中大部分内容受到版权的限制。《欧盟版权指令》规定，如果用户对输入的内容有合法访问权（例如没有破解付费墙），那么文本和数据挖掘就适用版权保护例外情况。版权持有者可以选择不适用该例外情况并收取费用。几家新闻出版商已经与有能力支付费用的大型人工智能开发者达成了许可协议。人工智能初创企业则正在等待几起悬而未决的法庭案件的结果，这些案件应该会澄清对人工智能应用版权法的解释。授权的数据集可能质量更高，并降低训练成本，但也可能使可用的数据集减少，导致有偏见的训练。此外，对训练输入内容授予版权使版权所有者的个人利益能够影响到人工智能模型产生的更广泛的社会福利。这些模型正迅速成为一种通用技术，被用于所有经济领域，远远超出了对版权有个人利益的创意媒体行业。

使用人工智能的创意艺术家开始要求对输出内容进行版权保护。《人工智能

法案》规定，人工智能的视听和文本输出内容应具有机器可读的水印，以区别于人类输出和深度伪造。水印技术还处于初级阶段，很容易被规避。如果人工智能只是协助人类，则不适用水印义务。在大多数国家，只有人类输出的内容才能获得版权，机器输出的内容则不行。混合输出内容中需要有多少人类贡献才能要求版权？一句人类写的“提示词”可能不够。《欧盟版权指令》可能需要重新考虑这些问题。

目前的《欧盟人工智能法案》只是漫长监管过程的开始。它将起草实施细则和指南的责任委托给欧盟委员会及其新成立的人工智能办公室。这些实施方案和指南将推动法案的执行，并决定法案的出台是未雨绸缪，能够促进值得信赖的人工智能创新，还是为时过早，反而会扼杀创新。

（王润潭摘译，归泳涛校）