**人工智能法专题讨论**

**姓名：付政烨 学号：2113203 学院：网络空间安全学院**

**一、立法背景**

**（一）世界角度**

随着智能化社会的迅速发展，全球范围内的国家纷纷加强对人工智能技术的研究与应用布局。人工智能技术已广泛应用于交通、教育、服务等多个领域，但随之而来的安全问题也愈发显著。研究指出，人工智能面临着包括数据风险（如数据投毒、泄露和异常）、算法风险、网络风险等多种挑战。此外，人工智能在处理复杂问题时可能带入种族和性别歧视的偏见。

例如，COMPAS 是一种在美国广泛使用的算法，通过预测罪犯再次犯罪的可能性来指导判刑，而这个算法或许是最臭名昭著的人工智能偏见。根据美国新闻机构 ProPublica 在 2016 年 5 月的报道，COMPAS 算法存在明显的「偏见」。根据分析， 该系统预测的黑人被告再次犯罪的风险要远远高于白人，甚至达到了后者的两倍。



可能你在直觉中也会认识黑人的再犯率会高于白人，但这并不和实际情况相符。在算法看来，黑人的预测风险要高于实际风险，比如两年内没有再犯的黑人被错误的归类为高风险的几率是白人的两倍（45% 对 23%）。而未来两年内再次犯罪的白人被错误认为是低风险的概率同样是黑人再犯将近两倍（48% 对 28%）。

当前，强化人工智能的安全问题已成为国际间共同关注的重点。全球主要经济体纷纷推出针对人工智能发展的政策和立法措施，以确保这一领域的健康与安全发展。

1. 2017年，中国国务院发布《新一代人工智能发展规划》，强调了建立健全的人工智能相关法律、伦理和社会问题的研究，提出建立一套完善的法律法规和伦理道德框架以及人工智能安全监管和评估体系的必要性，以保障人工智能技术的健康发展。
2. 美国在2016年和2019年分别发布了《国家人工智能研发战略规划》及其更新版，其中将人工智能安全建设作为国家战略的重要方向之一。这些文件体现了美国在确保人工智能技术发展同时，对潜在风险的防控给予高度重视。
3. 欧洲联盟在2021年迈出了更为具体的步骤，发布了《人工智能法案》（Artificial Intelligence Act），这是一个系统性且全面的立法框架，专门针对人工智能技术的风险预防提供保障机制和实施路径。该法案详细规定了不同风险级别的人工智能系统的监管要求，以及对高风险人工智能系统的严格管控。

**（二）欧盟角度**

1. 欧盟意识到单一国家的法规措施可能无法充分应对当下人工智能技术带来的挑战，因为AI的跨界性质和内部市场中产品与服务的自由流通特点使得单一国家的法规可能导致市场碎片化，增加法律不确定性，阻碍技术的均衡发展。
2. 各成员国法规的**差异**可能会造成AI系统的使用、责任、监管等方面的不一致，给AI提供商和使用者带来合规成本和法律风险。

**二、立法目的**

1. **建立统一的法律框架**: 通过建立一个统一的法律框架，解决欧盟成员国间在人工智能领域法规的分散性和法律不确定性问题。这一框架的建立旨在促进人工智能技术在整个欧盟范围内的健康发展，确保内部市场的正常运作，并确保人工智能系统相关产品和服务在整个欧盟内的无缝流通。
2. **保护成员国间的基本权利**: 该法律框架保证了在欧盟各成员国之间基本权利的一致尊重和保护。这包括个人数据保护、消费者权利和隐私权等核心价值。
3. **促进AI技术的接受度**: 统一规则的确立使人们和其他用户对基于人工智能的解决方案有更多的信心。这种信任是推动人工智能解决方案广泛接受和应用的关键因素。
4. **激励企业创新与开发**: 统一的法律环境为企业提供了一个更稳定和可预见的法律环境，从而鼓励企业投入资源进行人工智能技术的研发和创新。这不仅有助于技术进步，也促进了新技术的商业化。
5. **增强内部市场的效率与竞争力**: 通过消除内部市场中的法律壁垒，统一法律框架有助于提高市场效率，增强欧盟内部市场的全球竞争力。这样的市场环境有利于吸引更多的投资，推动经济增长。

**三、监管框架**

**（一）横向监管为基础**

目前，在AI监管的相关法律文件中，主要存在“横向”监管与“纵向”监管两种主要方式。

* 在**横向监管**方式中，监管机构将创建一个全面的法规，以尽可能涵盖AI可能产生的各方面影响；
* 在**纵向监管**方式中，政策制定者采取“定制”的方法，针对不同应用或类型的AI制定不同的法规。

欧盟《AI法案》则采用横向监管模式，具体而言，该法案以风险分级的方式将所有AI系统纳入监管范围（特殊AI系统除外），并允许监管机构随着AI的发展不断将新的应用领域纳入现有的风险类别，而没有针对特定AI应用领域制定具体的法律规范。**风险分级**方式使法案整体处于相对灵活的状态，既能够保持横向监管方式具有的统一性和协调性，同时，相对灵活的分类标准也弥补了传统横向监管方式下对具体的AI应用场景针对性不高的问题，兼具法律的确定性与灵活性，使得相关监管措施更易落地。

**（二）欧盟《人工智能法案》基于风险分类的监管方式[[1]](#footnote-1)**

|  |  |  |  |
| --- | --- | --- | --- |
| **风险等级** | **定义** | **例子** | **要求** |
| **不可接受风险** | 对人们的基本权利构成明显威胁的人工智能系统 | 1. 使用敏感特征（如政治、宗教信仰、性取向、种族）的生物识别分类系统； 2. 无目的地从互联网或闭路电视录像中抓取面部图像创建面部识别数据库； 3. 在工作场所和教育机构中进行情绪识别； 4. 基于社会行为或个人特征进行社会评分； 5. 设计用来操纵人类行为或利用人们的弱点（如年龄、残疾、社会或经济状况）的人工智能系统。 | 禁止 |
| **高风险** | 对健康、安全、基本权利、 环境、民主和法治具有重大潜在危害的人工智能系统 | 关键基础设施、医疗设备、招募以及执法、边境管制、司法等过程中使用的AI系统 | 全面的强制性合规义务，包括强制性基本权利影响评估、严格要求、风险缓解系统、高质量数据集、活动记录、详细文档、清晰的用户信息、人工监督、部署前合格评估、记录保存义务和强制性影 响评估等 |
| **有限风险** | 使用者在应用系统时能够意识到在与人工智能互动，且使用者仍可以自主判断做出决定 | 包括聊天机器人、某些情绪识别和生物识别分类系统以及生成式深度伪造等 | 受到更低透明度义务的约束，包括告知用户他们正在与人工智能系统进行交互，并将合成音频、 视频、文本和图像内容标记为人工生成 |
| **最小/无风险** | 所有不属于三个风险类别的其他人工智能系统 | 支持 AI 的推荐系统或垃圾邮件过滤器等 | 允许免费使用风险最小的人工智能系统，同时鼓励自愿行为准则 |

不同风险级别的AI系统对应不同的监管措施，高风险意味着更严格的监管。这种灵活的分类方法覆盖了所有AI系统，并随技术发展持续更新。该方法针对不同应用场景设定了具体监管措施，避免了规制失焦并为复杂场景提供了清晰的监管框架。

注意：对于多AI耦合系统，当较高风险与较低风险的AI系统同时存在时，可以采用“就高不就低”的原则，对风险进行较严的监管规制。

**（三）兼顾监管与发展的动态平衡**

《人工智能法案》通过多种灵活措施和例外情况来促进AI技术的创新发展，同时确保满足监管要求。具体来说，法案在其三个版本中均提到了AI监管沙箱机制，此措施旨在为企业提供一个受控的“安全空间”，使其能在此环境中测试和验证创新性AI系统，促进AI技术的开发和实验。

此外，法案对知识产权的保护也给予了充分重视。例如：

1. 第79条指出，国家监督机构应将其获取的任何信息，包括源代码、软件和数据（如适用）作为**机密信息**处理，并遵守欧盟关于保护知识产权和商业秘密的相关法律。
2. 第83条规定，参与该条例应用的所有相关方应致力于确保活动的透明性和公开性，同时也需要保护在执行任务过程中获取的信息和数据的**保密性**，通过制定技术和组织措施来确保信息的安全性和保密性，包括知识产权以及公共及国家安全利益。

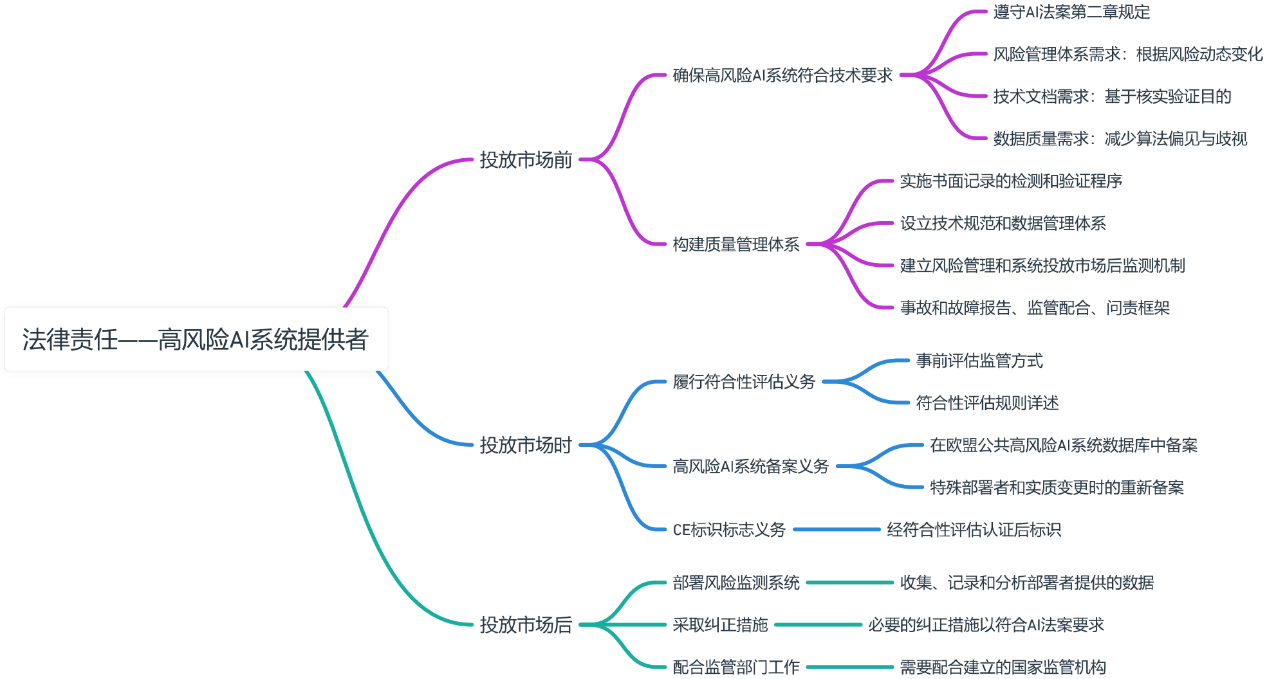
这种对知识产权的保护不仅是对创新活动的激励，也是推动科技发展和进步的关键因素。因此，《AI法案》在完善监管措施的同时，基于对个人基本权利的保障，努力实现AI技术的可信赖性与AI技术发展之间的动态平衡。

**（四）监管框架下的亮点制度**

|  |  |
| --- | --- |
| **亮点制度** | **内容概览及解读** |
| **禁止性应用方式** | 针对可能威胁公民权利的某些人工智能应用设定了禁令。这些禁令包括禁止基于敏感特征的生物特征分类系统等等（详见上表） |
| **执法豁免** | 禁止执法部门使用生物识别识别系统（RBI），除非在列出的严格定义的情况下才可使用。例如：   1. 在时间和地点上受到限制； 2. 需要获得事先的司法或行政授权； 3. 执法部门在事后使用此系统需与刑事犯罪相关联的司法授权。 |
| **高风险人工智能系统的安全义务** | 针对关键基础设施、教育和职业培训、就业、基本私人和公共服务、执法领域、移民和边境管理、司法和民主过程（如影响选举）的系统。  必须评估和降低风险、保持使用日志、确保透明度和准确性，并保证人类的监督。  公民对于可能影响其权利的人工智能系统的决策有权获得解释。 |
| **透明度要求** | 通用型人工智能（GPAI）系统及其基于的GPAI模型，需遵循特定的透明度要求。如人造或操纵生成的图像、音频或视频内容，如AI换脸技术或AI生成内容，必须进行明确的标记。 |
| **支持创新和中小企业的措施** | 在国家层面上建立沙箱[[2]](#footnote-2)和现实世界的测试环境，特别是针对中小企业在其产品进入市场之前，提供开发和训练创新AI技术的机会。 |
| **关于开源人工智能模型** | 关于禁止具有**不可接受风险**的AI系统以及限制**高风险**AI系统的条款，法案中并没有为开源AI系统设立特殊的规定。对于那些利用开源人工智能产品进行商业活动的实体，例如通过为开源模型提供付费技术支持或利用定向广告来支付成本，他们将不享有任何豁免权。法案鼓励开源开发者采用广泛接受的文档实践以提高透明度和可靠性。 |

通过这些亮点制度，展现了对技术发展的深思熟虑和对基本人权的坚持，试图在促进科技创新和保障人权之间找到一条可行的道路。这种新颖法律框架的设立体现了对技术发展潜在影响的深入理解及其对社会结构和个人生活的深远影响。通过设定明确的禁止和规范措施，法案不仅防范了人工智能可能滥用导致的风险，还激励了科技行业在遵守法律的前提下创新。

**四、法律责任——以高风险AI系统提供者为例**

****

**（一）投放市场前**

1. **确保高风险AI系统符合相关技术要求的义务：**根据《AI法案》第二章的规定，投放市场或投入使用的高风险AI系统应具备的要件，可进一步细化为基于风险动态变化形成的风险管理体系需求，基于核实验证目的形成的技术文档需求，以及基于减少数据训练引发的算法偏见与歧视等问题形成的数据质量需求。第三章则规定，在相关系统投入使用前，提供者需确保其提供的高风险AI系统已具有上述能力或履行相关要求。
2. **构建质量管理体系的义务：**《AI法案》2023年折衷草案第17条第（1）款规定，高风险AI系统提供者应建立、实施质量管理体系。具体而言，提供者应通过书面文件的形式（如政策、程序或指令等）系统有序地记录其检测和验证程序、适用的技术规范、数据管理体系、风险管理体系、系统投放市场后监测机制、事故和故障报告、监管配合、问责框架等。

**（二）投放市场时**

1. **履行符合性评估义务：**《AI法案》对高风险系统采取了事前评估的监管方式，规定在将高风险AI系统投放市场或投入使用前，提供者或第三方评估机构应当进行针对高风险AI系统的符合性评估。2023年《AI法案》折衷草案第43条详细规定了开展符合性评估的规则，其本质是为验证相关条款、对高风险AI系统的有关要求是否得到满足的过程。
2. **高风险人工智能系统备案义务：**根据《AI法案》提案第51条规定，高风险AI系统投放市场或投入使用前，应在欧盟委员会主导建立的欧盟公共高风险AI系统数据库中备案。折衷草案相较而言，除了新增了特殊部署者的备案义务，还进一步补充了在AI系统发生实质变更时的重新备案义务。
3. **CE标识标志义务：**对于前述符合性评估的认证，《AI法案》自2021年提案起，即引进了欧盟合格认证（Conformité Européenne, “CE”）机制。对存在高风险的AI产品和服务，经评估程序被认定合格后，可贴上CE标志投入使用。

**（三）投放市场后**

1. **部署风险监测系统的义务**：根据《AI法案》2023年折衷草案第61条规定，高风险AI系统提供者应采取与AI技术和高风险AI系统相称的方式，建立并执行上市后的监测系统。此监测系统应具有收集、记录和分析由部署者或其他来源提供的有关高风险AI系统在整个生命周期内的性能的相关数据。
2. **采取纠正措施等义务**：根据2023年《AI法案》折衷草案第21条和22条，如果高风险AI系统的提供者认为或有理由认为其投放市场或投入使用的高风险AI系统不符合《AI法案》的要求，应立即采取必要的纠正措施。
3. **配合监管部门开展工作的义务**：《AI法案》作为一项欧盟范围内的立法，将对每个欧盟成员国产生法律约束力。在各阶段的草案中，《AI法案》均提出各成员国须建立相应的国家监管机构，以监督《AI法案》的规则在国内的应用和实施。相应的，提供者也设有需要配合上述监管机构开展工作的义务。

**五、总结与展望**

本次专题讨论针对人工智能法律监管的现状与未来发展进行了全面梳理。随着人工智能技术的快速发展，全球主要经济体均已认识到加强AI法律和伦理规范的必要性。特别是欧盟的《人工智能法案》，它建立了一个全面的风险分类监管框架，旨在确保AI技术的健康发展，同时保护公民的基本权利与隐私。

**从全球视角来看**，AI法律立法的背景是技术的普及与风险的显现。例如，有关算法偏见的问题已经引发了广泛关注，如COMPAS系统在美国的使用就暴露了AI可能带来的种族歧视问题。因此，法律框架的构建不仅要解决技术应用中的即时问题，还要预防潜在风险，并在全球范围内促进法律与技术的协调发展。欧盟《人工智能法案》是对AI监管理念的一次重要实践，该法案通过设定不同风险等级来具体化监管措施，保障了从最高风险到最低风险的AI系统都在合理的监管之下。法案还强调了监管与发展的动态平衡，通过沙箱机制等措施支持技术创新，同时保护知识产权和促进开源模型的透明度和可靠性。

**从监管实践角度出发**，欧盟的法规模式可能为其他国家和地区提供了可借鉴的框架。统一的法律环境有助于降低成员国间的法律不确定性，推动AI技术在内部市场的自由流通，同时确保了对基本权利的尊重与保护。

展望未来，随着AI技术的进一步发展和应用扩展，立法与监管也需不断适应新的挑战和变化。全球合作在制定跨国AI监管标准中将扮演关键角色，以确保技术进步能够惠及所有国家，同时控制潜在的负面影响。在这一过程中，法律制定者、技术开发者和公众的共同参与和监督是实现健康发展的关键。通过这种多方面的合作与对话，我们可以期待一个更加安全、公正和创新的AI未来。

1. 《人工智能法案》的第5条 [↑](#footnote-ref-1)
2. 监管沙箱制度为英国金融行为监管局（Financial Conduct Authority）于2015年为监管FinTech（金融科技）及相关创新产品应运而生的一项监管措施，监管沙盒旨在“建立一个安全空间，企业可以在其中测试创新性的产品、服务、商业模式和提供机制，而不会因从事所述活动而立即招致通常的监管后果 [↑](#footnote-ref-2)