公益翻译小组荣誉出品

欧盟《人工智能白皮书》中译本

译校者：李芊晔、罗祥、袁俊、洪玮、甘丹、刘玲麟

2020年03月12日

译校者简介及工作分工

【组长】

李芊晔 车企IT合规经理

负责分工、组织翻译、全文校对及排版

【翻译】

李芊晔，车企IT合规经理（前言至第4章A节，原报告第1至6页）
罗祥，华侨大学法律硕士（第4章B节至第5章前言，原报告第7至10页）
袁俊，对外经贸大学法学研究生（第5章A、B节，原报告第11至15页）
洪玮，外资企业信息安全经理（亚太区）（第5章C节，原报告第16至18页）
甘丹，车企隐私合规（第5章D节，原报告第19至22页）
刘玲麟，资深法务经理（第5章E节至完，原报告第23至27页）

【校对】

所有翻译人员均参与第一轮校对
李芊晔负责第二轮全文校对

布鲁塞尔，2020年02月19日
COM(2020) 65 终版

白皮书

人工智能 ——追求卓越和信任的欧洲路径

人工智能白皮书：
追求卓越和信任的欧洲路径

人工智能发展迅速。它通过我们可以预见的不同方式改变着我们的生活，比如通过改善医疗保健（例如，更准确地诊断，更好地预防疾病），提高耕作效率，有助于缓解和适应气候变化，通过预测性维护提高生产系统的效率，提高欧洲人的安全等等。同时，人工智能也带来了一些潜在的风险，如决策不透明、性别歧视或其他形式的歧视、侵犯我们的私人生活或被用于犯罪目的。

在2018年4月提出的欧洲人工智能战略[1]的基础上，需要采取一种更切实的途径应对全球日趋激烈的竞争。为了应对人工智能的机遇和挑战，欧盟必须根据欧洲价值观一体行动，确定自己的方式，促进人工智能的发展和实施。

欧盟委员会致力于实现科学突破，以保持欧盟的技术领先地位，确保新技术为所有欧洲人服务——提高他们的生活质量，同时尊重他们的权利。

委员会主席Ursula von der Leyen在其政治指导方针[2]中提到了一种协调一致的欧洲方法——关于控制人工智能对人类和伦理意义，以及对更好地利用大数据进行创新的思考。

因此，委员会支持一种以监管和投资为导向的方法以实现双重目标，一方面促进人工智能的普及，另一方面解决伴随着这种新技术的使用而带来的风险。本白皮书旨在就如何实现这些目标提出政策选择，并不讨论人工智能在军事用途方面的发展和使用。委员会邀请成员国、其他欧洲机构和包含工业界、社会合作伙伴、民间社会组织、研究人员、公众和任何有关团体在内的所有利益相关者，对如下内容进行了反馈，并继续支持为委员会今后在这一领域的决策。

[1]欧洲的人工智能，COM/2018/237终版
[2]https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

1. 导言

随着数字技术成为人们生活中越来越核心的部分，人们应该能够信任它。值得信赖也是数字技术获得认可的一个先决条件。这对欧洲来说是一个机会，因为欧洲非常重视价值观和法规要求，且具备为从航空到能源、汽车及医疗设备行业建设安全可靠的先进产品和服务的能力。

欧洲当前和未来的可持续经济增长及社会福利越来越依赖数据创造的价值。人工智能是数据经济最重要的应用之一。如今，大多数数据都与消费者相关，并在集中的云基础设施上进行存储和处理。相比之下，未来更为丰富的数据中将有很大一部分来自工业、商业和公共领域，并存储在各种系统上，尤其是在网络边缘工作的计算设备上。这为在工业和“企业至企业间”数字化方面具备优势地位，但在消费者平台数字化上相对弱势的欧洲带来了新的机遇。

简而言之，人工智能是将数据、算法和计算能力结合起来的技术集合。因此，计算技术的进步和数据可用性的提高是当前人工智能热潮的关键驱动力。欧洲可以将其技术和工业优势与高质量的数字基础设施，以及基于其基本价值的监管框架结合起来，成为欧洲数据战略中提及的“数据经济及其应用创新的全球领导者”[3]。在此基础上，它可以发展一个人工智能生态系统，为整个欧洲社会和经济带来技术效益：

- 使公民享受新的福利，例如改善医疗保健、减少家用机械故障、更安全清洁的交通系统、更优质的公共服务；
- 促进商业发展，例如在欧洲的绝对优势行业（机械、交通、网络安全、农业、绿色及循环经济、医疗保健及时尚、旅游业等高附加值行业）；以及
- 为公共利益服务，例如降低服务（交通、教育、能源和废物管理）成本，提高

[3] COM(2020) 66 终版

产品可持续性[4]；在尊重公民权利和自由的情况下，为执法机关配备适当工具、合理确保公民安全[5]。

鉴于人工智能对我们社会的重大影响以及建立信任的必要性，欧洲人工智能必须立足于我们的价值观和基本权利，如人的尊严和隐私保护。

此外，人工智能系统的影响不仅应从个人的角度考虑，而且应从整个社会的角度进行考虑。人工智能系统的使用对实现可持续发展目标、支持民主进程和社会权利方面的影响举足轻重。随着最近欧洲绿色协议[6]的提出，欧洲在应对气候和环境的挑战方面处于领先位置。像人工智能这样的数字技术是实现绿色协议目标的关键因素。由于人工智能越来越重要，人工智能系统对环境的影响需要在其整个生命周期和完整的供应链中得到适当考虑，例如在算法培训和数据存储方面的资源使用。

为了达到足够的规模，避免单一市场的分裂，欧洲对人工智能采取共同的做法是必要的。引入国家干预可能危及法律确定性，削弱公民信任，并抑制一个充满活力的欧洲工业的出现。

本白皮书在充分尊重欧盟公民价值和权利的情况下，提出了使人工智能在欧洲得以安全可靠发展的政策选择。本白皮书的主要组成部分有：

- 规定了在欧洲、各成员国和地区层面实施一致措施的政策框架。在私营企业和公共部门通力合作下，该框架的目标是通过调动资源实现基于价值链的“卓越生态系统”，着手于研究和创新，通过建立适当的激励机制加速包含中小企业在内的人工智能解决方案的应用。

- 未来欧洲人工智能监管框架的关键要素为创造一个独特的“可信赖生态系统”。

[4]总体上看，人工智能和数字化是实现欧洲绿色交易宏伟目标的关键驱动力。然而，目前信息和通信技术（ICT）领域的环境消耗估计占全球排放总量的2%以上。本白皮书所附的欧洲数字化战略提出了数字化实现绿色转型的措施。

[5]人工智能工具可以为更好地保护欧盟公民免受犯罪和恐怖主义行为的威胁提供机会。例如，这些工具可以帮助识别网上恐怖主义的宣传内容，发现销售危险产品的可疑交易，查明危险的隐藏物或非法物质及产品，向紧急情况下的公民提供援助，并协助指导急救人员。

[6]COM(2019) 640终版

为此，它必须确保遵守欧盟的规则，包括保护基本权利和消费者权利的规则，特别是那些在欧盟运行的高风险人工智能系统[7]。建立一个可信赖的生态系统本身就是其政策目标，同时应该使公民有信心接受人工智能的应用，让公司和公共组织在法律上有把握利用人工智能进行创新。委员会强烈支持在“以人为中心的人工智能技术中构建信任的沟通函”[8]中提及的“以人为中心”的方法论，并将把人工智能高级专家组编写的道德准则试点阶段成果纳入工作范围。

本白皮书所附的《欧洲数据战略》旨在促使欧洲成为世界上最具吸引力、最安全、最充满活力的数据敏捷经济体——为欧洲提供数据以改善决策、改善公民的生活。该战略提出了实现这一目标所需的、包含调动私人及公共投资在内的若干政策措施。最后，本白皮书所附的委员会报告分析了人工智能、物联网和其他数字技术对安全和责任立法的影响。

2. 利用工业和专业市场优势

欧洲不仅可以作为人工智能的使用者，而且可以作为这项技术的创造者和生产者受益于AI的潜力。它拥有优秀的研究中心、创新型初创企业，并在机器人技术、竞争性制造业和服务业（从汽车到医疗保健、能源、金融服务和农业）领域处于世界领先地位。欧洲发展了强大的计算基础设施（如高性能计算机），这对人工智能的功能至关重要。欧洲还拥有大量的公共和工业数据，其潜力目前尚未得到充分利用。它在安全可靠的低功耗数字系统方面具有公认的行业优势，这对人工智能的进一步发展至关重要。

利用欧盟投资下一代技术和基础设施的能力，以及含数据素养在内的数字能力，将提高欧洲在数据经济关键扶持技术和基础设施方面的技术主权。基础设施应支撑欧

[7]尽管可能需要作出进一步安排，以防止和打击出于犯罪目的滥用人工智能，但这不在本白皮书讨论范围之内。
[8] COM(2019) 168

洲数据池的建立，以实现可信赖的人工智能，例如基于欧洲价值观和规则的人工智能。

欧洲应该利用自身的优势，扩大其在生态系统和价值链上的地位，从某些硬件制造部门到软件，再到整个服务业。这种情况已经在一定程度上出现。欧洲生产了全球超过四分之一的工业和专业服务机器人（例如，为精密农业、安全、卫生、物流），并在为公司和组织开发及使用软件应用程序（企业对企业的应用程序，如企业资源规划，设计和工程软件）、以及支持电子政务和“智能企业”的应用程序方面扮演重要角色。

欧洲在制造业应用人工智能方面处于领先地位。超过半数的顶级制造商在其制造运营中实施了至少一种人工智能[9]。

欧洲在研究方面处于优势地位的一个原因是，欧盟的资助方案已被证明有助于统筹行动、避免重复，并利用成员国的公共和私人投资。过去三年中，欧盟为人工智能研究和创新提供的资金已增至15亿欧元，与前一时期相比增长了70%。

然而，欧洲对研究和创新的投资扔进是世界其他地区公共和私人投资的一小部分。2016年，欧洲人工智能投资约32亿欧元，相比之下，北美人工智能投资约121亿欧元，亚洲人工智能投资约65亿欧元[10]。相应的，欧洲仍需大幅提高投资水平。事实将表明，与成员国共同制定的人工智能协作计划[11]是在欧洲建立更密切的人工智能合作、创造协同效应以最大程度优化人工智能价值链投资的良好起点。

3. 抓住未来机遇：下一波数据浪潮

尽管欧洲目前在消费者应用程序和在线平台上的地位较弱，这在数据访问方面造成了竞争劣势，但跨领域的数据价值和重复使用所带来的重大转变正在发生。世界上

[9]紧随其后的是日本（30%）以及美国（28%）。源自：凯捷咨询（2019）
[10]人工智能和自动化时代欧洲的十大要务，麦肯锡（2017）
[11] COM(2018)795

产生的数据量正在迅速增长，从2018年的33泽字节增长到2025年的175泽字节[12]。每一次新的数据浪潮都给欧洲带来了机遇，使其在数据敏捷经济中占据一席之地，并成为这一领域的世界领导者。此外，数据的存储和处理方式在未来五年内将发生巨大变化。如今，80%的数据处理和分析发生在云计算中的数据中心和集中计算设施中，20%发生在智能连接对象（如汽车、家用电器或制造机器人）以及靠近用户的计算设施中（“边缘计算”）。到2025年，这一比例将有显著变化[13]。

欧洲是低功耗电子产品的全球领导者，而低功耗电子产品是下一代人工智能专用处理器的关键。这个市场目前由非欧盟国家主导。在欧洲处理器计划（European Processor Initiative）和拟于2021年启动的关键数字技术联合项目（Key Digital Technology Joint engineering）等举措的帮助下，这种情况可能会有所改善。欧洲在神经形态解决方案[14]方面也处于领先地位，这种方案非常适合自动化工业过程（工业4.0）和运输模式。它们可以提高能源效率几个数量级。

量子计算的最新进展将使处理能力呈指数级增长[15]。得益于量子计算方面的学术优势和在量子模拟器及编程环境方面的强大地位，欧洲可以走在这项技术的前沿。欧洲旨在增加量子测试和实验设施可用性的倡议，将有助于将这些新的量子解决方案应用于一些工业和学术部门。

同时，欧洲将立足于自身的科学卓越性，在人工智能算法的基础上继续引领进步。需要在当前独立工作的学科之间建立桥梁，例如机器学习和深度学习（其特点为有限的可编译性，需要大量数据来训练模型并通过相关性进一步学习）和符号方法（其规则通过人工干预创建）。将符号推理与深层神经网络相结合可以帮助我们提高人工智能结果的可解释性。

[12]IDC (2019)
[13]盖特纳（2017）
[14]神经形态的解决方案是指大规模的集成电路系统，以模拟神经系统中神经生物结构的形式进行连接。
[15]量子计算机将有能力在不到秒的时间内，处理比当今性能最高的计算机大很多倍的数据集，从而允许跨领域开发新的人工智能应用。

4. 卓越的生态系统

为了建立一个卓越的生态系统，支持人工智能在整个欧盟经济和公共行政中的发展和吸收，需要在多个层面上加强行动。

A、 与会员国合作

根据2018年4月通过的人工智能战略[16]，欧盟委员会于2018年12月提出了一项与成员国共同制定的协同计划，以促进人工智能在欧洲的发展和使用[17]。

该计划提出了约70项联合行动项，以便成员国与委员会在诸如研究、投资、市场准入、技能和人才、数据和国际合作等关键领域开展更密切有效的合作。该计划预计将持续到2027年，并定期进行监控和审查。

其目的是最大限度地发挥对研究、创新和部署的投资的影响，评估国家人工智能战略，在此基础上建立并向成员国扩展其人工智能协作计划：

- *行动1：委员会将根据公众对白皮书内容的征询意见结果，向成员国提供一份修订后的协作计划，并期望其在2020年底前被采用。*

欧盟对人工智能领域的资金支持，应吸引并汇集一些领域投资，这是任何单一国家成员国无法实现的影响力目标是在未来十年内，每年吸引欧盟内超过200亿欧元的人工智能投资[18]。为了刺激私人和公共投资，欧盟将从“数字欧洲计划”、“地平线欧洲”以及“欧洲结构化及投资基金”中提供资源，以满足欠发达地区和农村地区的需求。

协作计划还将社会和环境福祉作为人工智能的一项关键原则。人工智能系统承诺帮助解决包括气候变化、环境退化等在内的、最紧迫的问题。同样重要的是，这以一

[16]欧洲的人工智能，COM(2018)237
[17]人工智能协作计划，COM(2018)795
[18]COM(2018)237

种环保的方式实现。人工智能能够而且应该严格审查自身的资源使用和能源消耗情况，并接受培训以便做出对环境有利的选择。委员会将考虑各种办法，来鼓励和促进与成员国一道解决人工智能问题。

B、 关注研究与创新界的努力

在当前分散的能力中心格局下，欧洲无法达到与全球领先机构竞争所需要的规模。尤为迫切的是，在多个欧洲人工智能研究中心之间形成更多协同和网络，共同努力以创造卓越、留住和吸引优秀的研究人员并开发最好的技术。欧洲需要一个集研究、创新和专业知识为一体的领先中心，吸引投资和该领域的最佳人才，成为人工智能领域的世界参考。

这些中心和网络应集中在欧洲有潜力成为全球引领者的领域，如工业、卫生、运输、金融、农产品价值链、能源/环境、林业、地球观测和宇航。在所有这些领域，争夺全球领导地位的竞争仍在继续，欧洲提供了巨大的潜力、知识和专长[19]。同样重要的是创建测试和实验站点，以支持新的人工智能应用的开发和后续部署。

- *行动2：委员会将助力卓越，促进结合了欧洲、国家和私人投资的测试中心，可能包括一项新的法律工具。欧盟委员会提出了一项雄心勃勃的专项拨款，以支持建立“数字欧洲”计划中所提及的世界领先测试中心的建立，并以“地平线欧洲”——委员会2021年至2027年多年财务框架的一部分——的研究和创新行动，作为辅助。*

C、 技能

人工智能的欧洲路径应立足于对技能的高度重视，以填补能力短缺。[20]欧盟委员会不久将提出加强技能议程，旨在确保欧洲所有人都能从欧盟经济的绿色数字化转型中受益。倡议还将包含对监管部门的支持，以提高其人工智能技能，更优质高效地实施相应规则。最新的数字教育行动计划将有助于更好地利用数据和基于人

[19]未来的欧洲国防基金和永久性结构合作（PESCO）也将为人工智能的研发提供机会。这些项目应与更广泛的欧盟专门针对人工智能的民间方案同步。
[20]https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced-profiles-eu

工智能的技术，如通过学习及预测分析改进教育和培训系统，使其适应数字时代。该计划还将在各级教育中提高对人工智能的意识，以帮助公民做好应对人工智能日益增加的影响的准备。

构建在人工智能领域工作所需的技能，提升劳动力，使之适应人工智能引领的转型，将是与成员国共同制定的、更新版人工智能协作计划中的一个优先事项。这可能包含将道德准则中的评估清单转变为培训机构资源，向人工智能开发者提供的指示性“课程”。应尤其致力于提高在这一领域接受培训、就业的妇女人数。

此外，欧洲人工智能研究及创新中心因其丰富的可能性，将吸引世界各地的人才，并在欧洲孕育、传播其在此领域的卓越技能。

- *行动3：通过建立和支持由顶尖大学和高等教育机构组成的欧洲数字计划网络的高新技术支柱，以吸引最好的教授和科学家，并在人工智能领域提供世界领先的硕士课程。*

除了提升技能外，工人和雇主还直接受到工作场所中人工智能系统的设计和使用的影响。社会伙伴的参与将是确保人工智能“以人为中心”进行工作的关键因素。

D、 关注中小企业

确保中小企业可以访问和使用人工智能也很重要。为此，应进一步加强数字创新中心[21]和按需人工智能平台[22]并促进中小企业之间的合作。“数字欧洲计划”（theDigital Europe Programme）将有助于实现这一目标。

在所有数字创新中心都应为中小企业提供支持，以帮助他们理解和应用人工智能的同时，重要的是每个成员国至少有一个在人工智能方面具有高度专业化的创新中心。

中小企业和初创企业将需要获得融资，以便调整其流程或使用人工智能进行创新。在即将到来的针对人工智能和区块链的1亿欧元试点投资基金的基础上，欧盟委员

[21] ec.europe.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities.
[22] www.Ai4eu.eu.

会计划根据欧洲投资计划（InvestEU）[23]进一步扩大人工智能的融资渠道。在欧洲投资计划确定的、可使用其资金的领域中明确提到了人工智能。

- *行动4：委员会将与成员国合作，以确保每个成员国至少有一个在人工智能方面具有高度专业化的数字创新中心。“数字欧洲计划”为数字创新中心提供支持。*

- *欧洲委员会和欧洲投资基金将在2020年第一季度启动1亿欧元的试点计划，为人工智能的创新发展提供股权融资。在就MFF达成最终协议之前，欧盟委员会拟从2021年起通过欧洲投资计划大幅扩大该规模。*

E、 与私营部门合伙

确保私营部门充分参与制定研究及创新议程，并向其提供必要的共同投资水平也很重要。这要求建立广泛的公私合作伙伴关系，并确保得到企业最高管理层的承诺。

- *行动5：在“欧洲地平线”（Horizon Europe）的背景下，委员会将在人工智能、数据和机器人技术方面建立新的公私合作伙伴关系，形成合力，确保在人工智能研究与创新的协作，并与“欧洲地平线”项目中的其他公私伙伴关系通力合作，与上述测试机构和数字创新中心共同工作。*

F、 推动公共领域应用人工智能

关键的是，公共管理部门、医院、公用事业和运输服务、金融监管机构以及其他公共利益领域迅速开始部署人工智能相关的产品和服务。应当尤其关注技术成熟且可大规模部署的医疗保健和运输领域。

- *行动6：委员会将发起公开透明的部门对话，优先考虑医疗保健、农村行政管理和公共服务运营商，以制定促进人工智能发展、试验和应用的行动计划。行业对话将用于准备特定的“应用人工智能计划”（Adopt AI*

[23] Europe.eu/investeu.

programme），以支持人工智能系统的公共采购，转变公共采购流程。

G、 安全访问数据和计算基础设施

本白皮书中提出的行动领域是对根据欧洲数据战略并行提出的计划的补充。改善数据访问和管理是根本。没有数据，人工智能和其他电子应用的发展无异于纸上谈兵。尚未生成的海量新数据为欧洲提供了立足于数据和人工智能转型前沿的机会。推进相应数据管理实践并促成数据的FAIR原则符合性，将有助于建立信任并确保数据的可重用性[24]。同样重要的是对关键计算技术和基础设施的投资。

欧洲委员会已根据“数字欧洲计划”提议投资40多亿欧元以支持高性能和量子计算，包括边缘计算和人工智能、数据和云基础设施。欧洲数据战略将优先实现这些领域的进一步发展。

H、 国际方面

欧洲完全有能力发挥其全球领导作用，围绕共同价值观建立联盟并促进人工智能应用合乎道德。欧盟在人工智能方面的工作已对国际讨论产生影响。在制定道德准则时，高级专家组中引入了一系列的非欧盟组织和一些政府观察员。与此同时，欧盟密切参与制定经合组织关于人工智能的道德原则[25]。20国集团（G20）随后在2019年6月关于贸易和数字经济的部长级声明（Ministerial Statement on Trade and Digital Economy）中批准了这些原则。

同时，欧盟注意到其他多边论坛也正在开展有关人工智能的重要工作，包括欧洲委员会（Council of Europe）、联合国教育科学与文化组织（UNESCO）、经济合作与发展组织（OECD）、世界贸易组织（WTO）和国际电信联盟（ITU）。在联合国，欧盟参与了数字合作高级别小组报告的后续行动，包含其对人工智能的推荐。

欧盟将以符合欧盟规则和价值观的方式（例如，促进更高的监管趋同、访问包括数

[24]如委员会专家组于2018年就FAIR数据在最终报告中所述，FAIR意为可查找、可访问、可互操作和可重用，https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf。
[25] https://www.oecd.org/going-digital/ai/principles/

据在内的关键资源、创造公平的竞争环境），继续与志同道合的国家及人工智能领域的全球参与者展开合作。委员会将密切监测限制数据流量的第三国政策，并通过世界贸易组织场景下的行动解决双边贸易谈判中的不当限制。委员会坚信，在人工智能事务上的国际合作方式必须基于促进尊重基本权利，包括人的尊严、多元化、包容、非歧视、保护隐私及个人数据[26]，委员会将致力于向全世界推广其价值观[27]。显然，负责任地开发和使用人工智能可成为实现可持续发展目标和推进2030年议程的推动力。

5. 可信任的生态系统：人工智能的监管框架

与任何新技术一样，人工智能的使用既带来机遇也带来风险。公民担心因算法决策信息不对称而无力捍卫自己的权利和安全，企业担心其法律上的不确定性。虽然人工智能可以帮助保护公民安全并使其享有基本权利，但公民也担心人工智能后果无法预料，甚至可能用于恶意目的。这些问题需要被解决。此外，除了缺乏投资和技能外，缺乏信任也是阻碍人工智能广泛应用的一个主要因素。

这就是委员会在2018年4月25日制定人工智能战略[28]来解决社会经济问题的原因，同时欧盟范围内也在提高对研究、创新和人工智能能力的投资。委员会与成员国制定了协作计划[29]来调整战略。委员会还成立了一个高级专家组，于2019年4月发布了有关可信赖人工智能的准则。[30]

委员会发表了一份沟通函[31]，充分肯定了高级专家组准则中确定的七项关键要求：

- 人事代理机构和监管；

[26]基于合作伙伴文书的记载，委员会将资助一个250万欧元的项目，支持与之志同道合的伙伴的合作，以促进欧盟人工智能道德准则，采取共同原则及操作结论。
[27]主席 Von der Leyen，驱动更多的联盟——我的欧洲日程，第17页。
[28] COM(2018) 237.
[29] COM(2018) 795.
[30] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
[31] COM(2019) 168.

- 技术鲁棒性和安全性；
- 隐私和数据治理；
- 透明性；
- 多样性、非歧视性公平；
- 社会和环境福祉；
- 问责制。

此外，本准则还包含了一份供企业实际使用的评估清单。在2019年下半年，超过350个组织对此评估清单进行了测试并发回反馈。高级专家组正在根据反馈意见修订其准则，并将于2020年6月之前完成这项工作。反馈意见过程的主要成果是，尽管部分要求已反映在现行法律或法规中，但在许多经济领域的现行立法中并未明确涵盖与透明性、可追溯性和人为监管有关的制度。

除了高级专家组定义的该组非约束性准则之外，根据总统的政治准则，明确的欧洲监管框架将在消费者和企业之间建立对人工智能的信任，从而加快对技术的应用。该监管框架应该与其他行动相一致，以提高欧洲在该领域的创新能力和竞争力。此外，必须确保在社会、环境和经济上达到最佳结果，并遵守欧盟法律、原则和价值观。这在公民权利可能受到最直接影响的领域中尤其重要，例如执法和司法领域中的人工智能应用。

人工智能的开发者和部署者已受到欧洲有关基本权利的法律（例如数据保护、隐私、非歧视）、消费者保护以及产品安全和责任规则的约束。不管是产品或系统，消费者都期望依赖人工智能达到相同水平的安全性以及权利尊重。但是，人工智能的某些特定功能（例如不透明性）会使该法规的应用和执行更加困难。因此，有必要检查当前的立法是否能够解决人工智能的风险并能够有效执行，是否需要对立法进行修改，或者是否需要新的立法。

在人工智能快速演变的情况下，监管框架必须给将来的发展留有余地。任何更改都应限于明确已识别的问题且具有可行性解决方案。

成员国指出目前缺乏共同的欧洲框架。德国数据道德委员会（The German Data Ethics Commission）呼吁建立基于风险的五级监管体系，从对最无害的人工智能系统的不设监管，到对最危险的人工智能系统的完全禁止。丹麦刚落地了数据道德保证的模型。马耳他也为人工智能引入了自愿认证体系。如果欧盟未能提供整个欧盟范围内的解决途径，这将引致确定的内部市场分裂风险，破坏信任、法律确定性和市场占有率的目标。

可靠的欧洲人工智能监管框架将保护所有欧洲公民、帮助创建无摩擦的内部市场，以进一步发展和推广人工智能、加强欧洲人工智能的产业基础。

A、 问题定义

虽然人工智能可以做很多好事，包括使产品和处理更安全，但它也可能造成伤害。这个危害可能是物质的（如个人安全和健康，包括失去生命，财产损害）和非物质的（如隐私泄露，言论自由限制，人格尊严损害，就业歧视），也可能涉及到多种风险。监管框架应集中于如何最大程度减少潜在危害的各种风险，尤其是特别重大的风险。

与使用人工智能有关的主要风险涉及用来保护基本权利（包括个人数据、隐私保护和非歧视）的规则的应用、安全[32]和与责任相关的问题。

基本权利面临的风险，包括个人数据、隐私保护以及非歧视

使用人工智能会影响为欧盟奠定基石的基本价值观，并侵犯诸多基本权利[33]，包括言论自由、集会自由、人格尊严，在特定领域不因性别、种族或民族、宗教或信仰、残疾、年龄或性取向而受到歧视的权利，个人数据和私人生活受保护的权利[34]，有

[32]这包括网络安全问题，与关键基础架构中的人工智能应用相关的问题或对人工智能的恶意使用。
[33] 欧洲委员会研究表明，人工智能的使用可能会影响大量的基本权利。https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.
[34]“一般数据保护条例”和“电子隐私指令”（新的电子隐私条例正在磋商中）在处理这些风险，但可能需要研究人工智能系统是否构成额外风险。委员会将持续监控和评估“一般数据保护条例”的应用情况。

效的司法救济和公平审判，以及消费者受保护的权利。这些风险可能源于人工智能系统设计过程中就存在的缺陷（包括人为监督），或者来源于在不纠正可能存在的偏见的前提下使用数据（例如，该系统仅使用或主要来自男性的数据进行训练，导致生成的与女性相关的结果不太理想）。

人工智能可以执行很多以前只能由人类完成的功能。因此，公民和法律实体将越来越多地受制于人工智能系统做出的或在其协助下做出的行动和决定，这些行动和决定有时可能难以理解，甚至在必要时难以有效挑战。此外，人工智能增加了跟踪和分析人们日常习惯的可能性。例如，存在一种潜在风险，即在违反欧盟数据保护和其他规则的情况下，人工智能可能会被政府当局或其他实体用于大规模监控，或被雇主用来观察员工的举止是否得体。通过分析大量数据并确定它们之间的联系，人工智能还可能被用于回溯和去匿名化个人数据。即使数据集本身不包括个人数据，也会产生新的个人数据保护风险。人工智能也被在线中介用于为用户排列被展示内容的优先顺序，并执行内容审核。经过处理的数据、应用程序的设计方式和人类干预的范围都会影响自由表达、个人数据保护、隐私和政治自由的权利。

某些人工智能算法在被用于预测刑事犯罪累犯的概率时，可能会引起性别和种族偏见，显示出女性与男性、本国民众与外国人的不同累犯预测概率。

资料来源：Tolan S.，Mron M.，Gomez E.和Castillo C.．“为什么机器学习可能导致不公平：来自加泰罗尼亚青少年司法风险评估的证据”，国际最佳论文奖，2019年人工智能与法律会议。

某些用于面部分析的人工智能应用显示出对性别和种族偏差的分析结果，在确定浅色较浅的男性时误差低，而在确定深色较深的女性时误差高。

资料来源：Joy Buolamwini，Timnit Gebru；第一次公平、问题和透明会议论文集，PMLR 81：77-91，2018年。

偏见和歧视是任何社会或经济活动的固有风险。人类决策也不能避免产生错误和偏差。然而，同样的偏差在人工智能中可能会产生更大影响，在规制人类行为的社会

控制机制尚未建立之际影响和歧视许多人。[35]当人工智能系统在其运行中“学习”时，也可能发生这种情况。在此种设计阶段无法预防或预测相应后果的情况下，风险不会来源于系统原始设计中的缺陷，而是产生于系统进行大量数据集训练时，其所识别出来的关联性或模式中带来的实际影响。

包含不透明性（黑盒效应）、复杂性、不可预测性和部分自主行为等在内的诸多人工智能技术的特点，使得人们难以验证其是否符合现行欧盟法律的规定，从而妨碍了这些规定在保护基本权利方面的有效执行。执法当局和受影响人士可能没有途经核实在人工智能的参与下决定是如何做出的，继而无法核实有关规则是否得到遵守。在受到这些决定可能带来的负面影响的情况下，个人和法律实体在有效诉诸司法方面可能会面临困难。

安全风险和责任制度有效运作的风险

当人工智能技术嵌入到产品和服务中时，可能会给用户带来新的安全风险。例如，由于物体识别技术的缺陷，自动驾驶汽车可能会错误识别道路上的物体，对人类造成人身伤害和物质损失。与基本权利面临的风险一样，这些风险可能由人工智能设计缺陷造成，亦或与数据可用性和数据质量有关，也可能与机器学习引发的其他问题有关。虽然其中一些风险并不局限于依赖人工智能的产品和服务，但人工智能的使用可能会增加或加剧风险。

除了个人面临相关风险外，缺乏应对此种风险的明确法律规定，可能会给在欧销售人工智能产品的企业带来法律不确定性困扰。市场监督和执法机构恐陷入一种境地，即由于没有明确授权采取行动，和（或）不具备适当的技术能力检验系统，从而无法对此进行干预。[36]法律不确定性可能会导致整体安全水平降低，并削弱欧洲公司

[35]欧盟男女平等咨询委员会正在编写一份人工智能意见书，分析人工智能对两性平等的影响，委员会预计将于2020年初通过该意见。欧盟男女平等战略（2020-2024）还处理了人工智能与男女平等间关系的问题，欧洲平等机构网络（Equinet）将发表一份报告（由罗宾·艾伦和迪·马斯特斯撰写），题为“监管人工智能：平等机构的新角色-迎接数字化和人工智能使用增加带来的平等和非歧视的新挑战”，预计将于2020年初发布。

[36]儿童智能手表就是一个例子。该产品可能不会对使用它的儿童造成直接伤害，但是由于缺乏最低级别的安全

的竞争力。

如果安全风险成为现实，由于缺乏明确要求和前述提及的人工智能技术特点，使得很难追认人工智能系统参与下做出的决定是具有潜在隐患的。相应地，这又可能使遭受损害的人很难依据现行欧盟和国家的责任立法获得赔偿。[37]

根据“产品责任指令”（Product Liability Directive），制造商应对缺陷产品造成的损害负责。

然而，在基于人工智能系统的产品，例如自动驾驶汽车的场景下，可能很难证明产品存在缺陷、已发生的损害以及两者之间的因果关系。此外，如何以及在多大程度上将“产品责任指令”适用到不同类型的缺陷也存在一定的不确定性，例如，如果这些缺陷是由于产品的网络安全漏洞造成的。

因此，追认人工智能系统做出的决定有潜在问题是存在难度的，且上述与基本权利的风险同样导致安全和责任相关的问题。遭受损害的个人可能无法有效地获得法庭立案所需的证据，且与传统技术造成损害的情况相比，可能无法那么有效地追偿。随着人工智能的使用越来越广泛，这些风险将会增加。

B、 有关人工智能的欧盟现行立法框架可做出的调整

大部分欧盟现行产品安全和责任立法[38]，包括由国家立法进一步完善具体行业规则，是与人工智能相关、可能适用于一些新兴的人工智能应用。

关于保护基本权利和消费者权利，欧盟有相应的立法框架包括种族平等指令[39]（the Race Equality Directive）、雇佣及职业平等待遇指令[40]（the Directive on equal treatment in employment and occupation）、关于在就业、获得商品和服

措施，它很容易被用作接触儿童的工具。市场监管机构可能会发现，在风险与产品本身没有关联的情况下，很难进行干预。

[37]白皮书附录了委员会关于人工智能、物联网和其他数字技术对安全和责任立法的影响的报告。

[38]欧盟产品安全法律框架包括作为安全网的“一般产品安全指令”（指令2001/95/EC），以及涵盖不同类别产品的不同行业规则，从机器、飞机、汽车到旨在提供高水平的健康和安全的玩具、医疗器械。产品责任法与不同的产品或服务损害产生的民事责任制度共同形成完整的产品责任体系。

[39]指令2000/43/EC。

[40]指令2000/78/EC。

务方面男女平等待遇的指令[41]、大量的消费者保护规则[42]以及个人数据保护和隐私规则，最为著名的当属《一般数据保护条例》，以及涵盖了个人数据保护要求的其他行业立法[43]，例如数据保护法律执法指令（the Data Protection Law Enforcement Directive）。而且自2025年伊始，《欧洲无障碍法案》中关于商品和服务的无障碍要求将会适用。[44] 此外，当执行其他领域的欧盟法规时，包括在金融服务、移居或者中介在线责任，都需要尊重基本权利。

尽管无论人工智能的参与度如何，欧盟立法原则上仍然适用，但重要之处在于需评估法规能够得到切实执行以应对人工智能系统产生的风险，或者是否需要调整特定的法律文本。

例如，经济运营者仍然要对人工智能是否遵守现行保护消费者的规则负全部责任，任何违背现行关于算法利用消费者轨迹的规则都是不被允许的，违反的话应受到相应惩罚。

委员会认为可以改进立法框架，以处理下列风险和情形：

- *有效应用和执行现行欧盟和国家立法：*人工智能的关键特性为确保现行欧盟和国家立法的合理应用和执行带来了挑战。人工智能缺乏透明度（不透明性）使得识别和证明可能的违法行为变得困难，包括保护基本权利、责任归属和满足索赔条件的法律条款。因此，为了确保有效的实施和执行，可能有必要对特定领域的现行立法进行调整或说明，比如在本白皮书随附报告中进一步详细阐释的责任问题。

- *欧盟现行立法的范围限制：*欧盟产品安全立法的一个重要关注点是产品的市场定位。在欧盟产品安全立法中，当软件是最终产品的组成部分时，必须遵守相关的产品安全规则，独立软件是否受欧盟产品安全立法管理是一

41指令2004/113/EC；指令2006/54/EC。
42 例如“不公平商业行为指令”（指令2005/29/EC）和“消费者权利指令”（指令2011/83/EC）。
43欧洲议会和欧洲理事会2016年4月27日关于自然人保护的指令（指令(EU)2016/680），提及关于主管当局为预防、调查、侦查或起诉刑事犯罪或执行刑事处罚而处理个人数据的，以及此类数据的自由流动。
44欧盟产品和服务无障碍条例（指令(EU) 2019/882）。

个公开的问题，虽然在一些行业有明确的规定。[45]欧盟现行的一般安全法规适用于产品而非服务，因此原则上也不适用于基于人工智能技术的服务（例如卫生服务、金融服务、交通服务）。

- *改变人工智能系统的功能：*将包括人工智能技术在内的软件集成到产品中，可以在产品和系统的生命周期中更新其功能。这对于需要频繁升级软件或依赖机器学习的系统更是如此。这些特点会产生新的风险，这些风险在系统上市时尚不存在。而现行立法主要关注产品上市时是否存在安全风险，并没有充分提及、解决这些新生风险的问题。

- *供应链中不同经济运营者之间责任分配的不确定性：*通常，欧盟产品安全立法将责任分配给市场上产品及其所有组件的生产者，例如人工智能系统。但如果人工智能技术是由其生产者之外的一方在产品上市后增加进产品中的，那这些规定可能会变得不甚明确。此外，欧盟产品责任立法规定了生产者的责任，却将供应链中其他主体的责任交由国家管理。

- *安全概念的变化：*在产品和服务中使用人工智能技术可能会产生欧盟立法目前没有明确提及的风险。这些风险可能与网络威胁、个人安全风险（涉及人工智能新应用，例如家用智能电器）、连接中断导致的风险等有关。这些风险可能在产品上市时出现，或者在使用产品时因软件更新或进行自我学习（self-learning）而出现。欧盟应充分利用其所掌握的工具，加强其对与人工智能应用相关潜在风险的证据基础，包括利用欧洲网络与信息安全局（ENISA）的经验评估人工智能潜在威胁。

[45]例如根据“医疗器械条例”（条例(EU)2017/745），制造商打算用于医疗目的的软件被视为医疗器械。

如之前指出的，一些成员国早已开始探索其国内立法的可行性来应对人工智能带来的挑战。这也带来了单一市场分散的风险。不同国家的法规很可能会限制那些想到单一市场销售和运营人工智能系统的公司。从欧盟层面统一使用一个共同的方法，将有助于欧洲公司顺利进入单一市场，并支持他们在全球市场中增加竞争力。

有关人工智能，物联网和机器人安全和责任影响方面的报告

随附在白皮书的报告分析了相关的法律框架，指出了将此框架应用于解决人工智能系统和其他数字科技方面所带来的特定风险的不确定性。

它总结得出目前的产品安全立法已经提供了扩展的安全保护的概念，以应对来自产品使用的风险。同时，那些明确涵盖新兴数字科技所带来的新风险的条款，也能被引用，以提供更多的法律确定性。

- 一些人工智能系统在其生命周期内的自主行为可能导致对安全有影响的重要产品变更，并可能需要进行新的风险评估。此外，也需要人类对人工智能产品及系统从设计、并贯穿其生命周期实施监管，保障安全。
- 适当情况下，生产者对于用户精神安全风险的明确责任也应被考虑进去。
- 欧盟产品安全立法能够规定在设计阶段解决使用错误数据导致的安全风险的具体要求，以及确保人工智能产品和系统使用过程中数据质量的机制。
- 有关算法系统的不透明性可以通过要求透明度来解决。
- 一个单独的软件投放到市场，或在上市后被下载到一个产品中，此情形下，如果其影响安全，现行的规则可能需要被调整和说明。
- 随着新技术使用导致供应链变得日益复杂，特别要求供应链上的经济运营者和用户之间能够协作的法律条款例可以提供法律确定性。

像人工智能、物联网和机器人之类的新兴数字科技，其特性可能会挑战责任框架的某些方面，并降低其有效性。由于其中一些特性很难把过失归咎到个人身上，而这在大多数国家中基于过失的索赔是必须的。这就导致了受害者的成本大大增加，也意味着除了生产者以外的责任追究更难提出和证明。

- 人类因人工智能系统使用而遭受的损害也应该和被其他技术造成的损害一样受到同等保护，同时也应允许技术革新继续发展。
- 所有确保这一目的的选择都应该被谨慎评估，包括可能修订产品责任指示和可能进一步针对国家责任规则的趋同化。譬如，委员会正在征求意见，是否以及应在多大程度上调整国家责任规则要求的、由人工智能应用所造成的损失举证责任，来减轻举证复杂性。

根据以上讨论，委员会得出结论，除了对于现行立法可能进行的调整外，还可能需要新增一个针对人工智能的立法来使得欧盟法律框架适用于当前及可预见的科技和商业的发展。

C、 未来欧盟监管框架的范围

对于将来的人工智能监管框架而言，关键之处在于确定它的应用范围。可行的设想是它将适用于依赖人工智能的产品和服务。人工智能因此应在本白皮书的目的及未来可能的政策制定参考中被清晰的定义。

在对欧洲有关人工智能的沟通函中，委员会首次定义了人工智能[46]。高级专家组[47]对这个定义又进行了完善。

在任何新的法律文书中，人工智能的定义应在足够精准、满足法律确定性需要的同时，还要足够灵活以适应科技发展。

> 以自动驾驶为例，其算法使用来自于车辆产生的即时数据（如车速、引擎消耗、减震器等）和感应器扫描到的车辆所处环境的数据（如路况、信号灯、其他车辆、行人等）来决定车辆到达指定目的地应采取的驾驶方向、加速度和速度。基于观察到的数据，算法能够适应路况和外部条件，包括基于其他驾驶员的行为来得出最舒服、安全的驾驶。

对于本白皮书的目的，和将来任何可能的政策提议讨论，明确“数据”和“算法”是组成人工智能的主要元素尤为重要。人工智能可以被集成到硬件中。被认为是人工智能子集的机器学习技术中，算法被训练成基于数据集来推导特定模型，从而确定完成既定目标所需的动作。使用中，算法会继续学习。虽然基于人工智能的产品可以通过感知其周边环境而不需通过预先设置的指令而行动，但它们的在行为很大程度上还是被开发者定义和约束的。人类决定并规划目标，而人工智能

[46] COM(2018) 237 最终版，第1页：“人工智能指的是通过分析周边环境后采取行动——一定程度的自动化——来达到特定的目的，从而展示智能行为的系统。
基于人工智能的系统可以是纯软件基础，在虚拟世界中展示（如，语音提示，图像分析软件，搜索引擎，语音和人脸识别系统）或是集成在硬件设备中的人工智能（如，高级机器人，自动驾驶汽车，无人机或是物联网的应用等）。

[47]高级专家组在人工智能定义一书的第8页中提及：“人工智能系统是人类设计的软件（也可能是硬件）系统，在给定一个复杂目标的情况下，通过获取数据而感知周围的环境，解读收集到的结构或者非结构化的数据，推理知识，或者处理从数据中得出的信息，从而在在物理或者数字维度行动，并确认最佳的行动来达到特定目的 。 人工智能系统既可以使用象征性的规则，也可以学习数字化模型，同时它们也可以通过分析环境如何受他们以前的行为的影响来对自身行为进行调整。

系统来对其进行优化。

欧盟有严格的法律框架来确保对消费者的特别保护，以消除不公平的商业行为并保护个人数据和隐私。此外，这些法律还包括了对于特定行业（如：健康，运输）的特别规则。这些现行的欧盟法律条款仍将继续适用于人工智能，尽管可能有必要对这些框架进行更新以反映数字化转型和人工智能的使用（参见章节B）。因此，此框架仍将继续监管那些目前已经被现行横向或部门内立法涉及的方面（如：医疗设备[48]，运输系统）。

原则上，针对人工智能的新的监管框架应该更有效地达到其监管目的，而不应增加过度的指令造成不合理的负担，尤其是对中小企业。为了实现这样的平衡，委员会认为应该采取基于风险的方法。

基于风险的方法对协助确认监管干预程度是否合理恰当非常重要。然而，它需要清晰的标准来区分不同的人工智能应用，尤其是那些关系到他们是否属于“高风险”[49]的问题。对高风险人工智能应用的判定应是清晰、易理解且对相关各方适用的。然而，尽管有的人工智能应用不被认为有高风险，它仍会受到现行欧盟法则的约束。

委员会认为应将特定的人工智能应用视为是高风险的，同时考虑其所应用的行业和预期用途是否牵涉重大风险，尤其应从安全保护、消费者权利和基本权利的角度出发。更确切地说，如果一个人工智能应用同时满足以下两个的判断标准，那它就应该被认为认为是高风险的：

- 首先，结合某行业的典型活动的特点，人工智能应用在该行业中使用时极有可能引发重大风险。这个判断标准为了确保监管的介入是针对那些大概率会发生风险的特定领域。被涵盖的行业应该在新的监管框架中被具体、详尽地列出。例如，健康、运输、能源和公共行业的多个部分。[50]此清单

[48]举例来说，对于人工智能系统向医师提供的特别的医疗信息，人工智能系统直接提供给患者的医疗信息和人工智能系统自行执行的对患者的医疗任务，有很多不同的安全考量和法律提示。委员会在逐一检查这些不同于健康的安全性和责任问题。
[49]欧盟立法会对在这里提到的“风险”根据不同方面进行分类，譬如，产品安全。
[50]公共行业包括救济所、移民局、边境防控和法院、社会安全以及就业服务等。

应被定期审阅，并根据实践的相关发展进行及时修正。

- 其次，在有争议的行业使用人工智能应用有可能会增加重大风险。这个标准反映出并非在特定行业中使用人工智能应用就都隐含了高风险。比如，健康行业可以被认为是存在这种情况的行业，但医院预约系统的缺陷并不会造成需要法律介入的重大风险。对人工智能应用在既定使用场景下的风险等级的评估应基于相关各方所受到影响的程度。例如，因人工智能应用的使用，给个人或公司造成了法律上的或类似的重大影响；造成伤亡或重大物质或非物质损失的风险；对个人或法律实体造成不可避免的影响。

这两个累计标准的使用能够确保监管框架范围的针对性，并提供了法律的确定性。原则上，新的人工智能监管框架所包含的强制性要求应仅适用于符合上述两条累计标准定义的高风险应用。

尽管有上述规定，但也可能出现一些例外情况，例如由于风险的存在，为某些特定目的使用人工智能应用仍会被视为是高风险的，且适用[51]如下规定，而与其所应用于的行业是否受关注无关。作为说明，可以特别考虑以下内容：

- 鉴于其对个人及解决欧洲平等就业问题的重要性，在招聘过程及影响劳工权益的情形下时下使用人工智能应用应始终被认为是“高风险”的。因此，下文中的要求应将始终适用。还可考虑影响消费者权益的其他具体应用场景。

- 出于远程生物识别[52]（remote biometric identification）及其它侵入式监控技术的目的而使用人工智能应用，应始终被认为是“高风险”的。因此，下文中的要求应将始终适用。

[51]尤其应该强调的是，其他欧盟法律在此处仍可能适用。例如，当包括消费品时，《一般商品安全指令》能适用于人工智能应用的安全性。

[52]远程生物识别应与生物特征认证相区别（后者是基于单一个体独特的生物特征的一种安全程序，用以保证某人即其所称）。远程生物识别是在生物识别标识（指纹、面部图像、虹膜、血管性状等）的帮助下，通过远程、公共空间和连续或持续的方式，将多重主体身份与数据库中存储的数据进行核验。

D、 要求的类型

在设计未来的人工智能监管框架时，有必要确定针对相关参与者的强制性法律要求的类型。这些要求可通过标准进一步明确。如上文章节C所述，除现行立法外，这些要求将仅适用于高风险的人工智能应用，从而确保所有监管干预措施的针对性和适当性。

考虑到高级专家组的指导方针以及上文所述，对高风险人工智能应用的要求可能包含如下几项关键特点，后续各小节将分点进行详细讨论：

- 训练数据；
- 数据及记录保存；
- 所提供的信息；
- 鲁棒性及准确性；
- 人类监督；
- 为某些特定人工智能应用所设立的具体要求，如那些以远程生物识别为目的人工智能应用。

为确保法律确定性，这些要求将被进一步详细说明，为所有需照此执行的参与者提供一个明确的基准。

a） 训练数据

促进、加强和捍卫欧盟价值观及规则，尤其是公民从欧盟法律中获得的权利，这比以往任何时候都更重要。这些努力也将毫无疑问地延伸适用于那些正考虑在欧盟市场上销售和使用的高风险人工智能应用上。

如前所述，没有数据就没有人工智能。许多人工智能系统的功能以及它们可能导致的行动和决策在很大程度上取决于对系统进行训练的数据集。因此，应采取必要的措施，确保在用于训练人工智能系统的数据能够符合欧盟的价值观和规则，特别是

与安全性和保护基本权利的现行立法规则相关联时。如下是有关用于训练人工智能系统的数据集的要求：

- 要求旨在合理保证人工智能系统支持的产品或服务的后续使用安全，基于此，它应满足欧盟安全规则（现行及可能的补充规则）中适用的一系列标准。例如，要求确保人工智能系统接受足够广泛的、涵盖了所有相关场景以避免危险情况的数据集培训。

- 采取合理措施的要求旨在确保人工智能系统在后续使用中不会产生禁止的歧视结果。这些要求可能需要在特定的义务下使用具有足够代表性的数据集，尤其确保这些数据集适当地反映了性别、种族和其他禁止的歧视原因的所有相关维度；

- 要求旨在确保使用基于人工智能的产品和服务的期间，充分保护隐私和个人数据。《通用数据保护条例》和《执法指令》就属于其各自范围内的问题进行了规定。

b） 保留记录和数据

考虑到许多人工智能系统的复杂性和不透明性等因素，以及在有效核实相关适用规则符合性和执行方面存在困难，因此要求对与算法编程、及用于训练高风险人工智能系统的数据有关的记录进行保留，在某些情况下，还需保留数据本身。这些要求基本上可以支持追溯和验证人工智能系统的有问题的操作或决策。这不仅有利于监督和执行，而且也可能激励相关经济运营者，使他们在早期阶段就考虑到尊重这些规则的必要性。

为此目的，监管框架可以要求对如下内容进行保存：

- 有关用于训练、测试人工智能系统的数据集的准确记录，包括对主要特征的描述及如何选择数据集；

- 在某些合理情况下，数据集本身；
- 有关编程[53]和训练方法，构建、测试和验证人工智能系统的过程和技术等的记录文件，包括在安全性和避免可能导致禁止性歧视的偏差的相关信息。

记录、文件和相关数据集应在有限的合理期间内进行保存，以确保有效执行相关法律。应采取措施确保对应记录及资料在需要时及时可用，尤其在监管当局进行测试或检查时。必要时，应做出安排以确保机密信息（例如商业秘密）受到保护。

c） 信息提供

除上文章节C中讨论的记录保存要求外，透明度也是必不可少的。为了实现既定目标，尤其为促进人工智能的负责任使用、建立信任并在需要时方便救济，主动提供高风险人工智能系统使用的充分信息至关紧要。

相应地，应考虑如下要求：

- 确保提供清晰的信息，说明人工智能系统的功能和局限性，特别是系统的预期用途、按预期运行所需的条件，以及实现该用途的预期精准度。这些信息对系统部署者尤其重要，但也可能与监管部门和受影响的各方相关。
- 另外，应明确告知公民与其互动的是人工智能系统而非人类。尽管欧盟数据保护法规已经包含某些此类规则[54]，但仍可能需要其他附加要求以实现上述目标。即便如此，仍应避免增加不必要的负担。因此，存在例外情况无需提供此类信息，例如在公民能很快意识到与他们互动的是人工智能系统的情况下，无需告知。更为重要的是，提供的信息必须客观、简明且易理解。提供信息的方式应根据特定的场景进行调整。

d） 鲁棒性和准确性

[53]例如，包括模型应优化的内容，针对某些参数起初设计权重等在内的算法文档。
[54]特别是，根据《通用数据保护条例》（GDPR）第13（2）（f）条的规定，控制者必须在获取个人数据时向数据主体提供必要的进一步信息，以确保对自动决策的存在以及某些其他附加信息进行公平和透明的处理。

人工智能系统——包含高风险人工智能应用在内——必须具备技术上的鲁棒性和准确性，才值得信赖。这意味着这些系统应以负责任的方式开发，并预先适当考虑其可能产生的风险。它们的开发和功能设计必须可以确保人工智能系统按照预期可靠运行。应采取合理措施最大程度降低可能造成损害的风险。

相应地，应考虑以下几方面：

- 要求确保人工智能系统在其生命周期的所有阶段都具有鲁棒性和准确性，或至少正确反映了它们的准确性程度；
- 确保结果可复现；
- 要求确保人工智能系统能够在其生命周期的所有阶段均能充分处理错误或不一致性。
- 要求确保人工智能系统能够抵御公开攻击和更隐蔽的数据或算法操纵企图，并在这种情况下采取缓解措施。

e）人类监督

人类的监督有助于确保人工智能系统不会违背人类自主性或造成其他不利影响。只有通过人类对高风险人工智能应用的适当管理，才能实现可信赖的、合乎伦理的和“以人为中心”的人工智能的目标。

尽管本白皮书中基于特定法律制度的人工智能应用被认为是高风险的，但需要进行人类监督的适当类型及程度因情况而异。这尤其取决于系统的预期用途以及该用途对受影响的公民和法人实体可能产生的影响。当人工智能系统在处理个人数据时，亦不得损害《通用数据保护条例》（GDPR）所确立的合法权利。例如，人类监督可能具有以下尚未详尽罗列的表现形式：

- 除非事先经过人工审核和验证确认，人工智能系统的输出将不会生效（例如，拒绝社会保障福利申请的决定只能由人类做出）；

- 人工智能系统的输出立即生效，但其后必须进行人工干预。（例如，信用卡申请被拒可能由人工智能系统处理，但之后必须进行人工审核）；
- 监控人工智能系统的运行，且人类有能力实时干预、并使其失效（例如，在无人驾驶模式中，当人类认为汽车运行不安全时，可以使用停止按钮或操作程序）；
- 在设计阶段，对人工智能系统施加操作限制（例如，当感应器不那么可靠时，无人驾驶模式应停止在某些能见度较低的条件下运行，或在任何情况下与前保持一定距离）。

f）远程生物识别的具体要求

诸如通过在公共场所部署人脸识别系统等，为了远程识别[55]而收集和使用生物特征数据[56]会给基本权利[57]带来特定的风险。根据使用目的、环境和范围的不同，使用远程生物识别人工智能系统对基本权利的影响可能会有很大差异。

欧盟数据保护规则原则上禁止单纯为识别自然人的目的而处理生物识别数据，但在特定条件下除外[58]。具体而言，根据GDPR该处理只能基于有限的理由，其中主要理由是出于实质性公共利益的考虑。该情况下应根据欧盟或本国法律处理，且遵守相

[55]在人脸识别方面，识别是指将一个人的面部图像的模板与数据库中存储的许多其他模板进行比较，以查明他或她的图像是否存储在数据库中。另一方面，身份验证（或验证)通常被称为一对一匹配。它可以比较两个通常被假定属于同一个个体的生物特征模板。两个生物特征模板进行比较，以确定两个图像上显示的人是否是同一个人。例如，这种程序在机场边界检查的自动边境控制（ABC）登机口处使用。

[56]生物特征数据被定义为"有关自然人的身体、生理或行为特征的特定技术处理所产生的个人数据，允许或确认该自然人的唯一认证或身份证明，如面部图像或指印[指纹]数据。（执法指令，第3（13）条；通用数据保护条例（GDPR），第4（14）条；条例（EU）2018/1725，第3（18）条。

[57]例如，关于人的尊严。与此相关的是，在使用面部识别技术时，尊重私人生活和保护个人数据的权利是基本权利关注的核心。对儿童、老年人和残疾人等特殊群体的非歧视和权利也有潜在的影响。此外，言论、结社和集会的自由不应因使用技术而受到损害。参见：人脸识别技术：执法中的基本权利考量，https://fra.europa.eu/en/publication/2019/facial-recognition。

[58]GDPR第9条，第10条执法指令。另请参见第10条法规（EU）2018/1725（适用于欧盟机构和组织）。

称性要求、尊重数据保护权的实质并采取适当的保障措施。根据《执法指令》须严格执行该处理，原则上应获得欧盟或国家法律的授权并采取适当保障措施。由于为单纯识别自然人而进行的任何生物识别信息的处理可能与欧盟法律中一项禁止条款的例外情形有关，因此这种情况仍受到欧盟《基本权利宪章》的约束。

因此，根据当前的欧盟数据保护规则和《基本权利宪章》，只有在用途合理、目的相称且保障充分的情况下，才可将人工智能用于远程生物识别。

为解决在公共场所使用人工智能可能产生的社会关注、避免内部市场分歧，委员会将在欧洲就合理使用人工智能的具体情况（如有）及普适性保障措施展开广泛讨论。

E、 接受者

关于适用上述高风险人工智能应用相关的法律要求的管理接受者，应考虑两个主要问题。

首先，如何在相关经济运营者之间分配责任。诸多角色参与在人工智能系统的生命周期中，包括开发者、使用人（使用配备了人工智能技术的产品或服务的人员）以及潜在人员（生产商、分销商或进口商、服务提供商、专业或私人用户）。

委员会认为，在未来的监管框架中，每项责任都应由最有能力应对对应潜在风险的角色承担。例如，人工智能的开发者可能最适合解决开发阶段所产生的风险，而他们在使用阶段控制风险的能力可能会受到更大限制，所以在使用场景下使用人应该承担相关责任。这并不妨碍承担对最终用户的责任，或在遭受损害的他方诉诸司法时确定哪一方应对损害承担责任。根据欧盟产品责任法，由生产者承担缺陷产品的责任，但不影响国家法律规定允许向他方追偿。

其次，存在立法干预的地域问题。委员会认为，最重要的是将该要求适用于在欧盟境内提供基于人工智能技术的产品或服务的所有经济运营者，无论其是否在欧盟境内设立。否则，就不能完全实现上述提及的立法干预的目的。

F、 合规与执法

为了确保人工智能可信、安全且遵守欧洲的价值观和规则，必须在实践中遵守适用的法律规定，并由国家和欧洲主管部门以及相关方有效执行。主管部门应有权调查个别案件，并相应评估其社会影响。

鉴于一些人工智能应用会对公民和我们的社会造成高风险（参见上文章节A），委员会认为在此阶段有必要进行客观的事前合规评估，以验证和确保高风险应用程序（参见上文章节D）符合上述某些适用的强制性要求。事前合规评估可能包括测试、检查或认证的程序[59]。可能包括检查开发阶段使用的算法和数据集。

针对高风险人工智能应用的合规评估应成为欧盟内部市场现有的诸多产品合规评估机制的一部分。若没有现成机制，可借鉴最佳实践、利益相关方和欧洲标准组织的相关输出以建立类似的机制。此类新机制应遵循国际义务，如相称、非歧视、使用透明及标准客观。

基于事前合规评估进行系统设计和实施时，应特别注意以下几点：

- 并非以上概述的所有要求均适合通过事前合规评估进行验证。例如，要求提供信息通常不适合通过此类评估进行验证。

- 应特别考虑某些人工智能系统进化发展并从经验中学习的可能性，可能需要在该人工智能系统的整个生命周期中反复进行评估。

- 需要验证用于训练的数据，以及用于构建、测试和验证人工智能系统的相关编程、训练方法、过程和技术。

- 若合规评估表明人工智能系统不满足要求（如与用于训练它的数据有关的要求），则需要纠正已识别的缺陷，例如，在欧盟对该系统进行再训练以确保满足所有适用要求。

[59]系统将基于欧盟的合规评估程序，请参阅第768/2008 / EC号决定或基于法规（EU）2019/881（《网络安全法》），考虑人工智能的特殊性。请参阅《欧盟产品规则实施蓝皮书指南》，2014年。

对于所有适用要求的经济运营者（无论其设立地点如何[60]），都必须进行合规评估。为了减轻中小企业的负担，可以设立支持性架构（包括通过数字创新中心）。此外，标准以及专用在线工具也可以协助合规。

任何事前合规评估均不应影响国家主管部门对合规情况的监控和事后执行。不仅是高风险的人工智能应用，其他人工智能应用也需要符合法律规定，虽然前者因其高风险特点而尤其引起国家主管部门的特别关注。事后控制应基于相关人工智能应用的充分文档记录进行（参见上文章节E），并在适当情况下由第三方（如主管部门）测试此类应用程序。因在某些场景下可能会对基本权利造成风险，所以上述措施尤其重要。相应的合规监控也应成为市场持续监控计划的一部分。下文章节H将进一步讨论与治理相关的内容。

此外，高风险人工智能应用和其他人工智能应用都应确保对受到人工智能系统负面影响的相关方实施有效的司法补救。与责任相关的问题将在本白皮书附录中“关于安全与责任框架的报告”中进一步讨论。

G、 无高风险人工智能应用的自愿标签

对于不属于“高风险”（参见上文章节C）且因此不受上述强制性要求约束（参见上文章节D、E和F） 的人工智能应用，除法律要求外，应建立自愿标签计划。

根据该计划，无需遵守强制性要求但对此感兴趣的经济运营者可以自愿决定遵从该要求或特别为自愿标签计划设立的具体类似要求。相关经济运营者也将相应获得其人工智能应用的优质标签。

自愿标签允许那些考虑了加强要求的经济运营者向市场释放信号，表明其人工智能产品和服务是可信赖的。它也方便用户意识到其心存疑虑的产品和服务其实是符合特定目的和欧盟范围内的标准化基准的，是超越了通常适用的法律责任要求的。这有助于增强用户对人工智能系统的信任，并促进该技术的大范围适用。

[60]关于相关治理架构，包括指定进行合规评估的机构，请参阅下文章节H。

此选项要求创建一个新的法律工具，来为非高风险人工智能系统的开发者和/或使用者制定自愿标签框架。尽管参与标签计划是自愿的，但一旦开发者或使用者选择使用标签，相关要求将自动产生约束力，并将结合事前和事后的执行确保其遵守所有要求。

H、 治理

应以国家主管部门合作框架的形式建立关于人工智能的欧洲治理架构，以避免职责分散、提升成员国能力，并确保欧洲逐步具备与其需求匹配的、测试和认证人工智能产品和服务的能力。在该场景下，支持国家主管部门在人工智能使用时完成其任务将是有益的。

欧洲治理架构可包含多个任务，例如作为定期交流信息和最佳实践、识别新兴趋势、提供标准化活动和认证建议的论坛。它还应在促进法律框架的实施中发挥关键作用，例如发布指引、意见和专门知识。为此，它将依赖于国家主管部门、国家和欧盟层面行业组织和监管部门的支持。此外，专家委员会也可以协助欧盟委员会。

治理架构应确保利益相关方最大限度地参与进来。应就框架的实施和进一步发展征询利益相关方（如消费者组织和社会合作伙伴、企业、研究人员和公民社会组织）的意见。

考虑到业已存在的架构，例如金融、制药、航空、医疗设备、消费者保护、数据保护，建议治理架构不应重复现有职能。相反，它应与各行业的其他欧盟和国家主管部门建立紧密联系，以补充现有的专业知识，并帮助现有部门监控和监督涉及人工智能系统以及支持人工智能的产品和服务的经济运营者的活动。

最后，若采用该方法，则可以将合规评估委托给成员国指定的机构。测试中心应根据上述要求，对人工智能系统进行独立审核和评估。独立评估将增加信任并确保客观性。它还可以促进有关主管部门的工作。

欧盟拥有出色的测试和评估中心，并应在人工智能领域发展其能力。在第三国设立

的经济运营者拟进入内部市场的，可使用在欧盟设立的指定机构；或者在与第三国达成互认协议的前提下，向指定进行该评估的第三国机构求助。

与人工智能有关的治理架构以及可能进行的合规评估不会影响相关主管部门在现行欧盟法律下关于特定行业或特定问题（财务、药品、航空、医疗设备、消费者保护、数据、相关法规等）的权力和责任。

6. 结论

人工智能是一项可以为公民、公司和整个社会带来诸多好处的战略技术，只要其立足于“以人为中心”、合乎道德、可持续发展并尊重基本权利和价值。人工智能带来了重要的效率和生产力收益，可以增强欧洲工业竞争力并改善公民福利。它还有助于找到最紧迫社会问题的解决方案，包括应对气候变化和环境退化、与可持续性和人口变化有关的挑战、以及保护我们国家民主，并在必要和适当的情况下打击犯罪。

为了充分抓住人工智能提供的机会，欧洲必须发展并加强必要的工业和技术能力。正如随附的欧洲数据战略所述，还需要采取措施使欧盟能够成为全球数据中心。

欧洲实现人工智能的途径旨在提升欧洲在人工智能领域的创新能力，同时支持整个欧盟经济中符合道德要求、且可信赖的人工智能技术的发展和使用。人工智能应该为人们服务，并成为造福社会的力量。

借本白皮书和随附的关于安全与责任框架的报告，委员会兹向成员国民间组织、行业和学术界广泛征询关于欧洲人工智能解决途径的具体建议。包括增加研究和创新投资、增强技能发展并支持中小企业使用人工智能的政策手段，以及针对未来监管框架关键要素的建议。该征询允许与所有有关方面展开全面对话，并为委员会的下

委员会通过以下网站公开邀请公众对白皮书提出建议并发表意见：https://ec.europa.eu/info/consultations_en.该意见的征询将开放至2020年5月19日。

委员会一般会发布关于征询公众意见的回复函。其中可能包含部分或全部建议的内容，但会确保内容的机密性。此情况下，请您在提交内容的首页明确标明禁止公开，并将您所提交内容的非保密版本发送至委员会以供发布。

一步工作提供参考。