布鲁塞尔，19.02.2020

COM（2020）最终版

欧盟委员会

关于人工智能———通往卓越和信任的欧盟道路

白皮书

译者：林鑫 刘嫱 万义龙 项一晴 姚凯 姚聪慧 曾晨
校对：王捷 王丹 朱莎
（以拼音排序，不分先后）

2020 年 3 月

译校人员简介

公益项目主编&领队：

王　捷：浙江垦丁律师事务所海外负责人 个人信息与数据安全合规/资深出海法律顾问（广州）

公益项目副领队：

朱　莎：浙江垦丁律师事务所 金融数据安全/资本市场跨境投资法律顾问

公益项目队员：

林　鑫：网络数据中心公司 法律顾问 系统集成网络、数据中心、大数据与人工智能等通信研究法律合规方向

刘　嫱：北京瑞银律师事务所律师

王　丹：重庆静昇律师事务所律师 商法（尤其是金融私法）、德国与欧盟私法方向，关注网络安全与数据法

万义龙：某公职单位，从事法律实务工作

项一晴：浙江理工大学政法学院 国际经济法方向

姚　凯：欧喜投资(中国)有限公司 IT 总监，信息管理、信息安全和数据保护方向

姚聪慧 某上海德资律师事务所律师 知识产权与食品合规方向

曾　晨：　西安交通大学 信息安全法律研究中心 网络安全与个人信息安全方向

关于人工智能的白皮书

通往卓越和信任的欧盟道路

人工智能发展迅猛，其将通过多种方式改变我们的生活，如改善医疗保健（例如：诊断更加精确、有能力更好地预防疾病）、提高农业耕作效率、促进气候变化的缓和与适应、通过预测性维护提高生产系统的效率、增强欧洲生产者的安全性，以及通过很多其他我们想象的方面来改变我们的生活。与此同时，人工智能（AI）也会带来若干潜在的风险，比如不透明的决策、基于性别或其他类型的歧视、对我们个人隐私的侵犯，或用于犯罪目的。

在激烈的全球竞争的背景下，需要在 2018 年 4 月提出的欧洲人工智能战略[1]的基础上，建立稳固的欧盟道路。为了应对人工智能的机遇和挑战，基于欧盟所追寻的价值观，欧盟必须团结一致，制定适合自己的方式，以促进人工智能的发展和部署。

欧盟委员会致力于实现科技突破、维持欧盟的科技领先地位，并确保新技术服务于所有欧洲人民，以改善他们的生活，同时尊重他们的权利。

欧盟委员会主席乌尔苏拉·冯·德莱恩在她的政治方针[2]中宣布，欧洲采取了协调一致的方法来处理人工智能对人类和道德伦理影响，并反思了如何更好地利用大数据进行创新。

因此，欧盟委员会支持以监管和投资为导向的方法，其双重目标是，促进实现人工智能的应用并解决与这项新技术的使用风险问题。本白皮书的目的，是就如何实现这些目的提供政策建议，但不涉及为居室目的的人工智能的开发和使用。欧盟委员会邀请了成员国、其他欧洲机构，以及所有利益相关者，包括工业、社会合作伙伴、民间社会组织、研究人员、普通公众和利益团体，对以下事项发表意见，并就此领域的决策为欧盟委员会的未来做出贡献。

1. 引言

随着数字技术在人们生活的方方面面都成为了更加重要的核心部分，人们应当能够信任数字技术。信任也正是实现此技术的先决条件。鉴于欧洲对价值观和法治的高度重视，以及其有着能制造和提供从航空到能源、汽车和医疗设备等安全、可靠和复杂的产品和服务的能力，这对欧洲来说是一个机会。欧洲当前和未来经济的可持续增长和社会福祉的维持，将越来越多地依赖利用数据所创造出来的价值。人工智能是数字经济的最重要的应用之一。如今，大多数数据都与消费者有关，并且在基于中央云的基础架构上进行存储和处理。相比之

[1] AI for Europe, COM/2018/237 final 欧洲的人工智能战略

[2] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

下，未来更为丰富的数据中，很大一部分将来自于工业、商业和公共部门，并且将会被存储在各种不同的系统中，尤其是存储在网络边缘的计算机设备上。这为欧洲打开了新的机遇，欧洲在数字化行业和企业对企业应用程序中拥有强大的地位，但在消费者平台中则处于相对弱势的地位。

简言之，人工智能是集数据、算法和计算能力的技术的集合。因此，计算的进步和数据可用性的提高是当前人工智能兴起的关键驱动力。正如欧盟数字方针所述，欧洲可以将其技术、工业力量优势以及高质量的数字基础设施和基于其基本价值观的监管架构[3]相结合，**从而成为数字经济和应用创新上的全球领先者。**在此基础上，它可以发展一个人工智能生态系统，从而将技术的优势带给整个欧洲社会和经济:

- 为**公民**获得新的利益，如改善医疗保健、减少家用机械故障，使交通系统更加安全和清洁，以及实现更好的公共服务；
- 促进**商业**发展，如欧洲有特殊优势的新一代产品和服务（机械、交通、网络安全、农业、绿色和循环经济、医疗保健以及时尚和旅游业等高附加值行业）；和
- **公共利益**服务，如通过降低服务成本（交通、教育、能源和废物管理），通过提高产品[4]的稳定性和通过执法机构配备适当的工具以确保公民的安全[5]，采取适当的保障措施，以尊重他们的权利和自由，。

考虑到人工智能可能对我们的社会产生重大影响，以及建立信任的需求，欧洲人工智能必须以欧洲的价值观和人类基本权利(如人的尊严和隐私的保护)为基础,这一点至关重要。。

此外，人工智能系统的影响力不仅应当从个人角度来考虑，还应当从社会整体来考虑。人工智能系统的应用可以在实现社会持续发展、支持民主化进程和社会权利上扮演重要角色。在最近欧洲提出的关于《欧洲绿色交易》[6]的提议中，欧洲在应对有关气候和环境挑战的道路上起主导地位。人工智能等数字技术是实现绿色交易目标的关键推动力。随着人工智能重要性的提高，需要在其整个生命周期和整个供应链中适当考虑人工智能系统对环境的影响，如关于算法培训和数据存储的资源利用上。

为实现规模化和避免单一市场的碎片化，欧洲必须采取一种通用的人工智能方法。采取国家层面的行动可能会危害法律稳定性，削弱了公民的信任和妨碍了亟需多元化有活力的欧盟工业发展。

- 在充分尊重欧盟公民的价值观和权利的基础上，本白皮书介绍了实现欧盟人工智能可信赖和安全发展的政策选择。本白皮书主要内容为：此政策框架阐述了整

[3] COM(2020) 66 final

[4] 一般而言，人工智能和数字化是欧洲绿色政策决心的重要驱动力。然而，信息通讯部门目前对环境的影响预计已经远超所有全球排放量的 2%。本白皮书随附的欧洲数据战略为数字化提供了绿色的前瞻性措施。

[5] 人工智能工可以提供机会，更好地保护欧盟公民免受犯罪和恐怖行动。例如：帮助识别网上恐怖宣传的工具，发现出售危险产品的可疑交易，识别危险隐藏物或违法物质和产品，在紧急时为公民提供帮助以及帮助指导急救人员。

[6] COM(2019) 640 最终版

合欧洲、国家和地区层面努力的措施。在个人和公共部门的合作下，本框架的目标是调动资源以实现整个价值链上的“**卓越生态系统**”，开启研究和创新；以及创造正确的激励措施以加快采取以人工智能为基础的方案，包括通过小中型企业(缩写为SMEs)。

■ 欧洲未来人工智能监管框架的主要关键点是创造独特的**“信任生态系统”**。为此，该监管框架必须遵守欧盟法律，包括保护基本权利和消费者权益的法规，尤其是在欧盟运行的具有高风险的人工智能系统[7]。建立信任生态系统本身就是一个政策目标，并且应当使公民有信心接受接受人工智能应用，并给公司和公共机构使用人工智能进行创新提供法律确定性。欧盟委员会大力支持走以人为本的道路，该以人为本的方式是在以人为本的人工智能[8]中建立可信交流为基础的，并且也会考虑参考人工智能高级别专家小组编写的《伦理准则》在试行阶段获得的有益成效。

本白皮书随附的《欧洲数据战略》（European strategy for data）旨在确保欧洲成为世界上最有吸引力、最安全、最具活力的、数据最敏捷的经济体——使欧洲能够利用数据改善决策和改善所有全体公民的生活。该战略提出了实现这一目标所需的很多政策措施，包括鼓励个人和公共投资。最后，在白皮书附带的欧盟委员会报告中对人工智能、物联网和其他数字技术对安全和责任法规的影响进行了分析。

2. 掌握工业和专业市场上的实力

欧洲处于有利位置，可从人工智能的潜力中获益，它不仅可以作为用户，而且可以作为该技术的创造者和生产者。欧洲拥有卓越的研究中心、创新型的初创企业，从汽车到医疗保健、能源、金融服务和农业等机器人技术和竞争性制造与服务领域均处于世界级领先地位。欧洲已经开发了强大的计算基础架构（例如高性能计算机），这对于人工智能的运行至关重要。欧洲还拥有大量的公共和工业数据，其潜能目前尚未被充分利用。欧洲在低功耗的安全可靠的数据系统方面具有公认的业界优势，这对人工智能的进一步发展至关重要。

利用欧盟在下一代技术和基础设施方面的投资能力，以及如数据素养等数字能力方面的投资能力，将增强欧洲在数据经济关键支持技术和基础设施方面的技术主权。这些基础设施应支持创建可信赖人工智能欧洲数据池，例如基于欧洲价值观和规则之上的人工智能。

从特定硬件生产行业到所有服务业的软件上，欧洲应当巩固其优势来扩大生态系统和价值链上的影响力。这已经在一定程度上发生了。在所有工业和专业服务（如精良农业、安全、健康、物流）方面，欧洲机器人的产能超过了四分之一，并且在公司和机构软件开发和

[7] 虽然，为防止和制止用于犯罪目的人工智能技术滥用，可能需要加入进一步的方案，但这已经超出本白皮书的范围。

[8] COM(2019)168 通讯

应用（B2B 应用如：企业资源规划、设计以及软件工程）以及支持电子政务和“智能企业”应用上扮演重要角色。

在制造业应用人工智能的部署上，欧洲处于领先地位。超过一半以上的顶级制造商在制造操作中至少实现了一个人工智能实力[9]。

欧洲在研究方面占据强势地位的原因之一是欧盟的资助计划已经被证明有助于共同采取行动，避免了重复和利用成员国的公共和私人投资。在过去的三年间，欧盟用于人工智能研究和创新的基金已经增加到了 15 亿欧元，也就是说与过去同期相比增加了 70%。

但是，欧洲在研究和创新方面的投资仍然只是世界其他地区公共和私人投资的一小部分。与 2016 年北美大约 121 亿欧元和亚洲 65 亿欧元[10]的投入相比，欧洲在人工智能上的投入是 32 亿欧元。对此，欧洲需要极大地增加投资比例。与成员国制定的人工智能[11]协作计划被证明是在欧洲建立更紧密的人工智能合作，以及在人工智能价值链投资最大化上创造协同效应的良好起点。

3. 抓住未来的机会：下一个数据浪潮

虽然欧洲目前在消费者应用程序和在线平台上还处于弱势地位，从而导致在数据获取上处于竞争劣势，但跨部门的数据价值和重用性正在发生重大转变。全球产生的数据量正在快速增长，从 2018 年的 33 兆字节到 2025 年预计 175 兆字节。数据每一次浪潮都为欧洲在数据-活跃经济上明确自身定位提供机会，并且在此领域成为领先者。此外，在未来五年中，数据存储和处理的方式将会发生极大的改变。如今，80%的数据处理和分析发生在数据中心和中心化的计算机设施的云端，20%发生在智能连接的对象，如汽车、家用电器或制造机器人以及与用户附近的计算设施中（“边缘计算”）。到 2025 年[12]，这些比例将发生显著变化[13]。

欧洲是低功耗电子产品的全球领导者，，这是下一代人工智能专用处理器的关键优势。这类市场目前是被非欧盟参与者所主导。在着力发展两端低耗能计算机系统和下一代高效能计算机方案如欧盟处理器方案的帮助下，以及预计在 2021 起实施的关键数据技术联合的工作下，可以改变这一现状。欧洲在神经形态解决方案中同样处于引领地位[14]，该解决方案非常适合于工业自动化发展（工业 4.0）和运输模式。其可以将数据效能提高几个数量级。

量子计算的最新发展将能够潜在地增加数据处理能力[15]。得益于欧洲在量子计算的科研优势和在量子模拟器和量子计算上的编程环境，欧洲可以在此技术走在前列。欧洲旨在增加量子检测和试验设备效能的方案将有益于量子解决方案在若干工业和科研部门的应用。

[9] 随后是日本（30%）和美国（28%），数据来源：: CapGemini (2019)
[10] 人工智能和自动化时代欧洲的当务之急，麦肯锡（2017）
[11] COM(2018) 795
[12] IDC (2019)
[13] Gartner (2017)
[14] 神经形态解决方案是指所有模仿神经生物结构的大型集成电路系统
[15] 量子计算机有能力在不到几秒钟内处理比目前最高性能计算机允许处理的数据还要大的数据量，这将促进跨部门新人工智能应用的发展。

同时，欧洲将继续凭借自身的科学卓越性，在以算法为基础的人工智能领域取得进步。在现有工作中，有必要在各单独学科之间建立规则桥梁，如机器学习和深度学习（特点是有限解释能力，需要大量数据空间来训练模型和相关性学习）和符号学方法（通过人为干预创建规则）。通过符号解释和深度神经网络的结合，可以帮助我们提供人工智能输出的解释能力。

4. 卓越的生态系统

要构建一个可以支持整个欧盟经济和公共管理领域人工智能发展和增长的卓越生态系统，需要在多个层面采取行动。

A. 与成员国协助

为了履行欧盟委员会在 2018 年 4 月通过的人工智能战略[16]，委员会在 2018 年 12 月发布了与各成员国共同拟就的协作计划，以促进欧洲人工智能的发展和使用[17]。

该计划提出了约 70 项联合行动，以期各成员国和委员会在研究、投资、市场增长、技能和人才、数据和国际合作等关键领域更紧密有效地合作。该计划预计持续到 2027 年，并将定期进行监督和审查。

计划的目的是最大限度地扩大在研究、创新和部署方面的投资的影响，评估本国人工智能战略，并与各成员国建立并扩展人工智能合作协作计划：

- ***行动 1：****委员会将参考白皮书的公众咨询结果，向各成员国提出修订的协作计划，并于 2020 年底前由各成员国采纳*

欧盟层面的人工智能资金平台应引导投资集中流向单个成员国无法独力支持的领域目标是在未来十年内每年在欧盟吸引超过 200 亿欧元的总投资[18]。为了激发私人和公共投资，欧盟将提供来自“数字欧洲计划”、“欧洲地平线”以及“欧洲结构和投资基金”的资源，满足欠发达地区和农村的需求。

协作计划协还将社会和环境福祉作为人工智能的关键原则。 人工智能系统有望帮助解决包括气候变化和环境退化在内的最紧迫问题。其过程也必须是环保的。 人工智能自身可以而且应该严格审查资源使用和能源消耗，并通过训练做出有利于环境的选择。委员会将与成员国一道考虑举措以激励和促进人工智能解决方案。

B. 关注研究和创新社区的努力

[16] Artificial Intelligence for Europe, COM(2018) 237
[17] Coordinated Plan on Artificial Intelligence, COM(2018) 795
[18] COM(2018) 237

目前能力中心分散，没有一个能达到与全球领先机构竞争所需的规模，欧洲不能维持这样的格局。必须在各个欧洲人工智能研究中心之间建立更多的协作和网络，并协调他们的工作以提升卓越水平，挽留和吸引最好的研究人员并开发最好的技术。欧洲需要一个研究、创新和专业能力的灯塔中心协调这些努力。该中心应成为人工智能全球典范，并可以吸引该领域的投资和最优秀的人才。

这些中心和网络应聚焦于欧洲有潜力成为全球冠军的行业，例如工业、卫生、运输、金融、农业食品价值链、能源/环境、林业、地球观测和太空。在这些领域争夺全球领导地位的竞争中，欧洲具有巨大的潜力、知识和专业能力[19]。 同样重要的是创建测试和实验场所支持新型人工智能应用程序的开发和后续部署。

- ***行动 2：****委员会将协助建立卓越中心和测试中心以整合欧洲、国家和私人投资，这可能包括一项新的法律工具。作为 2021 至 2027 年度财政框架的一部分，欧洲委员会提议一个雄心勃勃的专项资金以支持数字欧洲计划下的欧洲世界参考测试中心（world reference testing centres），并在适当情况下可通过欧洲地平线计划的研究和创新行动予以补充。*

C. 技能

欧洲人工智能的发展需要以对技能的高度重视为基础，弥补竞争力的缺失[20]。委员会将很快提出技能强化的议程，目的是确保欧洲每个人都能从欧盟经济的绿色和数字化转型中受益。举措还可以包括利用人工智能技术支持行业监管部门而有效并高效执法。更新的《数字教育行动计划》将有助于更好地利用学习和预测分析等数据和基于人工智能的技术改善教育和培训系统并使之适应数字时代。该计划还将在各级教育中提高人们对人工智能的认识，使公民对知情决策将会受到人工智能越来越多地影响做好准备。

发展与人工智能相关的必要技能并提高员工技能以适应由人工智能主导的转型，是与成员国一起制定的修订后的人工智能协作计划中优先事项。这可能包括将道德准则的评估清单转换为针对人工智能开发人员的指示性“课程”，并作为培训机构的资源提供。应作出特别的努力增加在这一领域接受培训和就业的妇女人数。

此外，欧洲人工智能研究与创新灯塔中心因其提供的可能性将吸引来自世界各地的人才。它还将发展和传播扎根于欧洲并得到发展的卓越技能。

- ***行动 3****：以领先的欧洲大学和高等教育机构“数字欧洲计划”网络支撑的高级技能为基础，建立在其上并得到其支持，吸引最优秀的教授和科学家，并提供人工智能领域世界领先的硕士学位课程。*

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设计和使用的影响。社会合作伙伴的参与将是确保在工作中以人为本的人工智能方法的关键因素。

[19] 未来的欧洲防卫基金和永久结构性合作（PESCO）也将为人工智能研发提供机会。 这些项目应与致力于人工智能更广泛的欧盟民用计划同步。

[20] https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced-profiles-eu

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]和人工智能需求平台[22]并促进中小企业间的合作。数字欧洲计划将有助于实现这一目标。虽然所有数字创新中心都应为中小企业提供支持，帮助他们理解和采用人工智能，但重要的是每个成员国至少有一个在人工智能方面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基础上，欧盟委员会计划根据欧盟投资计划[23]进一步扩大人工智能的融资渠道。欧盟投资计划中的担保所覆盖的合理使用领域明确提及了人工智能。

- ***行动 4：****委员会将与成员国合作，确保每个成员国至少有一个高度专业化的人工智能数字创新中心。 可以由"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。 根据 MFF 的最终协议，欧盟委员会意图从 2021 年起通过欧盟投资计划大幅扩大该规模。

E. 与私营部门合作

同样重要的是，确保私营部门充分参与制定研究和创新的议程并提供必要程度的共同投资。 这要求建立广泛的公私合作伙伴关系，并确保公司最高管理层的承诺。

- ***行动 5：****在"欧洲地平线"计划背景下，委员会将在人工智能、数据和机器人技术领域建立新的公私合作伙伴关系，共同努力，确保协调人工智能研究与创新，与"欧洲地平线"中的其他公私伙伴关系协助，并与上述测试设施和数字创新中心合作。*

F. 促进公共部门采用人工智能

公共行政部门、医院、公用事业和运输服务、财政监管以及其他公共利益领域必须迅速在他们活动中开始部署基于人工智能的产品和服务。 重点将放在技术成熟且可大规模部署的医疗保健和运输领域。

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

- ***行动 6：****委员会将发起开放和透明的部门对话，优先考虑医疗保健、农村行政管理和公共服务运营商，以提出一项促进发展、试验和接受的行动计划。 行业对话用于准备特定的"接受人工智能计划"，以支持人工智能系统的公共采购，并帮助转变公共采购流程。*

G. 强化数据访问和计算基础结构安全

本白皮书中列出的行动领域是在欧洲数据战略下并行提出的其他计划的补充，根本在于改善数据的访问和管理。没有数据，就不可能开发人工智能和其他数字应用程序。 尚未产生的大量新数据为欧洲提供了一个机会，使欧洲站在数据和人工智能转换最前沿。 提倡负责任的数据管理实务和基于 FAIR 原则的数据合规，将有助于建立信任并确保数据再利用性[24]。 同样重要的是对关键计算技术和基础架构的投资。

欧洲委员会已在"数字欧洲计划"下提议超过 40 亿欧元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。欧洲数据战略进一步确定了这些优先事项。

H. 国际方面

在建立价值共同体并促进符合伦理使用人工智能方面，欧洲处于充分的优势地位，可以发挥全球领导作用。 欧盟在人工智能方面的工作已经影响了国际讨论。在制定其伦理准则时，高级别专家组包括许多非欧盟组织和一些政府观察员。同时，欧盟密切参与了经济合作与发展组织人工智能伦理准则[25]的开发。20 国集团随后在 2019 年 6 月关于贸易和数字经济的部长级声明中认可了这些准则。

同时，欧盟认识到在其他多边论坛中正在开展的有关人工智能的重要工作。这些论坛包括欧洲委员会、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，这包括有关人工智能的建议。

在欧盟规则和价值观的基础上（例如，支持逐优的监管趋同、访问包括数据在内的关键资源、创造公平的竞争环境），欧盟继续与志同道合的国家以及人工智能领域的全球参与者合作。委员会将密切关注第三国对数据流的限制政策，通过双边贸易谈判和在世界贸易组织框架内采取行动解决不当限制。委员会坚信，在人工智能事务上的国际合作必须基于对基本权利的尊重包括人的尊严、多元化、包容、非歧视和隐私与个人数据保护[26]等，并将努力在全球范围内推广这些价值[27]。 显然，负责任地开发和使用人工智能可以推动实现可持续发展目标和推进 2030 年议程。

---

[24] 欧盟委员会公平数据专家委员会在 2018 年关于 FAIR 最终报告和行动计划（https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf）所述的可发现、可访问、可互操作和可重用。

[25] https://www.oecd.org/going-digital/ai/principles/

[26] 根据《伙伴关系项目》，欧盟委员会将资助一项 250 万欧元的项目，促进与志同道合的合作伙伴合作，推广欧盟人工智能伦理准则并采纳共同的原则和运营结论。

[27] President Von der Leyen, 《一个争取更多的联盟——我的欧洲议程》，第 17 页

5. 可信赖的生态系统：人工智能的监管框架

正如任何新技术一样，人工智能的使用也将同时带来机遇和挑战。面对算法决策的信息不对称，公民害怕这将使他们在捍卫自己的权利和安全方面变得无能为力，公司则担心法律的不确定性。尽管人工智能可以用于保障公民安全和确保他们行使基本权利，但他们仍然对于人工智能可能导致的某些意想不到的后果，以及被用于某些犯罪目的感到担忧。我们必须关切这些忧虑。除缺少投资和技术以外，缺乏信赖是阻碍人工智能更加广泛应用的主要因素。

为此，欧盟委员会于 2018 年 4 月 25 日制订了一项人工智能战略[28]，旨在解决社会经济方面的问题，同时，在欧盟范围内增加针对人工智能研究、创新和能力的投资。欧盟委员会同成员国达成了协调计划[29] 以统筹各国的相应战略。委员会还于 2019 年 4 月组建了高级别专家组，该专家组就可信的人工智能发布指导准则[30]。

委员会曾发文[31]对高级别专家组指导准则中定义的七项关键指标表示欢迎:

*人类行为和人类监管的优先性
*技术可靠性和安全性
*隐私和数据管理
*透明性
*多样性、非歧视性和公正性
*社会和环境有益性
*可评估性 （Accountability）

此外，该指导准则包含有一个供公司实际应用的评估清单。2019年下半年，超过350个机构对该评估清单进行测试并反馈了情况。高级别专家组正在根据反馈的情况对指导准则进行修正，这项工作将于2020年6月完成。反馈过程中一个重要的结果是，尽管许多指标已经在现行法律和监管体制中得到体现，但是诸如透明性、可溯性及人类监管，在许多经济体的现行法律中并没有被特别涵盖。

基于高级别专家组不具约束力的指导准则和欧委会主席的政治方针，一个清晰的欧洲监管框架将会在消费者和企业建立起对人工智能的信赖，并进而加速促进技术的推广采纳。这个监管框架将和其他措施保持一致，加强欧盟在这一领域的创新能力和竞争

[28] COM(2018)237.
[29] COM(2018)795.
[30] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
[31] COM(2019) 168.

力。并且，定能确保社会、环境和经济的良性发展，符合欧盟的法律、原则和价值观。尤其是在事关对公民权有重大直接影响的领域，比如在行政执法和司法活动中应用人工智能。

人工智能的开发者和使用者都遵守欧洲在基本权利（比如数据保护、隐私、反歧视）、消费者保护、产品安全和责任制度方面的立法约束。无论一种产品或者系统是否依赖于人工智能，消费者都期待着相同水平的安全和对他们权利的尊重。然而，人工智能的某些属性（比如模糊性）将会导致法律适用和执行变得更加困难。因此，有必要检视现行法律能否有效施行并消除人工智能的风险，以及是否有必要修正法律或者重新制订新的法律。

考虑到人工智能升级迅速，监管框架必须为适应将来发展留有余地。任何变更都应当限于针对已明确识别出的问题，而这些问题已有可行的解决之道。

各成员国认为当前尚缺少一个共同的欧洲框架。德国数据伦理委员会呼吁建立一个基于风险的五级管理体系，该体系层级从对无害的人工智能系统几乎没有监管，直到对危险的人工智能系统完全禁止。丹麦提供了一个数据伦理印章的示范方案。马耳他引入了一个基于自愿的人工智能认证系统。如果欧盟不能提出一个全欧适用的解决方案，那带来真正的风险，进而破坏建立信赖、法律确定性以及市场开放的目标。

针对可信赖的人工智能建立一个可靠的欧洲监管框架，将有助于建立一个畅通的高度发展的内部市场的，有助于促进人工智能的推广使用，有助于强化欧洲在人工智能方面的工业基础。

A. 问题定义

尽管人工智能可以带来很多益处，包括使得产品和生产过程更加安全，但它同样也能造成伤害。这种伤害可能是有形的（个人的安全和健康，包括死亡或者财产损失），或者无形的（隐私泄露、限制言论自由权、人格尊严、就业歧视），并且可能导致一系列的风险。监管框架必须致力于将潜在伤害的风险最小化，尤其是那些重要的方面。

人工智能应用存在最主要的风险，是关于那些保障基本权利的规则应用（包括个人数据和隐私保护以及反歧视），安全问题[32]和责任问题。

基本权利风险，包括个人数据和隐私保护以及反歧视

将人工智能应用于保护个人数据和私人生活[33]，司法救济和公正审判，及消费者保护等某些特定领域，可能会动摇欧盟赖以生存的价值观基础，并导致对基本权利的侵害

[32] 包括网络安全事件、重要基础设施中的人工智能应用事件，以及人工智能的恶意使用。

[33] 欧盟理事会的报告显示，人工智能的应用造成了大量的侵犯基本权利事件。
https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.

[34]，包括言论自由，集会自由，人格尊严，以及消除基于性别、族裔或种族出身、宗教信仰及世界观自由、残疾、年龄以及性取向等的歧视。这些风险可能是由于人工智能系统总体设计方面的缺陷所引起（包括在人员监管方面），也可能是基于数据滥用而导致的（比如系统训练只采用或者主要采用男性数据，从而导致女性处于劣势）。

人工智能可以承担很多以往只能由人类承担的工作。结果，公民和法人越来越受制于由人工智能系统做出的或者辅助做出的行为和决策，而这些行为和决策在某些必要的时候往往可能是难以理解和有效更正的。更甚者，人工智能将可能会更多的用于追溯和分析人们的日常习惯。举例来说，人工智能应用的一个潜在风险是，它将突破欧盟的数据保护规则或者其他规则，被行政当局或者其他组织用于监控民众，以及被雇主用于监视员工。通过大数据和其中关联性的分析识别，人工智能将被用于追溯和非匿名化某些个人数据，这将产生一个新的关于个人数据保护的风险，甚至此时数据集中并不包含个人数据。人工智能也可能被网络平台用于为用户进行信息和内容的优化推送。这些经过处理的数据，被设定的应用程序，以及人工所能够干预的程度，将会影响到人们的言论自由，数据保护，隐私以及政治自由。

用于预判刑事再犯的人工智能算法，会进行性别和种族偏见：性别不同、国籍不同，再犯率也就不同。参考文献:Tolan S., Miron M., Gomez E. And Castillo C. "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia", Best Paper Award, International Conference on AI and Law, 2019

用于人脸识别的人工智能程序，会进行性别和种族偏见：对浅肤色男性的性别识别错误率较低，对深肤色女性却较高的。参考文献: Joy Buolamwini, Timnit Gebru; Proceedingsofthe1stConferenceonFairness, Accountability and Transparency, PMLR 81:77-91, 2018.

偏见和歧视是任何社会和经济活动中内在固有的风险。人力决策也不能避免错误和偏见。但是，同样的偏见体现在人工智能身上将会产生更加严重的后果。对人类行为[35] 缺乏社会控制机制的时候，许多人会受到影响和歧视。这同样发生于人工智能边"学习"边运行的过程中。在此情况下，如果其后果不能被及时制止，或者在最初设计阶段没有准确预见到的话，那么，风险就不会来自最初的系统设计缺陷，毋宁源于系统于大数据中的交叉联系和模仿。

[34] 《通用数据保护条例》和《电子隐私保护条例》（新版的电子隐私保护条例正在制订中）都有涉及到这些风险问题，但是或许还是有必要考虑人工智能系统是否增加了额外的风险。欧盟委员会将会持续关注并评估通用数据保护条例的实施情况。

[35] 欧盟委员会男女机会平等咨询委员会正在准备一个"人工智能建议"，用以分析人工智能在性别平等方面带来的影响，该建议有望将于 2020 年初为委员会所采纳。欧盟性别平等战略 2020-2024 同样关注到了人工智能和性别平等之间的联系；欧洲平等实体网络将会在 2020 年初刊发报告，"管住人工智能:平等实体的新角色——直面人工智能和数字化给平等和非歧视带来的新挑战"。

许多人工智能技术的特定特征，包括模糊性（黑匣子效应），复杂性，不可预测性，以及部分的行为自主性，将会使得系统难以遵守欧盟现行规则，或者阻碍该规则的实施。这些规则原本是为了保障基本权利的。执法机关或者受到影响的个人，缺乏必要的手段以确认这些受到人工智能干预的决策是如何实施的，也无法确认这些决策是否遵守了相关规则。当这些决策对他们产生不利影响时，自然人或者法人都很难有效的寻求公正。

安全风险和侵权责任制度的有效运行

当人工智能技术嵌入到产品或者服务中时，可能会给用户带来新的安全风险。举例来说，当目标识别技术存在缺陷的时候，将会导致自动驾驶汽车错误识别路面物体，从而引发事故造成人员伤亡和财产损失。至于对基本权利的风险，这些风险可能产生于人工智能技术的设计缺陷，或源于数据取得与质量，或源于机器学习产生的问题。这些风险不仅仅局限于采用了人工智能技术的产品或服务，人工智能的使用也将进一步增加或者恶化这些风险。

如果缺乏明确的风险规避安全规则，不仅危及个人，还将会对那些在欧盟范围内经营人工智能相关产品的公司带来法律上的不确定性。市场监督管理机关也将会发现他们自己出于尴尬的境地。因为缺乏法律授权，并且／或者缺少足够的监管系统技术力量，他们将不知道自己能否进行干预[36]。法律不确定性将会整体的削弱欧洲公司的安全水平和竞争力。

如果安全风险成为现实，由于缺乏明确的指标，以及上述人工智能技术特征，追溯人工那些有问题的潜在的人工智能决策就会很困难。相应的，也会使得遭受损失的人们很难基于现行的欧盟和国际侵权责任法律主张损失赔偿[37]。

> 根据《产品责任指令》，制造商对缺陷产品造成的损害应当承担责任。然而，在自动驾驶汽车等基于人工智能的系统中，可能很难证明产品存在缺陷、已经发生的损坏或者两者之间的因果关系。此外，对于产品责任指令如何以及在何种程度上适用于某些类型的缺陷(例如，这些缺陷是否源于产品的网络安全缺陷)，还存在一些不确定性。

因此，如同上文有关基本权利的论述，追溯潜在的问题决策也会带来安全风险和责任承担的法律问题。受害人很难在法庭上进行有效举证以还原事实，因而，相对传统技术造成损害的情形，人们很难寻求救济。当人工智能技术广泛使用时，这些风险将进一步加剧。

B. 关于欧盟现行人工智能立法框架的可能性调整

[36] 一个关于儿童智能手表的例子。该产品或许不会对使用它的儿童造成直接的伤害，但是如果缺少最低限度的安全保障，它可能极易成为接近儿童的一个工具。市场监管当局则会发现在该产品和风险尚未发生联系的情况下，似乎很难进行干预。

[37] 此份白皮书在附件委员会报告中对人工智能、物联网及其他数字技术对安全和责任立法的影响进行了分析。

涉及产品安全及产品责任立法[38]的欧盟现行框架，适用主体广泛，其中包含行业特定规则，由国家立法机构进一步完善，可能适用于新兴人工智能应用且存在关联性。

在保护基本权利与消费者权益方面，欧盟立法框架囊括多方面的立法例如《种族与性取向平等指令》[39]、《就业平等指令》[40]、《男女就业平等指令》[41]、《商品与服务准入指令》[42]以及许多消费者保护法规，同样还有个人数据保护政策，尤其是通用数据保护条例及其他覆盖个人数据保护的行业性法律，例如《数据保护法执法指令》[43]。另外，从 2025 年开始，在《欧洲无障碍法案》中规定的商品和服务无障碍要求的规则将适用[44]。另外，当转换诸如金融服务、移民、网上中介责任等领域的其他欧盟指令时，应符合基本权利。

即使不考虑人工智能的参与，欧盟立法在原则上仍然完全适用，但尤为重要的是需要评估法律是其能够得到充分执行，以解决人工智能系统所带来的风险，或者是否需要对特定法律文本进行调整。

例如，经营者对人工智能现有保护消费者的规则负有全部责任，任何违反现有规则对消费者行为进行算法性利用的行为都将被禁止，若违反，将受到相应惩罚。

欧盟委员会认为，可以改进立法框架以应对以下风险和情形：

- 有效适用和执行现行的欧盟及国家立法：人工智能的关键特点是为确保欧盟和国家立法的适当适用和执行带来了挑战。缺乏透明度（人工智能的不透明度）很难查明和证明可能的违法行为，包含保护基本权利、界定责任、以及满足索赔条件的法律规定。因此，为了确保高效的应用及执法，可能有必要调整或澄清某些领域的现行立法，例如本白皮书所附报告中进一步详述的责任方面的立法。
- 现有欧盟法律的范围限制：欧盟产品安全法律的一个重要焦点即产品流通市场，而在《欧盟产品安全指令》中，软件在作为最终产品的一部分时必须遵守相关的产品安全规则，独立开发的软件是否被欧盟产品安全法覆盖仍是一个悬而未决的问题。除相关部门的明确规定外[45]，欧盟现行的安全规则通常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（如医疗服务、金融服务、运输服务）。
- 可改变的人工智能功能：将软件（包含人工智能）集成到产品后，这些产品和系统的功能可能在产品和系统的生命周期中改变。这对于需要频繁软件更新或依赖机

[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），以及涵盖从机器、飞机、汽车到玩具和医疗器械等不同类别产品的一些行业特定规则，旨在提供高水准的健康和安全保障。产品责任法中由于产品及服务所造成的损害赔偿，由不同的民事责任制度加以完善

[39] 指令 2000/43/EC

[40] 指令 2000/78/EC

[41] 指令 2004/113/EC; 指令 2006/54/EC

[42] 例如《不公平商业行为指令》（法令 2005/29/EC）以及消费者权益指令（2011/83/EC）。

[43] 指令（欧盟）欧洲议会以及理事会于 2016 年 4 月 27 日颁布的 2016/680 号指令，涉及主管当局在调查、侦查或起诉刑事犯罪、执行刑事处罚处理个人数据或个人数据自由流动时，起到保护自然人权益的目的。

[44] 指令（欧盟）2019/882 关于产品和服务的无障碍要求。

[45] 例如，制造商拟用于医疗目的的软件被视为医疗设备，需符合《医疗设备条例》(欧盟)2017/745 号规定。

器学习的系统尤其如此。这些特性会产生新的风险，而这些风险在系统最初流通于市场时并不存在。这些风险并不能于现有的法律中充分解决，毕竟现有的法律都聚焦于产品投放市场时存在的风险。

- 供应链中的不同运营商间责任分配的不确定性：通常，欧盟的产品安全法律将责任分配给流通市场产品的生产商，包含所有设备零件，如人工智能系统。但是，如果非生产商将产品流通于市场后，人工智能系统才被整合进来，此时是否可以援引规则并不明确。此外，欧盟产品责任法律规定了生产方的责任，并且规定了供应链的其他方的国家责任。
- 安全观念的改变：在产品以及服务中使用人工智能会产生风险，但欧盟现行法律并没有明确提及这一点。这些风险可能与网络威胁、个人安全风险（例如使用人工智能家用电器），连接中断等导致的风险有关。这些风险可能存在于产品流通市场的过程中，也可能是由于软件更新或产品使用过程中的自我学习。欧盟应当充分利用自己所掌握的法律工具，以加强人工智能应用可能潜在风险的事实基础，包括利用欧盟网络安全局（ENISA）的经验评估人工智能威胁状况。

正如上述所言，很多成员国已经在探索国家立法的备选方案，以应对人工智能带来的挑战。这就增加了单一市场可能碎片化的风险。不同的国家规则可能会给那些想在单一市场出售或运营人工智能系统的公司带来障碍。在欧盟层级采取共同的做法，可以使欧洲企业顺利进入单一市场从而受惠并且提高他们在国际市场的竞争力。

关于人工智能、物联网、机器人的安全及侵权责任影响报告

本报告，即本白皮书所附的报告，分析了相关的法律框架。它确认了当法律框架适用时所产生的人工智能系统及其他数字技术所带来的特定风险的不确定之处。

结论是，当前的产品安全法律已可以支持安全保护概念扩展来抵御人工智能产品使用中增加的风险。无论如何，规定明确涵盖了当前在数字科学中新的风险，从而可以提供更多法律的确定性。

- 某些人工智能系统在其生命周期内的自主行为可能对产品产生重要的安全性变化，这可能就需要新的风险评估。此外，只有人类的控制才能自始至终在人工智能系统及产品生命周期中提供保障。

- 为了解决用户心理安全风险问题，明确生产者的义务也应当适时地被考虑到。（前述：与仿真机器人合作）

- 欧盟产品安全法律既要对设计阶段的数据错误安全风险提出明确要求，同时该机制可以自始至终确保人工智能产品及系统使用的数据质量。

- 基于算法的系统的不透明度问题可以通过透明度要求来解决。

- 如果一个独立软件在市场上流通或在流通后才下载到产品中而对安全产生影响时，则可能需要对现有规则进行修改和澄清。

- 鉴于新技术在供应链中的复杂性，立法明确要求供应链中的经营者及用户间开展合作，以确保立法的确定性。

新兴数字技术如人工智能、物联网和机器人的特点可能会挑战侵权责任框架的某些方面，并可能降低其有效性。其中部分特征可能很难将损害追溯到个人身上。根据大多数国家规则，该种基于过失的索赔是必要的。这会显著增加受害者追索成本，也意味着难以对生产者以外的其他参与者提出或证明损害责任主张。

- 一方面因人工智能系统遭受侵害侵害的个人应与因其他技术而受害的个人享有同等程度的保护，另一方面，应保留技术创新发展的空间。
- 应仔细评估所有实现该目标的选项，包括产品责任指令可能发生的修改及国家侵权责任规范的相应协调。例如，委员会正在征求意见，是否以及在多大程度上需要调整国家侵权责任规则所规定的，关于人工智能应用的运营损害的举证责任，以减轻复杂性的后果。

从上述讨论看，委员会结论认为，除了尽可能调整现有法律，或许需要对于人工智能采取明确的立法，以使欧盟当前的法律框架，与当前和即将到来的科学技术及商业发展相匹配。

C. 未来欧盟监管框架的范围

未来有关人工智能智能的特定监管框架的关键问题是确定其应用范围。 可行的假设是，监管框架将适用于依赖人工智能的产品和服务。 因此，应在本白皮书以及未来任何可能的决策举措中明确定义人工智能。

> 例如，在自动驾驶中，算法使用来自汽车的实时数据（速度，引擎消耗，减震器等）以及来自扫描汽车整个环境的传感器的数据（道路，标志，其他车辆，行人等）计算出汽车到达特定目的地应采取的方向，加速度和速度。 根据观察到的数据，算法可以适应道路情况以及包括其他驾驶员行为在内的外界条件，并调整出最舒适，最安全的驾驶方式。

在其关于欧洲人工智能的通报中，委员会提供了人工智能[46]的第一个定义。高级别专家组[47]进一步完善了这一定义。

在任何新的法律工具中，人工智能的定义都需要足够灵活以适应技术进步，同时又要足够精确以提供必要的法律确定性。

对于本白皮书，以及未来任何有关政策倡议的讨论，阐明构成人工智能的主要要素（即“数据”和“算法”）都很重要。人工智能可以集成在硬件中。机器学习技术，属于人工智能的一个分支。机器学习技术是指基于一组数据，对算法进行训练以来推断某些规律，以找到实现某特定目标的行动。算法在使用的过程中会继续学习。尽管基于人工智能技术的产品可以通过感知其环境而自主采取行动，而无需遵循预定的指令，但其行为很大程度上受到其开发人员的定义和约束。 人类确定并编程目标，而人工智能智能系统会进一步优化。

欧盟拥有严格的法律框架，以确保对消费者的保护，解决不公平的商业行为并保护个人数据和隐私。此外，这个框架还包含某些行业（例如医疗保健，交通运输）的特定规则。欧盟法律的这些现有规定将继续适用于人工智能，尽管可能需要对该框架进行某些更新以反映人工智能的数字化转型和使用（请参阅 B 节）。 因此，现有水平或部门立法（例如医疗器械[48]，运输系统中）已经解决的那些方面将继续受该立法规制。

---

[46] COM(2018) 237 final，第 1 页:“人工智能(AI)是指通过分析其环境并采取行动-在一定程度上自治-来实现特定目标来显示智能行为的系统。基于人工智能的系统可以完全基于软件，在虚拟世界中发挥作用(如语音助手、图像分析软件、搜索引擎、语音和人脸识别系统)，也可以嵌入到硬件设备(如高级机器人、自动驾驶汽车、无人机或物联网应用)。”

[47] 高级别专家组，人工智能的定义，第 8 页:“人工智能(AI)系统软件(也可能是硬件)系统由人类设计,给出一个复杂的目标,在物理或数字维度通过感知环境数据采集、解释收集到的结构化或非结构化数据,知识推理,或处理信息,基于这些数据决定最佳动作，实现给定的目标。人工智能系统既可以使用符号规则，也可以学习数字模型，它们还可以通过分析环境如何受到之前行为的影响，来调整自己的行为。”

[48] 例如，对于向医生提供专业医疗信息的人工智能系统、直接向患者提供医疗信息的人工智能系统以及直接在患者身上执行医疗任务的人工智能系统，存在不同的安全考虑和法律含义。委员会正在审查这些安全与责任方面的挑战，这些挑战与医疗保健是截然不同的。

原则上，新的人工智能监管框架应能有效地实现监管目标，但又不应过度监管，以免造成不成比例的负担，尤其是对中小企业而言。为了达到这一平衡，委员会认为它应该采取一种以风险为基础的办法。

基于风险的方法对于衡量监管干预是否合比例很重要。但是，达到这个目的需要明确的标准来区分不同的人工智能应用程序，尤其是在鉴别什么属于"高风险"方面[49]。

高风险人工智能应用程序的判定应该清晰且易懂，并照顾各方利益关切。但是，即便一个人工智能应用程序不符合高风险条件，它也完全受制于现有的欧盟监管规则。

委员会认为，应同时根据其所属行业部门及预期用途，衡量人工智能应用是否存在高风险，特别是危及安全保障、消费者权利保护和基本权利的风险。具体地说，在同时满足以下两个标准的情况下，人工智能应用应视为高风险：

- 首先，考虑到通常进行的活动的特征，人工智能应用所属的行业一般预期会发生重大风险。第一个标准确保监管干预针对的是一般而言被认为最有可能发生风险的领域或行业。因而所涵盖的部门应在新的监管框架中明确而详尽地列出。例如，医疗保健、运输、能源和部分公共部门[50]。行业清单应定期进行审查，并根据实际情况在必要时加以修订；
- 第二，因人工智能在相关领域的应用的预期使用方式可能会产生重大风险。第二个标准反映了这样一种认识，即并非在选定行业中每次使用人工智能都必然涉及重大风险。例如，虽然医疗保健通常可能是一个高危行业，但医院预约排期系统的缺陷通常不会造成重大意义的风险，因而没有理由进行立法干预。对特定使用的风险水平的评估可以基于对关系人的影响。例如，使用人工智能应用对个人或公司的权利会产生法律或类似重大影响的；会造成受伤，死亡或重大财产或精神损害的风险的；会产生个人或法人无法合理避免的效果的等等。

同时应用这两项标准将确保监管框架的范围具有针对性，并提供法律确定性。有关人工智能的新监管架构所包含的强制性规定(见下文 D 节)，原则上只适用于根据这两项标准识别为高风险的应用程序。

尽管有上述规定，为了避免挂一漏万，在某些特殊情况下，即便不涉及相关高危行业，将人工智能应用程序用于某些特定目的仍被认为是高风险的[51]。作为举例说明，人们可能会特别考虑以下几点：

[49] 欧盟法规对"风险"的分类可能与此处所描述的有所不同，具体取决于具体领域，例如产品安全。
[50] 公共部门可以包括庇护、移民、边境控制和司法、社会保障和就业服务等领域。
[51] 需要强调的是，其他欧盟法规也可能同时适用。例如，当 AI 应用于消费产品时，其安全性也应受《通用产品安全指令》的约束。

- 鉴于对个人的重要性以及欧盟关于就业平等的规定，在招聘过程中以及在影响雇员权利的情况下使用人工智能应用将无例外地被视为“高风险”，因此，以下监管要求将始终适用。另外，影响消费者权利的其他特定应用也将被考虑进来。

- 远程生物特征识别[52]和其他侵犯私领域的监视技术的人工智能应用将始终被认定为“高风险”的，因此以下监管要求也将适用。

D. 要求的类型

在设计未来人工智能法规框架时，有必要就施加给相关人工智能参与者的强制性合法要求的类型进行讨论。此类要求可以通过国家标准进一步细化。根据上文 C 节部分所述以及除现有法律以外，此类要求只能适用于高风险的人工智能应用，以此确保所有监管干预合目的性、合比例性。

鉴于高级别专家组准则和前面所述的规则，对高风险人工智能应用的要求可以由下列关键特点构成，对此下面章节将会进一步讨论：

- 训练数据
- 数据保存与记录
- 待提供的信息
- 韧性（鲁棒性）与准确性
- 人为监督
- 对特定人工智能应用的特别要求，如：应用于远程生物特征识别的人工智能技术。

为保障法律的稳定性，此类要求将会被进一步细化，以便可以为需要遵守此类要求的所有参与者提供清晰的规范。

a) 训练数据

目前最重要的事情是促进、强化和捍卫欧盟的价值和法规，特别是欧盟法律赋予公民的权利。毫无疑问，这种努力也会延伸到这里关注的在欧盟面世和使用的高风险人工智能应用中。

正如之前的讨论，没有数据就没有人工智能。许多人工智能系统的应用及其引发的行为与决策在很大程度上取决于供系统训练的数据。为此，针对人工智能系统训练用数据，要考虑采取必要的措施来确保尊重欧盟的价值和法规，尤其是有关安全保障和基本权利保障的法律规定。针对人工智能系统培训数据集，可考虑下列要求：

[52]远程生物特征识别应当区别于生物认证（后者是一个依赖于个人特殊生物特性的安全程序，以保障其声称的某人正是其本人）。远程生物特征识别是指在远程、公共空间及以持续、正在进行的方式根据数据库中已收集的信息，借助生物特定（如：指纹、人脸、虹膜、血管分型）来识别多人。

- 要求确保后续使用人工智能系统所支持产品和服务安全性，因为这符合欧盟安全法规（现行和将来增补的法规）所设定的标准。例如：确保训练人工智能系统的数据集覆盖所有需要规避风险的场景；
- 要求采取合理手段，来确保人工智能系统的后续使用不会导致被禁止的歧视。这项要求包含一项义务，即必须采用具备足够代表性的数据集。尤其需要确保：此类数据集合理地顾及到可能导致被禁止歧视的方方面面，如性别、种族等；
- 要求确保使用人工智能产品或服务时，隐私与个人数据得到合理保护。《通用数据保护条例》和《执法数据保护指令》适用范围内的各个方面都应有法律行动进行规制。

b) 数据保存与记录

鉴于许多人工智能系统的复杂性、模糊性以及随之而来的合规审查和执法的困难，应当要求保存以下记录：算法编写记录、高风险人工智能系统的培训数据记录，以及某些数据的存储记录。这项要求主要是为了人工智能系统的潜在风险行为或决策能够被追溯和审查。这不仅便于监管与执法，也更加激励经营者尽早关注合规问题。

为此，法律框架内应规定保存以下记录的义务：

- 应用于培训和调试人工智能系统数据集的精确记录，包括对最重要的特征、类型、数据集拣选方式的记述；
- 在特定合理的情形，数据集本身的记录；
- 编写与训练[53]方法的记载，搭建、调试、校正人工智能系统的程序与技术的记载，要通过这些记载关注安全性，并避免发生可能导致的歧视偏见。

应在合理、有限的时段内保存这些记录和记载乃至数据集，以便有效执法；应当采取措施，保证这些记录和记载应询备查，尤其是有权机关的审阅与检查；必要时应采取预防措施，保护商业秘密等机密信息。

c) 信息提供

除了上述在(c)中讨论到的关于标识的要求以外，透明度也是必需的。为了实现所追求的目标--特别是促进负责任的使用人工智能，建立信赖并保证提供权利救济措施--重要的是以积极主动的方式在利用高风险人工智能系统时提供充分的信息。

据此，可以考虑下列要求：

[53] 如：记录算法，包括模型应该优化什么，哪些权重在一开始就被设定为特定参数等。

- 提供说明人工智能系统功能和限制的明确信息，特别是系统用途，它们可以在什么情况下按照预期运作以及达到指定用途的精确度。这些信息对系统的部署者尤为重要，也对主管机关和关系人有用。
- 另外，公民应当被清楚告知他们在于人工智能系统互动而不是与人类互动。虽然欧盟的数据保护法已经包含了一些此类规则[54]，为实现上述的目标，可能还需要一些额外要求。同时，要避免不必要的麻烦。例如公民明显在和人工智能系统交互等情况下，就不需要提供此类信息。此外，所提供的信息资料必须客观、简明、容易理解。提供信息的方式应根据具体情况进行调整。

d) 韧性（鲁棒性）和准确性

人工智能系统——当然还有高风险的人工智能应用——必须具备技术上的稳定性和精准性，才能值得信赖。这意味着需要以负责任的方式开发这种系统，并事先适当的考虑它们可能产生的风险。在开发阶段就要确保人工智能系统能够按照预期目的可靠运行。应采取一切合理措施来尽可能地降低造成损害的风险。

据此，可以考虑下列因素：

- 要求确保人工智能系统在其生命全周期都是稳定和精确的，或至少正确地反映其精确性水平；
- 要求确保结果的可再生产性；
- 要求确保人工智能系统能够在生命周期的所有阶段充分处理错误或不一致性。
- 要求确保人工智能系统对于针对操纵数据或算法本身的公开攻击和敏感攻击意图有防御能力，并且已经在一些案例中运用了该种缓解措施。

e) 人为监督

人为的监督有助于确保人工智能系统不会破坏人类自治性或造成其他不利影响。只有确保人类适当干预高风险的人工智能应用，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。

尽管本白皮书中所考虑的针对特定法律制度的人工智能应用都被认为是高风险的，但适当人类监督的类型和程度可能因情况而异。它应特别取决于系统的预期用途以及该用途对公民和法人可能产生的影响。人工智能系统在处理个人信息资料时，亦不得损害《通用数据保护条例》所确立的合法权益。例如，人类监督包括但不限于：

- 只有人工智能系统已经被一个人类提前审查和验证过，它才能输出结果（例如，只有人类才能拒绝一项社会福利申请）；
- 人工智能系统的结果立即生效的，要确保嗣后人类可以干预(例如，人工智能系统如果拒绝了一项信用卡申请，但这必须是嗣后人类可审核的);

[54] 根据 GDPR（2）f 部分，掌控者必须在获取个人数据时，必须向数据所有者提供进一步信息（确保数据自动处理决策是公平公开的）以及其他信息。

- 在运营过程中监控人工智能系统，以及实时干预和停用人工智能系统的能力(例如，当人类认为无人驾驶汽车操作不安全时，无人驾驶汽车上有一个停止按钮或程序可以使用);
- 在设计阶段就对人工智能系统运营进行限制(例如，在某些低能见度条件下，因为传感器可能变得不那么可靠，无人驾驶汽车应停止运行，或是在任何给定条件下与前方车辆保持一定距离)。

f) 远程生物特征识别的具体要求

收集和使用生物特征数据[55]进行远程身份识别[56]，如在公共场所部署面部识别，会对基本权利保障产生风险。使用远程生物识别人工智能系统对基本权利[57]的影响，可能因使用目的、环境和范围的不同而有很大差异。

欧盟数据保护规则在原则上禁止以单独识别自然人为目的处理生物特征数据，但在特定情况[58]下除外。具体来说，根据《通用数据保护条例》，这种处理只能在有限范围内进行，尤其是出于重大公共利益的考虑。在这种情况下，此数据处理必须根据欧盟或成员国法律进行，但须符合比例性的要求，数据保护权的实质内容及相关保障措施。根据《执法数据保护指令》，此类处理应当符合必要原则，在原则上必须得到欧盟或成员国法律的授权，以及适当的保障措施。由于任何以单独识别自然人为目的的生物特征数据处理都将涉及欧盟法律规定的禁令的例外情况，因此将受到《欧盟基本权利宪章》的约束。

因此，根据现行的欧盟数据保护规则和《基本权利宪章》，人工智能的使用只能以远程生物识别为目的，并且这种用途必须是正当的、适当的，并受到充分的保护。

为解决在公共场所使用的人工智能的潜在社会问题,并为避免内部市场的分裂,委员会将在特定情况下推出一个广泛的欧洲辩论,如果有的话,这可能会见证这样的使用,和其常见的保障措施。

E. 适用主体

[55] 生物特征数据被定义为"以自然人的物理、生理或行为为特征的经特定技术处理产生的个人数据，这些技术能处理准许或确认自然人的独特认证或识别，如面部图像或指纹数据"。(Law Enforcement Directive, Art. 3 (13); GDPR, Art. 4 (14); Regulation (EU) 2018/1725, Art. 3 (18).

[56] 在面部识别方面，识别意味着将一个人的面部图像模板与数据库中存储的许多其他模板进行比较，从而确定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)通常称为一对一匹配。它可以通过假定它们属于同一个体来比较两个生物特征模板，从而确定两个图像上显示的人是否是同一个人。例如，这种程序用于机场用于边境检查的自动边境控制(ABC)门。

[57] 比如人的尊严。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。不歧视儿童、老年人和残疾人等特殊群体以及对他们权利的保护也有潜在的影响。此外，言论、结社和集会自由不应因使用技术而受到损害。参见:面部识别技术:执法中的基本权利考虑，https://fra.europa.eu/en/publication/2019/facial-recognition.

[58]GDPR 第 9 条，《执行指令》第 10 条。另外参见欧盟 2018/1725(适用于欧盟机构和机构)规则第 10 条。.

关于上述所涉及的高危人工智能申请所适用的法律规定的适用主体，需要考虑两个主要问题。

第一个问题是如何在相关经营者之间进行分配。很多人员都参与了人工智能系统的生命周期。这些人包括了开发人员，运营人员（使用人工智能装备的产品或服务的人）以及其他潜在的人员（生产者，分销商或进口商，服务提供者，专业用户或私人用户）

委员会的观点认为，在未来的监管框架中，每一项义务都应该由最适合处理应对任何潜在风险的人员承担。例如，人工智能的开发者可能最适合处理开发阶段产生的风险，而他们对使用阶段所产生的风险的把控能力可能较为有限。在这种情况下，运营人员应是相关义务的主体。这并不影响当事人对最终用户是否应当赔付对最终用户或者其他遭受伤害并确保了有效的司法公正的当事人所造成的任何损害的问题。

根据欧盟产品责任法律，缺陷产品归责于生产商，不影响同样允许从其他方主张损害赔偿的国家法律。

第二个问题是关于法律的地理适用范围。按照委员会的观点，对于在欧盟中提供人工智能的产品或服务的所有相关经营者都是适用的，无论他们是否在欧盟设立住所。否则，前面提到的立法干预的目标无法充分实现。

F. 遵守和执行

为了确保人工智能是可靠、安全的，并与欧洲的价值观和规则相符合，适用的法律规则需要在实践中得到遵守，并由主管成员国国家和欧洲当局以及受影响的各方有效执行。主管当局的职责是调查个别案件，但也应能评估其对社会的影响。

鉴于某些人工智能应用对公民和我们的社会造成的高风险（见上文 A 节），委员会认为在现阶段有必要进行客观、事前的合规符合性评估，以核实和确保上述适用于高风险申请的某些强制性要求（见上文 D 节）得到遵守。它可以包括检查算法和开发阶段使用的数据集[59]。

针对高风险人工智能应用的合规符合性评估，应成为针对已经为欧盟单一内部市场上大量流通产品而存在的合规评估符合性评估机制的一部分。如果没有这样的现有机制可以依赖，可能就需要借鉴最佳实践、和利益相关者的实践成果及欧洲标准组织的可能意见以建立类似的机制。任何此类新机制都应是合比例的均衡和非歧视性的，并按照国际义务使用透明和客观的标准。

当依照事前合规符合性评估设计并运营执行一个系统时，应当特别考虑到下列各点：

- 并非所有上述要求都适合用事前合规符合性评估来验证。例如，关于应提供资料的要求一般不适合用这种评估来验证。
- 应该特别考虑到某些人工智能系统继续发展和机器从经验中学习的可能性，这可能需要在有关人工智能系统的整个生命周期中进行反复评估。

[59] 该系统基于欧盟合规评价程序（768/2008 EG 决议）、2019/881 欧盟条例（网络安全法律行动），并考虑人工智能运算的特殊性。见 2014 年转换欧盟产品法律指南（“蓝色指引”）。

- 有必要对于用于训练的数据集以及用于构建、测试和验证人工智能系统的相关编程和训练方法、流程和技术进行验证的需求。
- 假如合规符合性评估显示一个人工智能系统没有达到与训练数据集相关的要求，那么所发现的缺点就需要得到纠正，比如在欧盟内对该系统进行重新训练以确保其满足所有适用的要求。

对于符合要求的所有经营者，无论其住所在何处，都必须进行合规符合性评估[60]。为了减轻中小企业的负担，可以在数字创新中心框架内对为其进行支持设计一些支持组织，包括数字创新中心。此外，数据合规标准的和专用的在线工具可以促进规则对要求的遵守。

任何事前合规符合性评估都不应妨碍国家主管当局通过事中监管遵守情况和事后执法进行监管。这适用于高风险性的人工智能应用，但也适用于其他符合法律要求的人工智能应用，但是，所涉应用的高风险性质可能会成为是国家主管当局特别关注前者的理由。

通过审查相关人工智能应用的充分的文件记录，事后控制方得以实现（见上文 E 节），并在适当情况下允许第三方，如主管当局，对此类应用进行测试。在基本权利面临风险的情况下，测试这一点可能尤其重要，因为这些风险是基于环境随机而产生的。这种对合规性的监管应成为长期市场监管体系计划的一部分。下文 H 节将进一步讨论与治理有关的问题。

此外，对于高风险的人工智能应用和其他人工智能应用，都应确保受到人工智能系统负面影响的当事人得到有效的司法救济。与侵权责任相关的问题将在本白皮书附带的安全与责任框架报告中进一步讨论。

G. 自愿的无高风险人工智能人工智能应用无高风险的自愿标签

对于不符合"高风险"的人工智能应用（见上文 C 节），以及因此不受上述强制性规定（见上文 D、E 及 F 节）限制的应用，除了适用的法律立法之外，还可选择建立自愿标签机制。

根据该机制，不受强制性规定限制的有兴趣的经营者，可在自愿的基础上，决定使自己成为这些要求的规管对象，或成为一组为自愿计划而专门订定的类似规定的规管对象。有关经营者将因他们的人工智能应用而获发质量标签。

自愿标签将让有关的经营者发出信号，表明他们的人工智能产品和服务是值得信赖的。

它将让用户轻易识别地认识到，相关的产品和服务所满足的符合欧盟范围内的某些客观和标准化基准，已经超出了通常适用的法律义务范围。这将有助于增强用户对人工智能系统的信任，并促进该技术的全面应用。

这一选择将需要制定一个新的法律文件，为那些不被视为高风险的人工智能系统的开发人员和（或）运营部署人员确定自愿标签框架。虽然标签机制是自愿参与的，但一旦开发人员或运营部署人员选择使用标签，这些要求便具有约束力。结合事前和事后执法的结合将需要确保所有要求都能得到遵守。

[60] 关于所涉治理结构，包括合规实施机构，见 H 节。

H. 治理

为了避免管辖分散责任碎片化，提升成员国的治理能力和确保欧洲具备测试和认证人工智能的产品和服务所需的能力，建立基于成员国国家主管部门合作框架的人工智能欧洲治理框架是非常有必要的。实现上述这方面的要求，将有助于国家主管当局确保在人工智能应用领域的地方履行自己的责任。

欧洲治理框架可以有多项任务，作为一个定期交流信息和最佳实践的论坛，它可以识别新兴趋势并对就其标准化活动和认证的工作提供建议。它还应通过提供指导、意见和专门知识等方式，在促进法律框架的实施方面发挥关键作用。为此，它应该依靠国家当局网络，以及国家和欧盟层面的行业部门网络和监管当局。此外，一个专家委员会可以向委员会提供援助。

治理框架应该保证利益相关者最大程度的参与。在实施和进一步发展该框架时，应咨询消费者组织、社会团体、商界、研究人员和民间社会组织等相关人士的意见。

鉴于金融、制药、航空、医疗器械、消费者保护、数据保护等行业领域的现有情况，拟议的治理框架不应与现有监管职能相重叠。相反，它应该与欧盟其他部门和各国主管部门建立密切联系，以对现有专业领域进行补充，并帮助现有部门监测和监督涉及人工智能系统和支持人工智能的产品和服务的经营者的活动。

最后，如果采用了这种方法，就可以将执行合规符合性评估移转委托给会员国指定的公告机构。测试中心应根据上述要求对人工智能系统进行独立审计和评估。独立评估将增加信赖并确保客观性。它也可以促进有关主管当局开展工作。

欧盟已经拥有优秀的检测和评估中心，也应该发展其在人工智能领域的能力。以希望进入国内市场的第三国为设立地点的经营者可以利用在欧盟设立的指定机构，或者可以在与第三国达成相互承认协议的情况下，求助于指定进行这种评估的第三国机构。

与人工智能有关的治理框架和可能进行的一致性评估不会使现行欧盟法律下的相关主管机构的权力和责任在特定部门或具体问题上（金融、医药、航空、医疗器械、消费者保护、数据保护等）受到影响。

6. 结论

人工智能是一种战略技术，只要它是以人为本的、合乎伦理的、可持续发展的，并尊重基本权利和价值观，它就能为公民、企业和整个社会带来很多好处，人工智能显著提高了效率和生产率，这可以增强欧洲工业的竞争力，改善公民的福祉。它还有助于帮我们找到一些最紧迫的社会问题的解决方法，包括与气候变化和环境恶化作斗争，应对可持续性和人口变化有关的挑战，以及保护我们的民主，在必要时和适当时打击犯罪。

欧洲要充分抓住人工智能带来的机遇，就必须发展和加强必要的工业和技术能力。正如所附的《欧洲数据战略》（European strategy for data）所述，这还需要采取措施，使欧盟能够成为全球数据中心。

欧洲发展人工智能的方法旨在促进欧洲在人工智能领域的创新能力，同时支持在整个欧盟经济中发展和采用道德的和可信的人工智能。人工智能应该为人民服务，成为推动社会公益的力量。

有了这份白皮书和附带的关于安全与责任框架的报告，委员会开展了成员国、民间社会、工业界和学术界的广泛磋商，就欧洲发展人工智能方法的具体建议展开讨论。这些建议包括促进对研究和创新进行投资的政策手段、加强技能开发和支持中小企业采用人工智能，以及未来监管框架的关键要素。这次协商将使我们能够与所有相关方开展全面对话，这将为委员会今后的发展指明方向。

欧洲委员会欢迎各界人士通过公开咨询，就白皮书所载的建议提出意见，网址为https://ec.europa.eu/info/consultations_en。征求意见稿开放至2020年5月19日。

在公众咨询中，委员会一般会公布所收到的意见书。但是，公众可以要求提交的文件或其部分内容保密。如果需要的话，请在你提交的材料的首页清楚地说明不应公开，并将你提交的材料的非机密版本送交委员会用于公开。