

[输入关键字](#)[专栏首页](#)[文章](#)[专辑](#)[作家](#)[站长专栏](#)[王鹏 >> 正文](#)

超级人工智能：合作共赢，规范监管

王 鹏 2023年06月16日



王鹏

北京社科院研究员，北京习近平新时代中国特色社会主义思想研究中心特约研究员
工学院数字经济研究院院长、特聘教授

6月9日-10日，AI内行顶级盛会“北京智源大会”在北京举办。智源大会是智源研究院主办的年度国际性人工智能高端专业交流活动，业界学界共200余位人工智能顶尖专家到场。ChatGPT让公众直观领受到AI的力量，它展现出强大的翻译、理解、推理能力，在多项考试中超过人类平均水平，在带来AGI（通用人工智能，像人一样思考、从事多种用途的人工智能）曙光的同时，也带来大家对AI侵犯隐私、导致失业，甚至取代人类等问题的担忧。

一、超级AI所带来的近忧和远虑

近忧1：侵犯个人隐私

目前来说，人工智能至少可通过三种方式侵犯隐私，即非法监控、非法侵入个人空间或住宅以及非法干扰个人行为。

1.非法监控

监控不仅包括对物理空间的监控，也包括对虚拟空间的监控。虚拟空间是指网络空间，它独立于物理空间而存在。尽管非法监控与非法侵入个人空间或住宅存在交叉，但不论是对物理空间还是对网络空间的监控，非法监控所涉及的隐私领域主要是信息隐私，而非法侵入个人空间或住宅主要涉及空间隐私。在非法监控中，对物理空间或网络空间的侵入只是手段，终极目的在于获取未公开的个人信息。

2.非法侵入个人空间或住宅

这里所指的个人空间仅限于实体空间，并不包括网络空间。传统的侵入个人空间或住宅的方法比较有限，如监视、窃听、刺探、秘密进入等，这些方法都需要人力来完成。然而，人工智能可以通过侵入住宅内的监控摄像头、通讯网络设备、智能家居系统等方式侵扰个人空间或住宅，且不下任何外部物迹。

3. 非法干扰个人行为

众所周知，互联网上存在大量的实名或未实名的用户数据，人工智能通过一定的技术手段对用户数据进行收集与分析，可以准确分析出用户的兴趣爱好、行为踪迹等信息，并进行信息推荐，而这些信息用户并未主动提供，显而易见，用户是根本不愿意这些信息被其他主体知晓，这势必侵犯个人隐私。

近忧2：造成失业与垄断

从失业的角度来说，主要是因为超级人工智能的出现造成了劳动力的供大于求。人工智能的普及必然取代社会中原有的部分劳动力，且由于人工智能的低成本、高效率等突出特性，更易受到用人单位的青睐，造成以出卖劳动力为主的社会底层居民丧失劳动和就业机会，极大压缩其生存的空间，使其成为社会中的不稳定因素，进而引发不同群体间的对立和冲突。从形成行业垄断的角度来看，在市场中，资本与技术结合共谋实施行业垄断是时常发生的情形，由于掌握先进技术，部分企业可以借此获得垄断或支配地位，肆意打压中小企业或歧视性地对待合作商家而谋取巨额利益，进而严重破坏市场公平公正的竞争秩序。

远虑：存在毁灭性风险

总部位于旧金山的人工智能安全非营利组织（CAIS），一直在为“如何降低人工智能（AI）的社会规模风险”而努力。5月30日，该机构发布了一份《AI风险声明》，呼吁全世界应该像控制大流行病和核战争那样，降低AI带给人类灭顶之灾的风险，并且把优先级调到最高。在ChatGPT走红全球之后，75岁的辛顿从谷歌辞职，便拉响AI的警报，更自由地讨论AI给人类带来的风险。当时，他出席美国麻省理工大学

(MIT)的一场论坛时表示:“如果AI模型比人类聪明得多,它将善于操纵人类,你都不会意识到发生了什么。”在最坏情况下,“人类只是智能进化的一个过场”,在生物智能进化出数字智能后,数字智能将吸收人类创造的一切,并开始直接接触世界。当AI变得比人类更聪明时,就有可能导致人类面临生存威胁。

二、强化对超级人工智能监管的必要性分析

细化法律法规,监管技术齐驱

尽管新一代AI技术提出了诸多挑战,具有较大的监管难度,但是我们却不能听之任之和置之不理。如果任由ChatGPT等新一代AI技术如脱缰的野马一样横冲直撞,那么就可能出现难以挽回的损失甚至是灾难性后果。只有未雨绸缪地对新一代AI技术的发展趋势和应用前景进行研判和分析,提前做好监管准备,才能避免“兵来将挡水来土掩”式的被动响应。当我们建立和完善一揽子的新一代AI技术监管制度体系后,监管创新便能跟上技术进步的步伐,避免监管被技术甩在后面。否则,监管技术的落后将直接决定技术监管的失败,也必然会导致新一代AI技术的失控和无序发展。

新一代AI技术具有很强的不确定性,使政府监管难以下手。首先,新一代AI技术的高度不确定性意味着一般的政府部门难以预判其风险,政府部门的监管往往是针对既有问题而展开的,而新兴技术提出的监管挑战则可能不在其监管框架之内。在这种监管机制存在局限的情况下,政府部门对新兴技术的监管经常陷入被动响应的尴尬局面。其次,当新一代AI技术应用于各个领域时,它会成为新的操作系统、数字基础设施或技术平台,使其同相关业务难以切割。这使政府部门的监管不得不通盘考虑技术应用的广泛性和深入性,而不能采取一技一策的方式进行监管。而在这种情况下,由专门部门去细化相关的责任显得尤为必要。

加强国际合作,共度AI深秋

在全球化时代,由于人工智能技术呈现出跨国家、跨文化和跨领域特点,应对人工智能技术挑战并非完全取决于个别的国家日,因此加强国际间的相互合作也就成为人工智能监管的必然趋势。这种国际间的相互合作应该主要表现为两方面:一方面是主权国家之间的相互协调,以美国为主的部分西方国家以所谓的“意识形态”“共同价值观”为合作纽带,以应对中国在人工智能领域不断提升的影响,确保其在该领域的领导地位。如美国不仅计划与欧盟在跨大西洋合作方面积极展开关于人工智能技术发展、规制的讨论与合作,还在与韩国来往时多次强调希望促进韩美两国在人工智能等新兴技术方面的合作。另一方面是产业界、学术界等形成的行业宣言,如在首届全球数字经济大会人工智能产业治理论坛上发布的《人工智能产业担当宣言》,便是企业与学术机构在共同商讨下形成的倡议性文件。

三、加强国际合作具体规范设想

AGI拥有彻底改造文明的能力,因此,有意义的国际合作非常必要。作为地球上的一员,我们需要投资AGI安全、谨慎协调才能实现并享受这个未来。作为一项全球性的技术,轻率合作带来的后果,将影响所有人。因此,在国际合作中,这三个关键领域最为重要。

建立国际规范,注重过程包容

我国应注重运用在人工智能相关领域的技术优势,积极参与人工智能技术相关国际标准的制定工作,包括规范术语、技术框架、产品服务,逐步实现我国在人工智能技术标准方面的国际化。同时,鉴于人工智能国际标准在国际上具有一定敏感性,在标准化立项和推进过程中要秉承积极协商、灵活应对、推进共识的工作策略,保障国际标准工作的顺利推进。在推动共同国际规范和标准建立的过程中,应从相互依存、命运与共的“共同体”意识出发,同时注意各国之间的差异性,在制定规制的过程中,留下一定变通的空间。

加强人才互通,建立交流机制

千里之行,始于足下。我认为,最有建设性的第一步,是国际科学界、技术界展开合作。我们应该在AGI安全领域,推动建立增加透明度和知识共享的机制。发现紧急情况的研究人员应该为了共同的利益,共享新进展。另外,我们也应仔细考虑,如何在尊重和保护知识产权的同时鼓励这种做法。更广泛地说,我们应该推动、引导AI对齐和安全研究。

同时,我国应紧密跟踪研究国内外人工智能安全标准化工作动态和发展趋势,把握人工智能安全领域国际标准研究态势和重点方向。建立人工智能重点难点标准合作交流机制,借鉴各国在国际标准组织中的工作经验,参考形成我国人工智能安全国际标准化工作推进策略。充分吸收国外人工智能安全领域标准研究成果,借助国际力量丰富和推进我国人工智能安全标准制定,保持国家/行业标准与国际标准之间的协调一致,提高我国标准的先进性、规范性,确保人工智能安全标准能够及时满足人工智能安全保障和新兴技术产业发展的需求。借助于双边或多边的合作契机,邀请人工智能相关领域专家到我国开展交流研究,不断向世界输出人类命运共同体的合作理念和价值,搭建校企之间、科研院所之间等广泛的交流合作机制,推动构建人工智能国际监管的合作框架,为人工智能监管贡献中国方案,发出中国声音。



[专栏首页](#) [文章](#) [作家](#)

友情链接: [中国日报网](#) [新华网](#) [人民网](#) [中国网](#)

[关于我们](#) | [联系我们](#)

互联网举报中心 防范网络诈骗 法律顾问: 炜衡律师 视听节目许可证0108263 京公网安备110105000081号 京网文[2011]0283-097号