# 中美在“人工智能 X 生物安全”领域的讨论与措施及可能达成的共识

人工智能将生物技术带入一个全新时代，一方面给生命科学带来突破性发展的巨大潜力，另一方面又带来值得全球警惕的重大风险，包括危险合成病原体的研制、生物武器效能的强化等技术滥用或意外后果。为防范这些风险，美国人工智能企业已投入大量资金来构建“人工智能 X 生物安全”防护体系，同时因缺乏了解，他们又对中国在该领域的努力持怀疑和不满的态度。本文旨在梳理中美两国政府、企业、学界对“人工智能 X 生物安全”的讨论、建议与做法，并进行比较分析，最后提出两国有望在更广泛的人工智能安全领域达成的潜在共识。

## 一、“人工智能 X 生物安全”议题在美国

随着人工智能的迅猛发展，人工智能可能被恶意滥用的风险日益受到关注。“人工智能 x 生物安全”这一交叉领域就是可能会给人类带来灭顶之灾的风险源之一，近年来该议题在美国人工智能界引发了广泛担忧和深入讨论。业界认为，人工智能给生命科学领域带来的显著益处主要表现在两方面，一是可帮助人类研制新药和新疫苗，有助于促进公众健康；二是可加强迅速发现新传染病的能力，尽早阻止疫情传播。与此同时，人工智能也有可能导致有害生物制剂被故意或意外释放，从而引发全球性生物灾难。最大的风险体现在，一是设计、制造和部署危险生物制剂所需的知识和技术支持更容易获取，从而降低恶意行为者利用生物技术造成伤害的门槛；二是提升恶意行为者利用生物技术造成潜在危害的“可能性上限”。此外，AI 自主性的加强，可与实验室设备和机器人互动并操纵机器，若无监管，不知道人工智能会做出什么来。而且，被恶意使用的人工智能还有可能降低生物安全防护措施的效果。

美国人工智能业界大佬纷纷警告人工智能可能会被用于制造生物武器，并呼吁尽快制定安全规则。2022 年 9 月，前谷歌高管埃里克•施密特（Eric Schmidt）警告称，“人工智能的最大问题是会被用于生物战”。次年 7 月，Anthropic 公司 CEO 达里奥•阿莫迪（Dario Amodei）在国会参议院听证会上提醒，再过两三年，人工智能技术将有能力帮助不具备专业能力但有恶意的行为体制造生物武器。同年，OpenAI 的 CEO 萨姆•奥尔特曼（Sam Altman）呼吁对有助于创造新型生物制剂的人工智能模型制定管理规则。

美国政府已发布政策文件，要求关注人工智能在生物领域的潜在风险。2023 年 10 月，拜

登总统发布《关于安全、可靠和值得信任的人工智能开发与使用的行政命令》，提出人工智能必须确保安全，要求应对 AI 系统最为迫切的安全风险——包括生物安全风险。次年 10 月，白宫发布《关于推进美国在人工智能领域领导地位、利用人工智能实现国家安全目标、并促进人工智能安全性、保障性和可信赖性的备忘录》，要求商务部对与生物安全相关的风险进行评估。2025 年 4 月，美国国会新兴生物技术国家安全委员会发布《生物技术未来蓝图》，该报告虽然主要阐述如何赢得大国新兴生物技术竞争，但也强调“人工智能 x 生物技术”被误用、滥用可能带来的风险，并提出应与包括中国在内的国际社会共享生物安全的最佳做法和标准。

美国学界和智库界也已对“人工智能 x 生物安全”提出预警并进行了深入研究。2023 年 11 月 6 日，哈佛大学肯尼迪政府学院贝尔弗科学与国际事务中心发布政策简报《AI 时代的生物安全：有什么风险？》，提出“人工智能 x 生物安全”问题需要进一步研究，并建立新的检测手段和早期预警机制。同年 12 月，美国科学家联合会就应对这一前沿风险提出一系列政策建议，包括加强对生物设施工具研发工作的监管、建立评估人工智能时代生物风险的标准、确立负责任的研发准则等。2024 年 9 月，约翰霍普金斯大学、斯坦福大学、福特汉姆大学等高校的多名健康安全、医学、法学学者在《科学》期刊上联合发表《人工智能与生物安全需要治理》呼吁各国政府对先进 AI 模型进行评估，考虑是否需要采取安全措施。NTI 于 2024 年 11 月发布报告《为人工智能生物设计工具制定护栏》，接着又在次年 7 月发布《人工智能与生命科学交汇处的生物安全风险声明》，共有来自全球各地的 41 名专家联署。2025 年 3 月，美国国家科学院、国家工程院和国家医学院发布报告《生命科学的人工智能时代：收益与生物安全考量》，提出一系列建议，旨在帮助美国在生物技术领域利用 AI 的优势，同时将 AI 被滥用于开发有害生物制剂的风险降至最低。同年 8 月，2025 年 8 月，CSIS 发布《利用人工智能引发的生物恐怖主义加强美国生物安全的机遇：政策制定者应知事项》，报告探讨了人工智能可能被用于生物恐怖主义的风险，并提出美国决策者可采取的强化国家生物安全的关键措施。美国智库界指出，必须防范 AI 技术导致全球大流行病灾难，以及对国家安全、经济安全和公共卫生安全构成威胁。如何深化对这些潜在风险的共同认知，提高对最佳实践做法和有效措施的认识，以避免可能造成全球影响的极端后果，各国及民间社会应达成一定的共识。在这些方面，各国政府尤其担负着独特使命，必须在推动 AI 与生物技术创新发展的同时，建立健全防护机制，防控重大灾难性风险，确保 AI 赋能的变革性生物技术以负责任的方式研发应用，造福全人类。

美国人工智能相关企业已经开始采取一些早期安全措施，来防范人工智能的生物风险，包括开展 AI 生物安全评估、构建探测和预警系统、部署多层次安全策略、提升整体风险管理框

架、加强外部合作与行业引导等。例如，Ginkgo Bioworks 开发了 AI 驱动的“工程化核苷酸的检测与排序”（ENDAR）工具，用以识别是否存在人工工程痕迹的基因组。Anthropic 已于 2023 年 7 月开展“红队测试”（red-teaming），评估 AI 模型可能带来的生物安全风险。OpenAI 建立了一套风险管理制度，将 AI 模型的新能力按照生物、化学、放射性与核等类别分级评估。最新“GPT-5 thinking”模型被评定为在生物及化学领域的能力已达到“高风险”，为了充分降低相关风险，OpenAI 已导入强有力的防护机制。OpenAI 还建立了“智能体生物安全漏洞赏金计划”，邀请在人工智能红队测试、安全防护或化学与生物风险领域有经验的研究人员，尝试寻找能够突破其十级生物/化学挑战的通用越狱方法。

## 二、“人工智能 X 生物安全”议题在中国

中国在大力发展人工智能技术的过程中，非常重视人工智能安全问题。在“人工智能 x 生物安全”这一细分领域，中国虽尚未出台官方政策文件，但相关原则在有关人工智能安全和生物技术发展的指导性文件中均有体现。例如，《“十四五”生物经济发展规划》（2022 年）指出要关注“新型生物安全风险”。全国网络安全标准化技术委员会于 2024 年 2 月发布的《生成式人工智能服务安全基本要求》规定，生成式人工智能服务提供者应“重点关注生成式人工智能可能被用于编写恶意软件、制造生物武器或化学武器等安全风险”。该委员会于 2024 年 9 月发布的《人工智能安全治理框架 1.0》，将“核生化导”一并讨论，指出人工智能会“极大降低非专家设计、合成、获取、使用核生化导武器的门槛”，应“提高人工智能系统最终用途追溯能力，防止被用于核生化导等大规模杀伤性武器制造等高危场景”，包括“对训练数据进行严格筛选，确保不包含核生化导武器等高危领域敏感数据”。

中国学界和智库界对“人工智能 x 生物安全”的研究也在不断深入。中国疾病预防控制中心、中科院、军事科学院军事医学研究院，国务院发展研究中心等科研机构近年来发表了一系列分析人工智能与生物技术融合风险的论文。上文提到的由全球数十名专家联署的 NTI《关于人工智能与生命科学交叉领域生物安全风险的声明》，至少有 3 名在华工作的中国学者参与。安远 AI 和天津大学生物安全战略研究中心在 2025 世界人工智能大会上发布的《人工智能 x 生命科学的负责任创新》，是中国第一份聚焦“人工智能 x 生物安全”的智库报告。该报告认识到“人工智能的快速发展正在深刻重塑生命科学研究范式，为人类健康、经济发展和生态可持续性带来前所未有的机遇”，但同时也指出，“AI 的赋能也引发了有关生物安全的新兴挑战”，必须在推动技术融合和价值创造的同时，妥善应对潜在风险。上海人工智能实验室与安远 AI 联合发布的《前沿人工智能风险管理框架》，对生物安全也作了专门阐述。

与美国人工智能企业相比，中国人工智能企业虽然也非常重视生物安全，但相对比较低调，公开发声较少。例如，DeepSeek 在训练过程中采用了严格的内容过滤机制，确保模型不会生成涉及生物安全风险的有害信息；遵循伦理准则，拒绝提供可能被恶意利用的敏感信息，包括涉及生物实验安全、危险技术等内容，确保严格遵守国内外相关法律法规。阿里巴巴的 Qwen 大语言模型也采取了类似的防范机制。

不过，总体而言，中国政府、学界、企业对“人工智能 x 生物安全”领域的讨论和机制化成果仍少于美国。原因可能有以下几方面：

一是中国对“人工智能 x 生物安全”的讨论比美国起步晚，相关研究仍在进行中，现有成果尚未转化为具体政策或规则。中国式的思维是从整体到局部。过去几年，中国主要从如何建立整体的人工智能安全治理规范来思考人工智能安全问题。中国官方已先后发布《新一代人工智能治理原则——发展负责任的人工智能》（2019）、《人工智能安全标准化白皮书》（2019）、《新一代人工智能伦理规范》（2021）、《网络安全标准实践指南—人工智能伦理安全风险防范指引》（2021）、《生成式人工智能服务管理暂行办法》（2023），人工智能产业界发布《中国人工智能安全承诺框架》（2025）。这些文件都认识到人工智能带来的各种安全挑战，但没有专门提到“生物安全”。

“人工智能 x 生物安全”这类交叉领域研究，需要人工智能界与生物安全界通力合作。如上文提到的安远 AI 和天津大学生物安全战略研究中心的合作成果。在美国，该领域的研究主要由人工智能企业和智库主导；而在中国，似乎是生物学界和医学界对该领域关注更多。天津大学生物安全战略研究中心的一位专家在 2023 年表示，人工智能可能诱发更多生物安保风险，如加大生物安全风险概率、提升生物安全风险危害程度等，人工智能聊天机器人也可能被恶意利用，成为增强已知病毒毒性的工具。“人工智能 x 生物安全”领域的相关论文也更多是由生物学界和医学界的研究人员所撰写。

二是不同人工智能企业所关注的安全风险不同。对于开展基础研究和模型构建的企业来说，人工智能可能带来的安全风险林林种种，生物安全只是其中之一，“人工智能 x 生物安全”只是其中一个非常小众的领域。这类企业在探讨人工智能安全风险管理时，更多关注整体的安全风险识别与防范，但也把生物安全风险列为需要识别的具体“红线”之一。对于人工智能应用企业来说，若其应用业务不涉及生物相关领域，就没有考虑生物安全的需求。他们更多考虑人工智能可能带来的伦理、价值观、个人隐私方面的安全。因此，当中国人工智能界宣布要通过产业自律来促进人工智能稳健发展时，所做的承诺是人工智能安全防范和治理的一般性保证，

而未涉及生物安全领域这样的细分领域。对于专门关注人工智能安全的公司，由于目前没有合适的商业盈利模式，因此这类企业也没有对生物安全领域有太大投入。

三是中国在传统上更关注近期风险研究。总体而言，中国学者在全球中远期风险预测方面的成果较少，更多关注针对中国自身的近中期风险研究；而美国在中远期风险预测方面有着更完善的预测研究体系，由国家情报机构、军方、智库、学术机构等多方共同推动。“人工智能 x 生物安全”通常被视为中远期风险，当前人工智能在生物安全领域的直接风险还比较有限，这类风险已有一些研究案例出现，但距离大规模、现实的生物安全事故还有技术与合规屏障。因此，中国研究界对“人工智能 x 生物安全”这类中远期风险的关注也不如美国多。

鉴于以上，对于人工智能在生物合成领域滥用可能带来的全球生存威胁，中美及全球的人工智能研究界、生物学界、医学界等必须加强跨国、跨学科的交流与合作，增进相互了解，逐步建立信任，相互学习、相互借鉴，共同应对人工智能在生物安全领域带来的挑战。

## 三、中美两国领导人有望在人工智能安全领域达成的共识

鉴于人工智能的迅速发展给包括生物安全在内的各个领域都带来显著的安全挑战，中美两国应加强在人工智能安全领域的合作，并有望就以下方面达成一定共识：

1. 中美应在人工智能安全领域做负责任的大国。
2. 中美应针对人工智能安全问题建立常态华沟通机制，鼓励包括中美在内的研究人员互相交流，逐步建立互信。
3. 在人工智能安全等全球性挑战上，有必要加强国际合作，通过合作实现共赢；通过合作防止人工智能对人类带来的负面影响。
4. 在基本伦理、价值观方面达成一定共识，例如维护人类尊严、确保对人工智能的控制等。
5. 人类应维持对核武器/战略武器/大规模杀伤性的控制。
6. 促进研发人工智能向善技术，合作防止先进人工智能和未来超级智能给人类带来的生存性威胁，包括：针对先进人工智能模型采取防护措施，加强“人类控制”。
7. 联合防止恐怖分子或非国家行为体利用先进人工智能模型来制造核生化威胁。