北京大学
PEKING UNIVERSITY

Deepseek内部研讨系列

DeepSeek应用场景中需要关注的十个安全问题和防范措施

AI肖睿团队

(傅峥、王俊鑫、秦天、李娜、谢安明、陈钟)

2025年2月26日

• 北大青鸟人工智能研究院
• 北大计算机学院

摘要

1. 本次讲座为DeepSeek原理和应用系列研讨的讲座之一。随着人工智能技术的快速发展，DeepSeek作为前沿的AI平台，其安全性和可靠性成为关注焦点。本讲座探讨了DeepSeek在实际应用中面临的安全挑战，为不同类型的用户揭示使用DeepSeek的潜在风险并提出防范策略。
2. 本讲座的内容分为四个主要部分：

   ① 首先，介绍了DeepSeek安全问题具有威胁难以预测、攻防非对称的特点，以及存在数据隐私、知识产权、责任归属、伦理道德等法律问题；并从内生安全和外延安全描述了安全方案框架，帮助大家对DeepSeek的安全建立**整体认知**。

   ② 其次，帮助大家理解DeepSeek**模型自身的5个安全问题**，包括DDoS攻击、无限推理攻击、漏洞探测与利用、投毒问题和越狱问题，还演示了漏洞探测与利用的过程，让大家形象的理解DeepSeek被黑客利用的过程。

   ③ 再次，帮助企业用户和技术爱好者说明了DeepSeek**私有化部署的2个安全问题**，包括DeepSeek本地化部署工具的风险、针对DeepSeek本地化部署实施网络攻击的风险。并演示了从利用部署工具的漏洞到最终获取模型服务器管理权的全过程，帮助大家能安全的使用私有化部署。

   ④ 最后，为普通用户介绍DeepSeek**外延的3个安全问题**，包括仿冒DeepSeek官方APP植入木马、仿冒DeepSeek官方网站和域名收集用户信息、DeepSeek辅助实施攻击。还演示了仿冒网站收集用户信息、DeepSeek辅助渗透攻击的方法。让普通用户在使用DeepSeek的时候了解常见的防范措施。
3. 在技术学习的道路上，优质学习资源至关重要。推荐大家参考《人工智能通识教程（微课版）》这本系统全面的入门教材，结合B站“思睿观通”栏目的配套视频进行学习。此外，欢迎加入ai.kgc.cn社区，以及“AI肖睿团队”的视频号和微信号，与志同道合的AI爱好者交流经验、分享心得。

目录
CONTENTS

PART 01

DeepSeek安全问题的特点及目前的安全方案框架

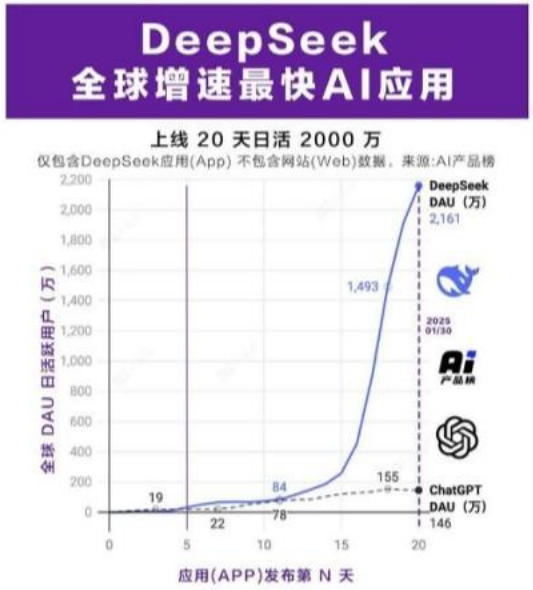
DeepSeek快速出圈，硬控全中国

北京大学 PEKING UNIVERSITY

deepseek

2025年2月17日，中共中央在北京召开民营企业座谈会，DeepSeek创始人梁文锋出席了会议。这不仅是对DeepSeek的认可，也意味着国家对民营企业AI创新、AI应用的鼓励和高度重视。

2025年1月20日下午，中共中央政治局常委、国务院总理李强主持召开专家、企业家和教科文卫体等领域代表座谈会，听取对《政府工作报告（征求意见稿）》的意见建议。DeepSeek公司创始人梁文峰作为企业家代表之一参加了此次座谈会。



DeepSeek安全问题的特点

安全问题特点1：威胁难以预测

传统技术工具与人工智能技术特性迥异。

传统技术工具即便复杂，其设计原理和运行逻辑也能被人掌握，行为可预测。

传统技术工具的逻辑相对固定
行为可预测、容易控制