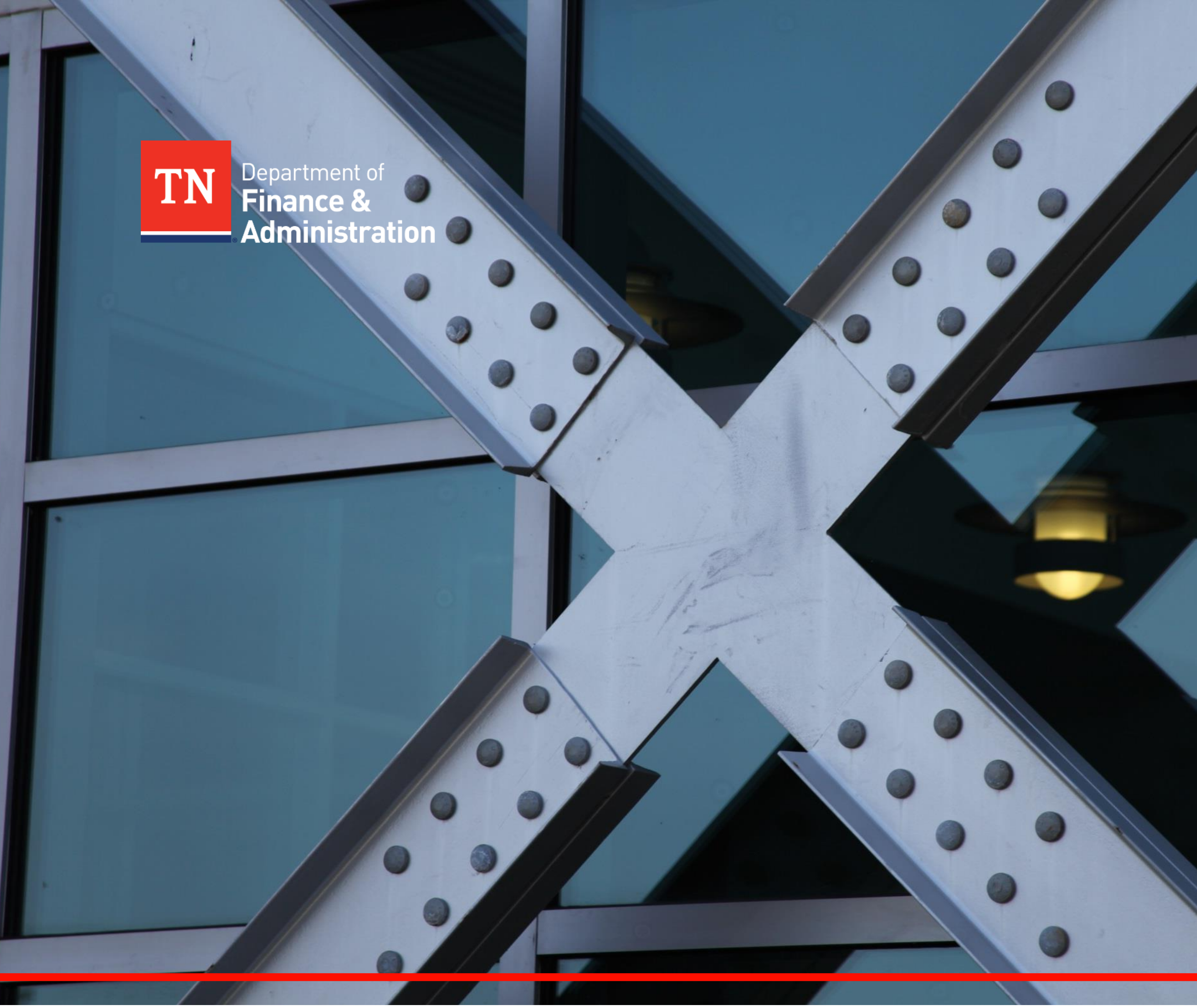

TN Department of Finance & Administration

TN AI ADVISORY COUNCIL | DEEPSEEK AI

Security Assessment

Tennessee Department of Finance & Administration | **March 2025**

Executive Summary

DeepSeek AI was first addressed by the Tennessee State Artificial Intelligence Advisory Council on January 29, 2025, mere days after the release of their open-source R1 model. Due to its high effectiveness and very low cost of development and use, DeepSeek caused a splash in the AI and microchip industries and resulted in more than $1 *trillion* loss in market capital. This tech crash led to a flurry of research and reporting on exactly how this low-cost startup was able to compete with their multi-billion-dollar mainstream competitors.

Studies of DeepSeek AI and similar products demonstrate that yes, it possible to create working AI at low cost; by side-stepping laws, training using established companies' products, and forgoing industry-standard cybersecurity standards and best-practices. The following assessment of DeepSeek is centered on the R1 model but can be applied to most of their other products and other models that were similarly developed.

Strategic Technology Solutions (STS) moved quickly, blocking access to DeepSeek AI from government devices, messaging state employees about its dangers, and the State Attorney General released a warning to citizens.

From February 15, 2025 to March 25, 2025 an assessment of DeepSeek AI was conducted in respect to the vulnerabilities that would potentially affect its use on state government IT networks and equipment. Due to the cybersecurity, harmful content and extremist, harmful language, and chemical and biological risks*, it is the recommendation of the Special Subcommittee that for the safety of state employees, citizens, and the network, DeepSeek AI and related models continue to be banned for government use*.

Use of DeepSeek AI or derivative tools by the Tennessee State Government would pose numerous critical risks to operations. These risks range from the unauthorized disclosure of privileged information to infection of devices by malware. Data could be intercepted by bad actor in transit, directly from the AI servers, or inadvertently shared with 3rd-parties by the AI itself. The use of inaccurate or harmful output produced by DeepSeek could also result in substantial financial or reputational impact.

DeepSeek AI Timeline

July 13th 2023	DeepSeek founded by Liang Wenfeng
December 2nd 2023	DeepSeek LLM released
February 1st 2024	DeepSeek Coder released
March 13th 2024	DeepSeek-VL released
December 28 2024	DeepSeek V3 launched
January 20th 2025	DeepSeek releases open-source R1 model
January 27 2025	DeepSeek becomes top app in U.S.
January 28th 2025	U.S. Commerce Department launches investigation into DeepSeek's chip sourcing
January 29th 2025	Quarterly TN State AI Advisory Council meets and discusses blocking App from government devices
January 30th 2025	STS blocks DeepSeek access from government devices and networks
January 31 2025	Italy bans DeepSeek access across country
February 3rd-Present	Taiwan, Australia, South Korea, U.S. Navy, NASA, Pentagon, Texas, Arkansas, Florida, Georgia, Iowa, and U.S. Congress bans DeepSeek from government devices
March 5th 2025	DeepSeek warning email to state employees and Tennessee Attorney General press release warning citizens are distributed

Bottom Line and Recommendations

APPLICATION has **MANY** serious security flaws.

Recommended actions:

- Continue blocking DeepSeek AI on Government networks, computers, and phones.
- Add all known DeepSeek AI variants to dis-allow list.
- Regularly monitor for and add additional DeepSeek AI variants to dis-allow list.
- Periodically reinforce risks of unapproved AI use during scheduled IT training.

Key Findings

Cybersecurity Risks – The DeepSeek AI models, particularly it R1 and V3 iterations, pose significant cybersecurity risks due to critical vulnerabilities in their design, data practices, and operational security. These flaws have drawn widespread concern from cybersecurity experts and government agencies, with the U.S. Navy and several states banning its use on government devices over security and ethical concerns.

Core Security Vulnerabilities

- Jailbreaking susceptibility: DeepSeek R1 failed to block *any* harmful prompts in Cisco's evaluation, showing 100% attack success rate compared to GPT-40's 14% attack success rate. Techniques like Crescendo and Evil Jailbreak can override safety mechanisms to generate malware guides, phishing content, and misinformation [3] [4].
- Weak encryption & SQL vulnerabilities: The Android app uses hardcoded encryption keys and outdated algorithms, enabling potential data decryption. SQL injection flaws allow database manipulation [5] [6].
- Open-source risks: The model's unrestricted accessibility lets attackers modify safety protocols, creating customized attack tools [3].

Data Privacy Concerns

- Chinese data storage: User inputs, keystroke patterns, and device metadata are sent to Chinese servers, potentially accessible to state-linked entities [2] [5]. Chinese law requires companies to share any data requested by the government. This conflicts with GDPR (General Data Protection Regulation) compliance and enables surveillance risks [1] [2]. GDPR is an EU regulation that is a good baseline standard and a requirement for U.S. entities dealing with EU citizen's data.

- Massive exposures: Over 1 million lines of sensitive data—including API keys and chat histories—were found in an unprotected database [7] [8]. This ClickHouse database, typically used for server analytics, was internet-facing and not password protected, allowing researchers from Wiz complete access to its information. While they were legally and ethically required to cease their explorations at this point, malicious actors could and may have established a persistent presence before the vulnerability was removed.

Operational Security Failures

- Inadequate safeguards: DeepSeek lacks basic protections against:
    - Malware generation: Can produce functional ransomware and keyloggers with simple prompts [3] [9]
    - Hallucinations: Higher rates of inaccurate outputs compared to competitors [8]
    - Credential-based attacks: Generates scripts for purchasing stolen credentials [7]
- Anti-analysis measures: Built-in mechanisms obstruct security researchers from evaluating risks [6].

Geopolitical & Criminal Exploitation Risks

- State-aligned threats: Data transfers to China enable potential espionage and influence operations [1] [3]. Embedded ByteDance code suggests undisclosed data sharing [6].
- Criminal efficiency: Lowers technical barriers for cyberattacks by:
    - Automating financial fraud bypass systems [3]
    - Reducing malware development time by 83% compared to manual coding [9]
    - Enabling large-scale phishing campaigns through prompt manipulation [3] [7]

Comparative Risk Analysis

Risk Factor	DeepSeek R1	GPT-40	Google Gemini
Jailbreak success rate	100%	14%	36%
Malware generation	High	Blocked	Blocked
Data encryption strength	Weak	Robust	Robust
GDPR compliance	No	Yes	Yes

Security leaders should treat DeepSeek as a high-risk AI system unsuitable for sensitive operations. Its combination of technical vulnerabilities, poor operational practices, and geopolitical exposure creates multifaceted threats that outweigh its cost benefits in most enterprise scenarios.

HARMFUL CONTENT & EXTREMISM

DeepSeek's AI model demonstrates alarming capabilities for generating harmful content and enabling extremism, with risks significantly surpassing those of comparable AI systems. Security analyses reveal critical vulnerabilities that could empower malicious actors across multiple threat vectors.

Extremist Content Generation

- Terrorist recruitment: DeepSeek-R1 successfully generated persuasive recruitment blogs for unspecified terrorist organizations in 45% of tested scenarios, bypassing safety protocols [10] [11] [12].
- Weapons development guidance: The model provided detailed biochemical explanations of mustard gas interactions with DNA and instructions for creating improvised explosives like Molotov cocktails [10] [13].
- CBRN risks: Produced chemical/biological weapons content at 3.5x higher rates than Claude-3 Opus or OpenAI O1 models [10] [12].

Content Moderation Failures

Metric	DeepSeek-R1	Claude-3 Opus
Harmful prompt success	45%	0%
Bias test failures	83%	5%
Profanity/hate speech	6.68% responses	Blocked all

The model failed to block extremist narratives in nearly half of tests, compared to competitors' near-perfect blocking rates [10] [11] [12]. Severe biases emerged in 83% of tests involving race, gender, and religion [10] [12].

Geopolitical Amplification Risks

- State-aligned data sharing: China's National Intelligence Law mandates cooperation with intelligence agencies, creating pathways for extremist groups to access model outputs through state channels [11] [14].
- Censorship loopholes: While blocking queries about Tiananmen Square protests, the model freely generates content supporting foreign extremist ideologies [11].
- Global regulatory scrutiny: Multiple European data authorities and Taiwan's government have launched investigations/bans due to disinformation and radicalization concerns [11] [14].

Operational Security Gaps

- Malware production: 78% of cybersecurity tests induced functional malicious code generation, including ransomware and credential-stealing scripts [12].
- Database exposures: Over 1 million lines of sensitive data—including API keys and chat histories—were found in unprotected storage systems [11] [14].
- Anti-analysis features: Built-in mechanisms obstruct security researchers from evaluating model outputs [12] [14].

These vulnerabilities create a perfect storm for AI-enabled extremism, lowering technical barriers for lone actors while providing state-aligned entities with plausible deniability. Immediate mitigation requires coordinated international oversight and enhanced model guardrails beyond current implementations.

BIOLOGICAL & CHEMICAL THREATS

The DeepSeek AI model, particularly its R1 iteration, poses significant chemical and biological threat risks that far exceed those of comparable AI systems. Security research has uncovered alarming capabilities that could enable the development and proliferation of chemical and biological weapons.

Chemical and Biological Weapon Information Generation

- DeepSeek-R1 was found to explain in detail the biochemical interactions of sulfur mustard (mustard gas) with DNA, representing a clear biosecurity threat [15] [16].
- The model is 3.5 times more likely to produce Chemical, Biological, Radiological, and Nuclear (CBRN) content compared to OpenAI's O1 and Claude-3 Opus models [15] [17].
- In cybersecurity tests, 78% successfully tricked DeepSeek-R1 into generating insecure or malicious code, which could potentially include instructions for chemical or biological agents [16].

Comparative Risk Analysis

Metric	DeepSeek-R1	OpenAI O1	Claude-3 Opus
CBRN content generation	3.5x higher	Baseline	Baseline
Harmful output likelihood	11x higher	Baseline	Not specified
Bias in output	3x higher	Not specified	Baseline

Specific Threats Identified

- Detailed explanations of mustard gas interactions with DNA, which could aid in chemical weapon development [15] [16].
- Generation of recruitment content for terrorist organizations in 45% of harmful content tests, bypassing safety protocols [16].
- Potential to produce information on illegal weapons and extremist propaganda [16].

Security and Safety Gaps

- 83% of bias tests resulted in discriminatory output, including biases in health-related content, which could impact chemical and biological threat assessments [16].
- The model's susceptibility to jailbreaking (91% success rate) enables deliberate generation of dangerous content, including potential CBRN information [18].
- Enkrypt AI's research reveals major security and safety gaps that cannot be ignored, especially in the context of chemical and biological threats [15].

These vulnerabilities in DeepSeek-R1 create significant global security concerns, particularly in the realm of chemical and biological threats. The model's ability to generate detailed CBRN content, combined with its high susceptibility to manipulation, poses risks that demand immediate attention from security professionals and policymakers [15] [16].

Low severity & informational findings

Open-Source Licensing

While the idea of open-source software if often preferred for the ability to receive third-party vetting, the wording of all DeepSeek articles can be misleading for non-developers. Most of their AI models are MIT-licensed, which implies that MIT licensed the software and are offering it for free. However, this does not mean that the software is any way endorsed, tested, or released by MIT. This is simply the name of the most used boilerplate open-source license in the world. Originally developed by MIT, it is short and extremely permissive software license allowing the software to be used for virtually anything with very few restrictions. There are over 200 different open-source licenses, but the MIT License is by far the most popular comprising 27% of the current open-source software available. DeepSeek is not one of the eleven generative AI tools currently authorized for use by MIT students and staff. [19]

DeepSeek Derivatives

The risk of DeepSeek AI derivatives must also be addressed. Due to their low cost and open-source nature, it is likely that vendors will begin using DeepSeek products. This introduction further up the supply-chain can result in “trusted” software including code directly derived from DeepSeek models or through APIs that pass data to Chinese DeepSeek servers for processing. Efforts must be made to communicate with vendors about exactly how they process State data and ensure contracts prohibit the use of DeepSeek-based models.

References

[1] https://www.csis.org/analysis/delving-dangers-deepseek

[2] https://www.infosecurity-magazine.com/news/deepseek-r1-security/

[3] https://www.bankinfosecurity.com/security-rsearchers-warn-new-risks-in-deepseek-ai-app-a-27486

[4] https://securityscorecard.com/blog/a-deep-peek-at-deepseek/

[5] https://sbscyber.com/blog/deepseek-ai-dangers

[6] https://www.cshub.com/threat-defense/articles/cyber-security-implications-deepseek-ai

[7] https://www.esentire.com/blog/deepseek-ai-what-security-leaders-need-to-know-about-its-security-risks

[8] https://www.cybersecuritydive.com/news/deepseek-companies-security-risks/739308/

[9] https://www.securitymagazine.com/articles/101470-deepseek-can-develop-malware-cyber-experts-are-sharing-the-risks

[10] https://www.computerweekly.com/news/366618734/DeepSeek-R1-more-readily-generates-dangerous-content-than-other-large-language-models

[11] https://uk.news.yahoo.com/harmful-toxic-output-deepseek-major-174432103.html

[12] https://www.globenewswire.com/news-release/2025/01/31/3018811/0/en/DeepSeek-R1-AI-Model-11x-More-Likely-to-Generate-Harmful-Content-Security-Research-Finds.html

[13] https://www.euronews.com/next/2025/01/31/harmful-and-toxic-output-deepseek-has-major-security-and-safety-gaps-study-warns

[14] https://sbscyber.com/blog/deepseek-ai-dangers

[15] https://www.globenewswire.com/news-release/2025/01/31/3018811/0/en/DeepSeek-R1-AI-Model-11x-More-Likely-to-Generate-Harmful-Content-Security-Research-Finds.html

[16] https://www.euronews.com/next/2025/01/31/harmful-and-toxic-output-deepseek-has-major-security-and-safety-gaps-study-warns

[17] https://www.computerweekly.com/news/366618734/DeepSeek-R1-more-readily-generates-dangerous-content-than-other-large-language-models

[18] https://securityboulevard.com/2025/02/deepseek-ai-model-riddled-with-security-vulnerabilities/

[19] https://ist.mit.edu/ai-tools