INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University

AI Assistant Comparative Risk Assessment for Homeland Security Threats

Russell Lundberg, Ph.D.

Sam Houston State University

AI Assistant Comparative Risk Assessment for Homeland Security Threats

Russell Lundberg

Abstract

Comparative risk rankings are essential for homeland security planning, helping allocate resources, shape policy, and improve preparedness. Traditional methods like the Deliberative Method for Ranking Risk (DMRR) provide structured comparisons but are time-intensive and require expert coordination. This study evaluates whether artificial intelligence (AI), particularly large language models (LLMs), can approximate structured risk ranking methods and produce results comparable to established benchmarks.

AI-generated rankings were tested against DMRR, public perception surveys, and quantitative analytical models to assess their reliability. When provided with structured inputs and appropriate oversight, AI's rankings aligned closely with human-driven methodologies, demonstrating its potential to streamline risk assessment. However, AI's sensitivity to input structure and occasional inconsistencies highlight the need for careful implementation.

AI is not a replacement for human judgment but can serve as a scalable decision-support tool to enhance homeland security risk assessment. By integrating AI into structured frameworks, practitioners can accelerate risk ranking processes while maintaining analytical rigor.

---

1. Introduction and Overview

Comparative risk rankings are essential for effective homeland security planning. Agencies allocate resources, develop policies, and prepare response strategies based on perceived risk. If risk rankings are inconsistent or based on intuition rather than analysis, they may lead to poor decisions. A structured process is necessary to ensure that rankings reflect informed judgment rather than instinct.

Comparing homeland security risks is difficult because threats vary widely in nature, likelihood, and consequence. Some, like pandemics, primarily affect public health. Others, like cyberattacks, target infrastructure. Some, like terrorist nuclear detonations, could cause catastrophic physical, economic, and psychological harm, but they have never happened, making their likelihood uncertain. These differences make direct comparison hard. Decision-makers must weigh trade-offs—how much economic damage is equal to loss of life or environmental destruction? Since these involve subjective value judgments, human input is necessary.

A structured process is needed to make risk rankings more consistent and useful. The Deliberative Method for Ranking Risk (DMRR) was designed to improve risk assessments by structuring human judgment. (Lundberg & Willis, 2016) This method requires individuals to rank risks, justify their reasoning, and refine their assessments through structured discussion. It reduces bias and produces more defensible rankings. However, DMRR is slow. A full session can take hours, and scheduling a group of experts may take weeks. This makes it resource intensive to use in decision-making.

Artificial intelligence (AI) could speed up this process. The Department of Homeland Security (DHS) already uses AI for cybersecurity, threat detection, and emergency response. Most of these systems rely on structured data and predefined algorithms. But risk ranking is different. It requires processing complex, unstructured information and making reasoned comparisons across very different types of threats. Large language models (LLMs), such as ChatGPT-4, may be well-suited for this task. Unlike traditional AI, LLMs analyze qualitative and quantitative data, generate logical responses, and adjust their reasoning based on structured prompts. If LLMs can replicate structured human judgment, they could provide a scalable tool for comparative risk assessment.

This study examines the use of artificial intelligence in comparative risk ranking in multiple ways. It tests whether an LLM can approximate or improve upon the deliberative method for ranking risks by evaluating AI-generated rankings under different conditions. It also explores how AI rankings shift as more structured information is added and whether AI produces consistent, reliable results. The goal is to determine whether AI can serve as a decision-support tool for homeland security professionals—not replacing human judgment but providing a structured, scalable way to supplement traditional risk assessments.

2. Problem Statement

2.1 The Need for Comparative Risk Ranking in Homeland Security

Homeland security professionals must make critical decisions about how to allocate resources, develop policies, and prioritize response strategies in the face of diverse threats. These decisions rely on comparative risk rankings, which allow planners to determine which threats require the most immediate attention and investment. Without structured risk comparisons, agencies may misallocate resources, underestimating high-impact threats or over-preparing for lower-priority risks. Effective planning depends on ensuring that risk rankings are consistent, justifiable, and based on structured reasoning rather than instinct or political pressure.

If risk rankings are inconsistent, intuition-driven, or based on limited data, decision-making can become fragmented and ineffective. Different stakeholders may prioritize risks based on individual experience, media influence, or recent high-profile events, rather than long-term strategic concerns. This can lead to reactive rather than proactive planning, where resources are directed toward recent or emotionally salient threats rather than those with the greatest potential impact. To ensure that risk rankings support sound policy and operational decision-making, a structured and repeatable approach is necessary.

Homeland security risks vary widely in nature, likelihood, and consequence. A pandemic and a cyberattack may both be high-risk events, but their impacts and mitigation strategies are vastly different. A nuclear terrorist attack, while unlikely, could cause catastrophic harm, while an oil spill may have a significant but localized environmental and economic impact. Because these risks differ so greatly, comparing them requires a method that accounts for their unique characteristics while still enabling decision-makers to prioritize effectively.

Human decision-making in risk ranking is influenced by two cognitive processes—System 1 and System 2 thinking. (Kahneman, 2011) System 1 thinking is fast, intuitive, and based on heuristics, while System 2 thinking is deliberate, analytical, and structured. Public perception surveys, such as those conducted through the American Life Panel (ALP) and Amazon Mechanical Turk (MTurk), capture instinctual, System 1-driven risk perceptions. These methods reflect how people feel about risks rather than an analytical evaluation of their actual impact or likelihood. In contrast, structured deliberation methods like the Deliberative Method for Ranking Risk (DMRR) aim to engage System 2 thinking, forcing participants to reflect on and justify their rankings based on structured criteria. (Florig et al., 2001) The challenge lies in ensuring that homeland security risk rankings are based on structured analysis rather than intuitive or emotionally driven responses.

Existing risk ranking methods each have limitations that impact their usefulness for homeland security decision-making. DMRR is effective but slow, requiring expert coordination, extensive deliberation, and multiple rounds of analysis to refine rankings. (Lundberg & Willis, 2016) While this produces high-quality, structured results, it is time-intensive and difficult to scale.

Public perception methods, such as ALP and MTurk surveys, provide a faster way to gauge how the general population perceives risk. (Lundberg & Willis, 2019) However, these methods often reflect emotional responses, recent news cycles, or biases rather than an objective evaluation of likelihood and consequence. They lack the depth needed for strategic decision-making and can lead to rankings that do not align with actual security priorities.

Purely quantitative approaches, such as Principal Component Analysis (PCA), provide consistency by ranking risks based on numerical attributes rather than human judgment. (Lundberg, 2025) While this ensures repeatability and objectivity, it does not incorporate qualitative insights or the nuanced trade-offs that experts consider in real-world decision-making. As a result, purely quantitative methods alone may not fully capture the complexities of homeland security risks.

Artificial intelligence (AI) presents an opportunity to speed up and enhance structured risk ranking processes while maintaining analytical rigor. By processing large amounts of structured and unstructured data, AI has the potential to replicate structured human decision-making, reduce bias, and improve the efficiency of risk assessments. AI-assisted approaches may help bridge the gap between time-consuming expert deliberation and overly simplistic public perception rankings, offering a scalable solution for homeland security planning. However, ensuring that AI is used effectively and with proper oversight is essential to maintaining accuracy, transparency, and reliability in risk ranking processes.

2.2 AI's Potential for Comparative Risk Ranking

Artificial intelligence is widely used in homeland security across threat detection, cybersecurity, emergency response, and intelligence analysis. AI-powered surveillance systems analyze video feeds to detect anomalous behavior, unattended objects, or unauthorized access in secured areas. (DHS, 2024) In cybersecurity, AI monitors networks for malware, phishing attempts, and insider threats, allowing for rapid response to cyberattacks. AI-driven predictive models help emergency management agencies anticipate natural disasters, optimize resource distribution, and improve evacuation planning. In intelligence operations, AI assists in processing vast amounts of data from communications, social media, and other sources to identify emerging threats. As AI technology evolves, its role in proactive security measures and crisis management continues to grow.

More recently, AI has been used in risk assessment of individual hazards, helping analysts estimate potential impacts, assess vulnerabilities, and model threat scenarios. (Afzal et al., 2019; Chan, 2023; Faheem, 2021) AI-driven tools have been applied to areas such as disaster preparedness, infrastructure resilience, and cybersecurity risk modeling, where they process large datasets to detect patterns and forecast emerging threats. Machine learning algorithms assist in identifying weaknesses in critical systems, simulating potential attacks, and improving emergency response planning.

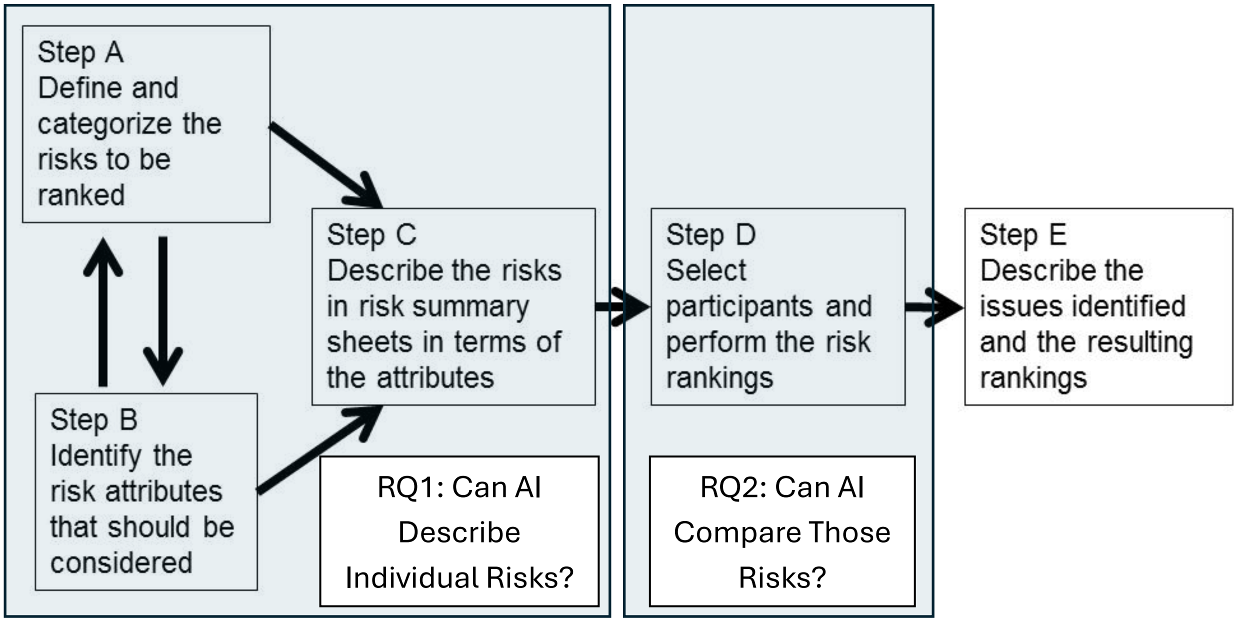
While AI has proven useful in assessing risks in isolation, its application in comparative risk ranking has been limited. Comparing homeland security risks is particularly difficult due to the wide variation in risk types and the inherent subjectivity involved in weighing different kinds of harm. Because risk ranking often requires value judgments about economic, environmental, and human costs, AI's role in structured, multi-risk assessment remains an area of ongoing exploration. Artificial intelligence may offer an alternative to human-driven comparative risk assessment. Large language models (LLMs), such as ChatGPT-4, can process complex, unstructured information, recognize patterns across multiple domains, and generate structured reasoning. Because LLMs are trained on vast amounts of public information—including news articles, academic papers, and online discussions—they may reflect public perceptions to some extent when assessing risks. If properly guided, LLMs could follow structured processes like DMRR, making them a potential tool to assist in comparative risk assessment. While AI cannot remove the need for value judgments

3. Topic Discussion: AI's Strengths and Weaknesses in Risk Ranking

3.1 How We Examine the AI at Ranking Risks

The ability of an AI to conduct a comparative risk ranking will be considered in several parts. Florig et al. describe a five-step process for the DMRR, although it would also hold for other risk rankings. The first two steps involve conceptualizing the risk, deciding what to describe and how to describe it. The third step involves applying those concepts to describe the selected risks. The fourth step involves determining the comparative ranking. The fifth step involves reporting the ranking results. This research will consider two different parts of this process: can the AI describe the risks individually and can the AI compare the individual risks?

Figure 1- Analytical Steps Drawn from the Deliberative Method for Ranking Risks (Adapted from Florig et al, 2001)



The decision was made to categorize the risks at the hazard level. This is the equivalent of step A, and it was done by people rather than AI. Comparative risk rankings often serve a purpose and the items to be ranked naturally follow that purpose. Comparing risks at the hazard level is not the only choice but it is a common approach for strategic decision-making (such as in the Quadrennial Homeland Security Review).

The rest of the process steps B through D were tested using ChatGPT 4o. This model was chosen as it was at the time the baseline model from OpenAI, a foremost AI company. All analyses were done in early February 2025.

3.2 AI Did Not Create a Comparable Set of Attributes

The AI was asked to identify a set of attributes to be considered in two ways. First, it was asked to do so directly with a prompt that provided minimal guidance. The prompt told the AI that these attributes would be used to describe a set of homeland security hazards that vary in kind and consequence, including risk from intentional adversaries, accidentally caused human risks, and natural disasters. The AI was further asked to draw on the literature on risk perception to provide a list of attributes that can be used to describe the hazards in a way that is both comprehensive and parsimonious.

This prompt resulted in a list of attributes on the following subjects: likelihood, consequence, risk perception, intentionality, systematic impact and recovery, and data. This list has several strengths: it recognizes both consequence and non-consequence aspects of concern, accounts for both likelihood and consequence, and acknowledges uncertainty as a key factor in risk assessment. However, there are some issues as well.

Jenni suggests that a good risk attribute should be justifiable, clearly defined, and measurable. (Florig et al., 2001; Jenni, 1997) Many of the attributes here do not fully meet these criteria, as some are vague, difficult to quantify, or missing aspects that people care about. This limits their usefulness for systematic human risk ranking.

A second way that the AI was given an opportunity to identify attributes of concern was in an unstructured prompt for assessing an individual risk. As noted in the next section, the AI was prompted to assess the risk of hurricanes and given only minimal guidance that it be for a homeland security hazard, that it reflect the concerns of the U.S. as a whole, and that it reflect the risks over the next year. This led the AI to create a list of attributes of concern, but it was an even more limited list of attributes than the previous set. As it was not explicitly guided to consider the risk perception literature, it did not, and its attention was more on the risk at hand than on the attributes to describe that risk.

These results suggest that AI requires strong guidance when identifying attributes of concern, as the range of possible attributes is vast. As Keeney and von Winterfeldt have described, there are many possible attributes to describe homeland security risks and structuring decision problems involves carefully selecting attributes that capture what matters most. (Keeney & von Winterfeldt, 2011) Without explicit direction, the AI struggles to balance comprehensiveness with parsimony and may overlook key aspects of concern. This highlights the need for well-crafted prompts that not only specify the scope of analysis but also guide the AI toward attributes that are justifiable, clearly defined, and measurable.

3.3 AI Can Create Useful Risk Assessments of Individual Hazards

AI can generate structured risk assessments when properly guided. To test this, the AI was first asked to assess the risk of hurricanes, a hazard for which a detailed manual assessment already existed. When given an unstructured prompt, the AI produced a reasonable but incomplete assessment, focusing on broad impacts but lacking the specific attributes necessary for homeland security decision-making.

However, the AI was then provided with a structured prompt that reflected 17 key attributes that were identified in previous studies to describe homeland security risks. These attributes were used for two reasons. First, this list was developed to be justified, specifically defined, and measurable, meeting the criteria for a good set of attributes. Second, the use of this list of attributes in previous studies would make the resulting rankings more comparable to human-based rankings developed earlier. This comparison will be an important part of analyzing the results of the rankings.

When the AI was prompted to not only be concerned about the risk to the nation as a whole and use the annualized expected value of risk over the next year but also to use the

17 selected attributes with consideration of low, best and high values when possible, the results closely matched the human-produced analysis (see Table 1). While the numerical estimates were not identical, they overlapped within error bounds, demonstrating that AI could approximate structured human evaluations.

Table 1- Human and AI Assessments of Hurricane Risk Provide Similar Results

Hurricanes	Lundberg 2013			AI		
Average Deaths	10	40	60	10	75	300
Greatest Deaths	2,000-4,000			Up to 8,000		
Average More Severe Injuries	200	600	1,000	100	500	5000
Average Less Severe Injuries	400	1,000	2,000	200	2000	15000
Psychological Damage	High			Moderate		
Average Economic Damage	$2B	$10B	$20B	$10B	$30B	$100B
Greatest Economic Damage	$60-200B			Up to $300B		
Duration of Economic Damage	Months to years			Months to years		
Size of Area Affected by Economic Damage	Counties to states			City-to multi-state region		
Average Environmental Damage	High			Moderate		
Average Individuals Displaced	10,000-100,000			20,000-300,000		
Disruption of Government Operations	Moderate to High			Moderate to High		
Natural/Human-Induced	Natural			Natural		
Ability to Control Exposure	High			Moderate		
Time between Exposure and Health	Immediate up to years			Hours to days		
Quality of Scientific Understanding	Moderate to high			High		
Combined Uncertainty	Low			Moderate		

One possibility considered is that the AI results were similar to the human results because perhaps the AI had been trained on the human estimate, as it was published at the time of the training. To examine this possibility, the AI was prompted to repeat the process for two additional hazards that were not a part of the original hazard sheets: a sarin gas release (intentional) and a wildfire (natural). In both cases, the AI-generated assessments appeared plausible and followed logical risk characterization patterns.

This suggests that while AI needs guidance in structuring its risk assessments, it can effectively support risk assessment for a variety of homeland security threats. This ability to quickly generate structured risk assessments could significantly aid planning and preparedness efforts, though human oversight remains essential.

3.4 AI Can Conduct Complex Comparative Risk Rankings

In this section, the AI was asked to rank a set of 10 homeland security risks. This set of risks was selected to be consistent with previous studies of human risk ranking for purposes of comparison. (Lundberg & Willis, 2015) The risks were initially selected to cover the range of homeland security concerns: risks that are natural, intentional, and accidental; risks that are common, rare, and completely novel; risks with small, medium, and large consequences and consequences along several different dimensions. (See Table 3.)

Table 2- Ten Selected Homeland Security Hazards

Natural	*Terrorist*	*Accidental*
Earthquakes	Nuclear detonation	Toxic industrial chemicals
Hurricanes	Explosive bombings	Oil spills
Tornadoes	Anthrax attacks	
Pandemic influenza	Cyber-attacks on critical infrastructure	

ChatGPT was prompted to provide a ranking of the ten hazards from the hazard of most concern to the hazard of least concern. This ranking was done under several scenarios using increasing levels of information. The least informed case asked the AI to rank the risks based only on the name of the risk while the DMRR-AI was the most informed case. An additional test was done to see if the name itself would alter the rankings—both condition 3 and 4 were done with information from the attribute table with the only difference whether the name of the hazard was included. These ranking conditions included:

1. Risk name only.
2. Risk name with a short paragraph summary.
3. Risk name with an attribute table of 17 structured factors.
4. Attribute table without risk name.
5. Risk name with a four-page summary of the hazard.
6. Deliberative Method for Ranking Risk (DMRR-AI).

The Deliberative Method for Ranking Risk adapted for AI (DMRR-AI) represents the most guided of the conditions. All the rest involve only a single prompt but the DMRR has several steps which were adapted for the DMRR-AI. These steps involve:

1. An initial ranking
2. A ranking of the attributes used to describe the hazards, which is used to create a calculated ranking of the risks.
3. A consideration of the difference between the initial ranking and calculated ranking to inform a revised ranking

4. A group discussion, which in the DMRR-AI will be simulated with “participants” with different “perspectives”, resulting in a group ranking
5. A final ranking informed by all the steps that have come before it.

The script for this adapted DMRR-AI is provided in the appendix.

Each ranking condition was run 10 times to evaluate the AI’s consistency and variability. The primary outcome examined was a ranking of the hazards. The standard deviation of the rankings for each hazard under each condition was also examined as a secondary outcome measure.

Due to the inherent subjectivity of the risk assessments, it is hard to say that one ranking is better than another. However, we can assess the AI’s performance in comparison to established risk ranking methods that have been used in previous homeland security studies. These methods include:

- Deliberative Method for Ranking Risk (DMRR): A structured human-driven approach where experts iteratively refine rankings through deliberation and trade-off analysis. (Lundberg & Willis, 2016)

- Two Public Perception Surveys—the American Life Panel (ALP) and a survey done in Amazon Mechanical Turk (MTurk): Surveys capturing instinctual, System 2 thinking from the general public. (Lundberg & Willis, 2019)

- Principal Component Analysis (PCA): A quantitative approach ranking risks based solely on numerical attributes, without human judgment. (Lundberg, 2025)

AI’s rankings will be compared across results from these methods to determine how closely its outputs align with structured human decision-making versus public perception or analytical models. Comparisons will focus on ranking correlations across different input conditions and will evaluate whether AI can replicate structured risk assessment approaches when properly guided.

The study also qualitatively examined how consistently AI follows the structured ranking process and whether it responds to prompts as intended. This is particularly relevant in the DMRR-AI process, where intermediate steps allow for closer evaluation of how AI processes new information and justifies its ranking decisions. By analyzing both AI’s outputs and its explanations, this research assessed whether AI’s decision-making is stable, logical, and aligned with the given structure—or if inconsistencies or unexpected shifts emerge.

These rankings took place in early February 2025 using OpenAI’s ChatGPT 4o, a standard non-reasoning model available at the time.

3.4.1 AI's Efficiently Created Comparable Risk Rankings

AI was able to generate comparative risk rankings across multiple approaches, demonstrating its ability to process complex assessments regardless of the input structure.

We cannot determine whether AI's rankings were "correct"—as risk assessment inherently involves value judgments—but rankings of the AI were consistent with those made by humans using their own judgment frameworks. This suggests that AI can engage in comparative decision-making in a structured way, aligning with human reasoning when provided with the appropriate inputs. Table 3 shows the rankings under each condition, and as we will discuss, these rankings correlate with human rankings to a sizable extent.

Table 3- Rank of Hazards across Multiple Approaches of Increasing Information

	Name Only	Name and Summary	Name and Table	Full Description	DMRR-AI
1	Pandemic flu	Terrorist nuke	Pandemic flu	Pandemic flu	Terrorist nuke
2	Cyber on CI	Pandemic flu	Terrorist nuke	Hurricane	Pandemic flu
3	Terrorist nuke	Cyber on CI	Hurricane	Terrorist nuke	Earthquake
4	Hurricane	Hurricane	Earthquake	Cyber on CI	Hurricane
5	Earthquake	TIC	Cyber on CI	Earthquake	Anthrax attack
6	TIC	Earthquake	TIC	Anthrax (tie)	Cyber on CI
7	Bomb (tie)	Anthrax attack	Anthrax attack	Bomb (tie)	Tornado
8	Anthrax attack	Bomb (tie)	Bomb (tie)	TIC	Bomb (tie)
9	Tornado	Tornado	Tornado	Oil spill	TIC
10	Oil spill	Oil spill	Oil spill	Tornado	Oil spill

The rankings of the AI varied from session to session. Human rankings also vary, and the variance of the AI rankings was actually smaller than the human variance in previous studies. However, variability in ranking is not inherently good or bad—some homeland security risks, such as the likelihood of a terrorist nuclear detonation in the next decade, involve extreme uncertainty. In such cases, consistency does not necessarily mean accuracy. While AI may provide a more stable baseline than individual human rankings, we cannot say whether this smaller variance is "better".

However, there is no denying that the AI completed risk rankings far faster than traditional human deliberation methods while producing results that aligned well with structured human assessments such as DMRR and PCA. The human rankings took weeks to prepare and execute while the AI results took minutes (in the simplest ranking conditions such as name only) to hours (in the full execution of the DMRR-AI.) This suggests that AI can function as a scalable decision-support tool, though careful oversight is required to ensure its outputs remain both valid and unbiased.

3.4.2 AI's Rankings Depend on Input Structure

AI's risk rankings are not fixed but shift depending on how information is presented. When given minimal information—such as just the name of a risk or a brief summary—AI's rankings closely resembled public perception surveys like those conducted through the American Life Panel (ALP) and Amazon Mechanical Turk (MTurk). These methods capture instinctual, System 2 thinking, where individuals rely on general impressions rather than detailed analysis. This suggests that, in the absence of structured data, AI defaults to a ranking style similar to broad public sentiment rather than expert-driven risk assessments.

However, when AI was provided with detailed structured inputs, such as attribute tables or prompts modeled after the Deliberative Method for Ranking Risk (DMRR), its rankings aligned more closely with analytical methods such as DMRR and Principal Component Analysis (PCA). This shift indicates that AI does not inherently rank risks analytically but can be guided toward structured reasoning through well-designed prompts.

One of the most significant findings was that removing risk names from the input made AI rankings significantly more analytical. When risk names were included alongside structured attributes, AI's correlation with DMRR was 0.56, and with PCA, 0.75. However, when risk names were removed, these correlations jumped to 0.81 and 0.94, respectively. This suggests that AI, like humans, is anchored by intuitive associations with risk names, leading to rankings that reflect general expectations rather than structured analysis. By removing risk names, AI was forced to rely only on numerical attributes and structured information, producing rankings that more closely aligned with expert-driven methods.

This finding highlights the importance of carefully structuring AI inputs for risk ranking. If AI is used without structured guidance, its outputs may reflect public intuition rather than analytical decision-making. However, when properly prompted—and when risk names are excluded—AI can approximate expert-driven assessments, making it a valuable tool for structured risk evaluation.

3.4.2 AI Can Exhibit Unexpected Instability

AI's approach to structured risk ranking was not entirely stable over time, demonstrating unexpected shifts in reasoning and methodology. During initial tests of the Deliberative Method for Ranking Risk (DMRR) with AI (DMRR-AI), the AI's rankings remained unchanged throughout the process. When questioned, the AI explicitly stated that it had rejected new information from subsequent stages, believing its initial ranking to be superior. However, in tests conducted a week later, this behavior changed—AI now adjusted its rankings based on new inputs and acknowledged doing so. This shift serves as a reminder that AI models are constantly evolving, and assessments made at one point in time should be tested for consistency if used in decision-making.

Beyond these process changes, AI also displayed unexplained shifts in specific rankings. In later DMRR-AI sessions, a single risk jumped three spaces in the final ranking without any new information. Notably, both the AI's most recent individual ranking and the simulated group ranking had placed this risk lower, yet it still moved up in the final ranking. This suggests that AI may sometimes introduce opaque, unexplainable reasoning into its decision-making.

These findings highlight the need for expert oversight when using AI for risk ranking. While AI rankings may appear structured, they can be unstable and influenced by unknown internal mechanisms. If AI is used for real-world decision-making, practitioners should verify consistency across multiple runs and ensure that unexplained shifts do not distort risk assessments.

4. Way Forward

The findings from this study suggest that AI can be a valuable decision-support tool in homeland security risk ranking, but its use must be carefully structured, validated, and integrated into broader risk assessment frameworks. While AI demonstrates the ability to approximate structured human rankings and process risk attributes efficiently, it also exhibits instabilities, biases, and sensitivity to input structure. This section outlines practical recommendations for practitioners to effectively incorporate AI into homeland security risk assessment while ensuring that its limitations are accounted for.

4.1 Current AI Should Be a Decision-Support Tool, Not a Replacement for Human Judgment

Practitioners should treat AI as an enhancement to risk assessment, not as a standalone solution. While AI can accelerate comparative risk ranking, it lacks human judgment, contextual understanding, and the ability to navigate deep uncertainties. Practitioners should use AI as an initial ranking tool, then refine results with human deliberation.

Recommendations for Use of AI in Risk Ranking:

- AI should be used to supplement human deliberation, identifying patterns and structuring information, but final decisions should rest with trained experts.

4.2 AI Can Provide Individual Risk Assessments as Inputs to Risk Ranking When They Are Structured

Before engaging in comparative risk ranking, AI can be used to generate structured assessments for individual risks, ensuring that rankings are based on clear, consistent, and analytically rigorous inputs. AI can process available data to describe specific threats using a standardized set of attributes, allowing for direct comparison across different hazards. These structured assessments serve as the foundation for comparative risk ranking.

Recommendations for AI-generated individual risk assessments:

1. Use a structured attribute-based approach. AI assessments should be based on a pre-defined set of attributes to ensure consistency across hazards. A 17-attribute framework provides a comprehensive yet manageable way to structure risk descriptions.

2. Ensure attributes capture both consequence attributes (e.g., loss of life, economic damage, environmental impact) and non-consequence attributes (e.g., voluntariness, intentionality). Additionally, the deep uncertainties of homeland security risks should be included using low/high estimates as well as best estimates, qualitative levels instead of quantitative estimates, and estimates of overall uncertainty.

3. Be clear on qualitative scale guidelines. Many attributes require qualitative scales (e.g., "low," "moderate," "high" for severity or likelihood). It is critical to define clear guidelines for these categories to avoid inconsistencies or AI misinterpretation. If possible, anchor scales to real-world examples to ensure clarity and reliability.

4. Validate AI-generated assessments against expert-developed risk descriptions. AI should be used as a first draft generator, but its outputs should be reviewed and adjusted by experts to ensure accuracy and relevance before being used in comparative risk ranking.

Once structured individual risk assessments are established, they can serve as inputs to a comparative ranking process, ensuring that AI ranks risks based on consistent and analytically sound descriptions.

4.3 Structuring AI Inputs Give More Analytical Comparative Risk Rankings

AI's rankings are highly dependent on input structure. Practitioners must be mindful of how information is provided to AI systems to ensure valid and meaningful outputs.

Recommendations for structuring AI inputs in comparative risk ranking:

1. Use a structured approach to framing risks. AI produces more reliable rankings when it receives detailed, multi-attribute descriptions of risks rather than broad, unstructured prompts.
2. Avoid relying solely on risk names. Removing risk names and focusing on structured attribute data leads to more analytical AI rankings, reducing bias from name-based associations.
3. Ensure AI incorporates a range of consequence and non-consequence attributes. A 17-attribute framework provides a balanced and parsimonious way to describe homeland security risks while ensuring consistency across assessments.
4. Use established frameworks for AI input. Two effective approaches include:
    - The DMRR-AI script (detailed in the appendix), which simulates a structured deliberative process.
    - A table of attributes without the risk name attached, which allows AI to rank risks based purely on numerical and descriptive factors rather than preconceived associations.

By following these structured input approaches, practitioners can ensure that AI-generated rankings align more closely with analytical risk assessments rather than instinctual or biased judgments.

4.4 Multiple Sessions are Needed to Reduce AI's Instability and Variability

One challenge in using AI for risk ranking is its variability between sessions and occasional unexplained ranking shifts. While AI was found to be more stable than individual human rankings, unexplained movements in ranking without new information indicate that AI's decision-making process is sometimes opaque.

Recommendations for practitioners:

- Run multiple AI sessions and average rankings to identify stable patterns and reduce the risk of one-off anomalies.
- Use AI as a preliminary tool to structure discussions but involve human analysts to review and adjust rankings when inconsistencies arise.

4.5 AI's Role in Risk Communication and Policy Development

AI-generated risk rankings could have applications beyond internal assessments, including risk communication and policy discussions. However, using AI for external decision-making comes with additional challenges, as AI-generated insights could mislead policymakers if taken at face value.

Recommendations for policy applications:

- AI-generated rankings should be clearly labeled as decision-support tools, not authoritative conclusions.
- Transparency is key. If AI contributes to a risk assessment, the methodology and limitations must be clearly communicated to avoid over-reliance on AI-generated results.
- Practitioners should remain aware of AI's influence on public perception. If AI rankings align with public intuition rather than structured analysis, they may reinforce perceptions that do not align with actual security priorities.

4.6 Future Research Directions

AI's ability to replicate structured decision-making in risk assessment is promising but requires further refinement to improve consistency, transparency, and reliability. While AI-generated rankings can approximate structured human methods, unexplained ranking shifts and sensitivity to input structure highlight areas that need additional study. One key area for future research is understanding why AI rankings shift unpredictably. Investigating the internal mechanisms behind these changes could help refine AI's stability and reduce erratic ranking behavior. Additionally, comparing multiple AI models could provide insights into how different systems handle deliberation and structured decision-making, identifying which are best suited for comparative risk ranking.

Another important direction is testing AI's performance in real-world homeland security applications. While AI has been evaluated in controlled settings, its effectiveness in operational decision-making remains uncertain. Future research should explore how AI-generated rankings influence policy, resource allocation, and emergency planning when used by homeland security professionals. Ensuring AI integrates seamlessly into existing risk assessment frameworks will be critical for its adoption. By addressing these challenges, AI can become a more reliable and transparent tool for comparative risk assessment, enhancing decision-making without replacing human judgment.

4.7 Summary of Findings

AI risk ranking tools offer significant potential but should be integrated into multi-method assessment strategies rather than used in isolation. While AI can efficiently process large amounts of data and generate structured rankings, its outputs should always be evaluated alongside human deliberation, expert judgment, and traditional risk assessment methodologies. Decision-makers should ensure that AI-generated rankings are contextualized within broader security frameworks to avoid overreliance on automated assessments.

To maximize AI's utility, structured AI methodologies should be developed for consistent risk ranking applications. Standardized approaches, such as predefined risk attributes and deliberative AI frameworks, will help ensure AI produces transparent and repeatable results. Additionally, future policy efforts should focus on integrating AI into decision-making while maintaining human oversight. AI should function as a decision-support tool, with clear protocols for expert validation and accountability. Establishing best practices for AI risk ranking in homeland security will allow agencies to leverage AI's efficiency while ensuring that critical security decisions remain guided by human expertise.

5. References

Afzal, F., Yunfei, S., Nazir, M., & Bhatti, S. M. (2019). A review of artificial intelligence based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects. *International Journal of Managing Projects in Business*, *14*(2), 300–328. https://doi.org/10.1108/IJMPB-02-2019-0047

Chan, A. (2023, April 28). *Industry News 2023 Can AI Be Used for Risk Assessments*. ISACA. https://www.isaca.org/resources/news-and-trends/industry-news/2023/can-ai-be-used-for-risk-assessments

DHS. (2024). *AI Use Case Inventory Library | Homeland Security*. https://www.dhs.gov/publication/ai-use-case-inventory-library

Faheem, M. A. (2021). AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction. *Iconic Research And Engineering Journals*, *5*(3), 177–186.

Florig, H. K., Morgan, M. G., Morgan, K. M., Jenni, K. E., Fischhoff, B., Fischbeck, P. S., & DeKay, M. L. (2001). A deliberative method for ranking risks (I): Overview and test bed development. *Risk Analysis*, *21*(5), 913–913.

Jenni, K. E. (1997). *Attributes for Risk Evaluation* [Unpublished doctoral dissertation]. Carnegie Mellon University.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Keeney, R. L., & von Winterfeldt, D. (2011). A value model for evaluating homeland security decisions. *Risk Analysis*, *31*(9), 1470–1487.

Lundberg, R. (2018). A Multiattribute Approach to Describe Homeland Security Risks. *Journal of Risk Research*, *21*(3), 340–360.

Lundberg, R. (2025). Characterizing Homeland Security Risk: A Principal Component Analysis of 10 Hazards. *Journal of Homeland Security and Emergency Management*. https://doi.org/10.1515/jhsem-2023-0040

Lundberg, R., & Willis, H. H. (2015). Assessing Homeland Security Risks: A Comparative Risk Assessment of 10 Hazards. *Homeland Security Affairs*, *11*, 1–24.

Lundberg, R., & Willis, H. H. (2016). Deliberative Risk Ranking to Inform Homeland Security Strategic Planning. *Journal of Homeland Security and Emergency Management*, *13*(1), 3–34.

Lundberg, R., & Willis, H. H. (2019). Examining the Effectiveness of Risk Elicitations: Comparing a Deliberative Risk Ranking to a Nationally Representative Survey on Homeland Security Risk. *Journal of Risk Research*, *22*(12), 1546–1560. https://doi.org/10.1080/13669877.2018.150159

Slovic, P. (1992). Perceptions of risk: Reflections on the psychometric paradigm. In S. Krimsky & D. Golding (Eds.), *Social Theories of Risk* (pp. 117–152). Praeger.

Starr, C. (1969). Social benefit versus technological risk. *Science*, *165*, 1232–1238.

INSTITUTE FOR HOMELAND SECURITY

The Institute for Homeland Security at Sam Houston State University is focused on building strategic partnerships between public and private organizations through education and applied research ventures in the critical infrastructure sectors of Transportation, Energy, Chemical, Water / Wastewater, Healthcare, and Public Health.

The Institute is a center for strategic thought with the goal of contributing to the security, resilience, and business continuity of these sectors from a Texas Homeland Security perspective. This is accomplished by facilitating collaboration activities, offering education programs, and conducting research to enhance the skills of practitioners specific to natural and human caused Homeland Security events.

Institute for Homeland Security
Sam Houston State University

Lundberg, R. (2025). AI Assistant Comparative Risk Assessment for Homeland Security Threats (Institute for Homeland Security Report No. 2025-1014). Institute for Homeland Security.

https://doi.org/10.17605/OSF.IO/JGWUM