**How STRIDE elements can apply to GenAI and LLM threat modeling (The AI-based Wedding/Matchmaking Site)**

| **STRIDE Element** | **Application to GenAI/LLM** |
| --- | --- |
| **Spoofing** | An attacker could impersonate a legitimate user or system to feed malicious input to the AI model. |
| **Tampering** | The training data or the model parameters could be manipulated to affect the model’s behavior or outputs. |
| **Repudiation** | There might be a lack of auditing controls to track the interactions with the AI system, allowing attackers to deny their actions. |
| **Information Disclosure** | Sensitive data used in training or interacting with the model could be exposed due to insufficient data handling practices. |
| **Denial of Service (DoS)** | The AI service could be overwhelmed with requests or given complex inputs to consume computational resources, causing it to fail. |
| **Elevation of Privilege** | Exploiting vulnerabilities in the AI system could give an attacker unauthorized access or control over the system’s functions. |

For example, a matchmaking site using an LLM for generating personalized interactions could suffer from:

* **Spoofing**: If user verification mechanisms are weak, an attacker could provide input as a user and train the model to produce undesirable responses.
* **Tampering**: An insider could inject biased data into the training set, leading to skewed results that favor certain outcomes.
* **Repudiation**: Without proper logging, it would be difficult to trace back and verify the actions of users who might provide harmful inputs.
* **Information Disclosure**: The model might inadvertently reveal personal details about users in its responses if the training data isn't properly anonymized.
* **Denial of Service**: Sending numerous complex queries could slow down or crash the service, denying legitimate users access.
* **Elevation of Privilege**: Flaws in the system could potentially be exploited to gain administrative access, allowing for broad manipulation of the AI model.

**STRIDE Threat Modeling for AI-based Online Wedding Site**

| **STRIDE Element** | **Potential Threats** | **Mitigation Strategies** |
| --- | --- | --- |
| **Spoofing** | An attacker could impersonate a legitimate user to manipulate the AI’s behavior or access sensitive information. | Implement strong authentication mechanisms such as multi-factor authentication (MFA) and regular user validation checks. |
| **Tampering** | Malicious users might modify the AI’s training data or code to influence its behavior, potentially causing it to provide incorrect or harmful suggestions. | Use secure coding practices, regularly audit and monitor code and data integrity, and implement access controls to prevent unauthorized changes. |
| **Repudiation** | Users might deny actions they performed on the site, such as making bookings or sending messages, which could lead to disputes or fraudulent claims. | Implement comprehensive logging and monitoring to track user activities and changes made to the system, ensuring actions can be verified. |
| **Information Disclosure** | Sensitive data, such as personal details of the couples, payment information, or vendor contracts, could be exposed through the AI system. | Encrypt sensitive data, both at rest and in transit, and apply strict data access controls and privacy settings. |
| **Denial of Service (DoS)** | Attackers could overload the AI system with requests, rendering the wedding planning services unavailable to legitimate users. | Deploy rate limiting, robust traffic management, and distributed denial-of-service (DDoS) protection mechanisms to maintain service availability. |
| **Elevation of Privilege** | Attackers could exploit vulnerabilities in the AI system to gain unauthorized access or control over the site’s functionalities. | Regularly patch and update the AI system and underlying infrastructure, conduct vulnerability assessments, and use the principle of least privilege for access control. |