# TB S/W

|  |  |
| --- | --- |
| TB01 | Entity (LLM, Endpoint, Server, etc ) |
| Spoofing | Strengths | Weaknesses |
| Tampering |  |  |
| Repudiation |  |  |
| Information disclosure |  |  |
| Denial of service |  |  |
| Elevation of privilege |  |  |

## Example for TB S/W

|  |  |
| --- | --- |
| TB01 | External Endpoints |
| Spoofing | Strengths | Weaknesses |
| Tampering | ….. | VULN01: Modify System prompt(prompt injection) |
| Repudiation | Proper authentication and authorisation (assumed) | …. |
| Information disclosure | …. | …. |
| Denial of service | … |  |
| Elevation of privilege | … | …. |

# Vulnerability list

|  |  |  |
| --- | --- | --- |
| Vuln ID | Description | Examples |
| VULN01 | Modify System prompt (prompt injection) | Users can modify the system-level prompt restrictions to "jailbreak" the LLM and overwrite previous controls in place |
| VULN02 |  |  |
| VULN03 |  |  |
| VULN04 |  |  |

# Recommendations

|  |  |
| --- | --- |
| REC ID | Recommendations for Mitigation |
| REC1 | void training GenAI models on non-public or sensitive data. Treat all GenAI output as untrusted and apply restrictions based on the data or actions the model requests.  |
|  |  |
|  |  |
|  |  |