QBE Cyber Services

Understanding LLMjacking

An AI Security Threat

What is LLMjacking?

LLMjacking is the unauthorised hijacking of Large Language Model (LLM) resources, where attackers exploit vulnerabilities to steal computational power — the core asset behind modern AI systems. This misuse can take several forms, all of which result in excessive or unauthorised consumption of compute, including:

- Running unauthorised AI workloads using stolen access
- Training or fine-tuning large models for commercial or malicious use
- Extracting proprietary training data or model weights through compute-heavy probing
- Manipulating model outputs to serve malicious content — particularly when such misuse leads to sustained or abnormal resource usage

Key Risk Indicators

Your organisation may be vulnerable if you:

- Use cloud-based AI/ML services (OpenAI, Anthropic, Google AI, etc.)
- Deploy self-hosted LLMs without proper access controls
- Share API keys across teams or applications
- Lack monitoring for abnormal API usage

Recommended Protection Strategies

While each organisation may use AI differently, the following are key strategies specifically aimed at reducing the risk of LLMjacking:

Access Control & Authentication

- Implement strict API key rotation policies (30-90 days)
- Use separate keys for development, testing, and production
- Restrict remote administrative access to allow listed IP addresses where possible
- Implement usage alerts and spending caps

Monitoring & Detection

- Set up real-time alerts for potentially malicious usage spikes
- Monitor for rapidly increasing query patterns or volumes
- Log all API calls with user attribution

Technical Safeguards

- Audit applications for hardcoded credentials
- Implement and audit prompt injection defences
- Regular security audits of AI integration

Incident Response Planning

- Define clear escalation procedures for suspected hijacking
- Maintain ability to quickly revoke compromised credentials
- Document AI system dependencies

Contact Us

Visit qbe.com/cyber/cyber-services to find out more about our QBE Cyber Services complimentary and preferred pricing offerings.

For further assistance, contact QBE Cyber Services at QCyberServices@qbe.com.

QBE Insurance