Generative Backdoor Optimization and LLM-Based Recommendation for Robust Intelligent Systems

Abstract

Robust intelligent systems must address both external adversarial manipulation and internal data contamination. Universal transferable attacks show that vision-language models can be vulnerable to targeted adversarial examples, while clean-image backdoors demonstrate that malicious triggers can be optimized without obvious visual artifacts. These risks are particularly relevant for systems that combine multimodal models, personalized recommendation, and lifelong learning. LLM-based sequential recommendation introduces dynamic indexing mechanisms that improve user modeling but also expand the attack surface for poisoned data, biased inputs, and unreliable outputs. Biomedical evidence on remote cardiac injury provides an example of complex high-stakes reasoning where trustworthy learning is essential. Flood disaster relocation research adds a public-policy context in which intelligent systems may support decisions affecting households and communities. This literature cluster emphasizes adversarial robustness, dynamic recommendation, biomedical reasoning, and disaster-oriented decision support.

Keywords

pre-disaster relocation; agent-based model; flood disaster; household behavior; disaster governance; relocation planning; climate adaptation

References

Xu, B., Dai, X., Tang, D., & Zhang, K. (2025, November). One surrogate to fool them all: Universal, transferable, and targeted adversarial attacks with clip. In Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (pp. 3087-3101).

Xu, B., Yang, F., Tang, D., Dai, X., & Zhang, K. (2026, March). Breaking the stealth-potency trade-off in clean-image backdoors with generative trigger optimization. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 32, pp. 27197-27205).

Yin, J., Zeng, Z., Li, M., Yan, H., Li, C., Han, W., ... & Wang, S. (2025, April). Unleash llms potential for sequential recommendation by coordinating dual dynamic index mechanism. In Proceedings of the ACM on Web Conference 2025 (pp. 216-227).

Zhang, Z., Duan, Y., Gu, F., Wei, J., Huo, H., Wang, Y., ... & Gao, Z. (2026). Circulating ATP from hepatic ischemia-reperfusion drives remote cardiac injury via macrophage inflammasome activation. Pharmacological Research, 108306.

Zhou, Y. (2022). Pre-disaster relocation and agent-based model for flood disaster [Doctoral dissertation, University of Wisconsin–Madison].

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. In International Conference on Learning Representations.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In International Conference on Learning Representations.

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 39-57). IEEE.

Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019, November). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 1441-1450).

Eltzschig, H. K., & Eckle, T. (2011). Ischemia and reperfusion—from mechanism to translation. Nature Medicine, 17(11), 1391-1401.