References

Al-Dujaili, A., Huang, A., Hemberg, E., & O'Reilly, U.-M. (2018). Adversarial deep learning for robust detection of binary encoded malware. *2018 IEEE Security and Privacy Workshops (SPW)* (pp. 76-82). IEEE. https://doi.org/10.48550/arXiv.1801.02950

Anderson, H. S., & Roth, P. (2018). Ember: an open dataset for training static PE malware machine learning models. *arXiv preprint arXiv:1804.04637*, 10.48550/arXiv.1804.04637. https://doi.org/10.48550/arXiv.1804.04637

Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., & Rieck, K. (2024). Pitfalls in Machine Learning for Computer Security. *Commun. ACM, 67*(11), 104–112. https://doi.org/10.1145/3643456

Aryal, K., Gupta, M., Abdelsalam, M., Kunwar, P., & Thuraisingham, B. (2025). A Survey on Adversarial Attacks for Malware Analysis. *IEEE Access, 13*, 428-459. https://doi.org/10.1109/access.2024.3519524

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition, 84*, 317-331. https://doi.org/10.1016/j.patcog.2018.07.023

Cortellazzi, J., Quiring, E., Arp, D., Pendlebury, F., Pierazzi, F., & Cavallaro, L. (2025). Intriguing Properties of Adversarial ML Attacks in the Problem Space [Extended Version]. *ACM Trans. Priv. Secur., 28*(4), Article 42. https://doi.org/10.1145/3742895

Demetrio, L., Coull, S. E., Biggio, B., Lagorio, G., Armando, A., & Roli, F. (2021). Adversarial exemples: A survey and experimental evaluation of practical attacks on machine learning for windows malware detection. *ACM Transactions on Privacy and Security (TOPS), 24*(4), 1-31. https://doi.org/10.1145/3473039

Doan, B. G., Yang, S., Montague, P., De Vel, O., Abraham, T., Camtepe, S., Kanhere, S. S., Abbasnejad, E., & Ranashinghe, D. C. (2023). Feature-space bayesian adversarial learning improved malware detector robustness. *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 14783-14791). https://doi.org/10.48550/arXiv.2301.12680

Grosse, K., Papernot, N., Manoharan, P., Backes, M., & McDaniel, P. (2016). Adversarial perturbations against deep neural networks for malware classification. *arXiv preprint arXiv:1606.04435*, 10.48550/arXiv.1606.04435. https://doi.org/10.48550/arXiv.1606.04435

Jedrzejewski, F. V., Thode, L., Fischbach, J., Gorschek, T., Mendez, D., & Lavesson, N. (2024). Adversarial Machine Learning in Industry: A Systematic Literature Review. *Computers & security., 145*, 103988. https://doi.org/10.1016/j.cose.2024.103988

Kreuk, F., Barak, A., Aviv-Reuven, S., Baruch, M., Pinkas, B., & Keshet, J. (2018). Deceiving end-to-end deep learning malware detectors using adversarial examples. *arXiv preprint arXiv:1802.04528*, https://doi.org/10.48550/arXiv.1802.04528. https://doi.org/https://doi.org/10.48550/arXiv.1802.04528

Li, D., Li, Q., Ye, Y., & Xu, S. (2023). Arms Race in Adversarial Malware Detection: A Survey. *ACM Computing Surveys, 55*(1), 1-35. https://doi.org/10.1145/3484491

Liu, H., Sun, W., Niu, N., & Wang, B. (2022). MultiEvasion: Evasion Attacks Against Multiple Malware Detectors. *2022 IEEE Conference on Communications and Network Security (CNS)* (pp. 10-18). https://doi.org/10.1109/CNS56114.2022.9947227

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016). The limitations of deep learning in adversarial settings. *2016 IEEE European symposium on security and privacy (EuroS&P)* (pp. 372-387). IEEE. https://doi.org/10.1109/EuroSP.2016.36

Raff, E., Barker, J., Sylvester, J., Brandon, R., Catanzaro, B., & Nicholas, C. (2017). Malware detection by eating a whole exe. *arXiv preprint arXiv:1710.09435*, 10.48550/arXiv.1710.09435. https://doi.org/10.48550/arXiv.1710.09435

Suciu, O., Coull, S. E., & Johns, J. (2019). Exploring adversarial examples in malware detection. *2019 IEEE Security and Privacy Workshops (SPW)* (pp. 8-14). IEEE. https://doi.org/10.48550/arXiv.1810.08280

Vassilev, A., Oprea, A., Fordyce, A., & Anderson, H. (2024). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (10.6028/NIST.AI.100-2e2023). https://doi.org/10.6028/NIST.AI.100-2e2023

Wang, Y., Liu, J., & Chang, X. (2019). *Assessing transferability of adversarial examples against malware detection classifiers* Proceedings of the 16th ACM International Conference on Computing Frontiers, Alghero, Italy. https://doi.org/10.1145/3310273.3323072

Yan, S., Ren, J., Wang, W., Sun, L., Zhang, W., & Yu, Q. (2023). A Survey of Adversarial Attack and Defense Methods for Malware Classification in Cyber Security. *IEEE Communications surveys and tutorials, 25*(1), 467-496. https://doi.org/10.1109/COMST.2022.3225137

Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE transactions on neural networks and learning systems, 30*(9), 2805-2824. https://doi.org/10.48550/arXiv.1712.07107