* Backdoor Learning: A Survey. [[pdf]](https://www.researchgate.net/publication/343006441_Backdoor_Learning_A_Survey)
  + Yiming Li, Yong Jiang, Zhifeng Li, and Shu-Tao Xia. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
* Backdoor Attacks and Countermeasures on Deep Learning: A Comprehensive Review. [[pdf]](https://arxiv.org/pdf/2007.10760.pdf)
  + Yansong Gao, Bao Gia Doan, Zhi Zhang, Siqi Ma, Anmin Fu, Surya Nepal, and Hyoungshick Kim. arXiv, 2020.
* Data Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses. [[pdf]](https://arxiv.org/pdf/2012.10544.pdf)
  + Micah Goldblum, Dimitris Tsipras, Chulin Xie, Xinyun Chen, Avi Schwarzschild, Dawn Song, Aleksander Madry, Bo Li, and Tom Goldstein. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
* A Comprehensive Survey on Poisoning Attacks and Countermeasures in Machine Learning. [[link]](https://dl.acm.org/doi/10.1145/3551636)
  + Zhiyi Tian, Lei Cui, Jie Liang, and Shui Yu. *ACM Computing Surveys*, 2022.
* Backdoor Attacks and Defenses in Federated Learning: State-of-the-art, Taxonomy, and Future Directions. [[link]](https://ieeexplore.ieee.org/abstract/document/9806416)
  + Xueluan Gong, Yanjiao Chen, Qian Wang, and Weihan Kong. *IEEE Wireless Communications*, 2022.
* Backdoor Attacks on Image Classification Models in Deep Neural Networks. [[link]](http://cje.ejournal.org.cn/article/doi/10.1049/cje.2021.00.126)
  + Quanxin Zhang, Wencong Ma, Yajie Wang, Yaoyuan Zhang, Zhiwei Shi, and Yuanzhang Li. *Chinese Journal of Electronics*, 2022.
* Defense against Neural Trojan Attacks: A Survey. [[link]](https://www.sciencedirect.com/science/article/pii/S0925231220316350)
  + Sara Kaviani and Insoo Sohn. *Neurocomputing*, 2021.
* A Survey on Neural Trojans. [[pdf]](https://eprint.iacr.org/2020/201.pdf)
  + Yuntao Liu, Ankit Mondal, Abhishek Chakraborty, Michael Zuzak, Nina Jacobsen, Daniel Xing, and Ankur Srivastava. *ISQED*, 2020.
* A Survey of Neural Trojan Attacks and Defenses in Deep Learning. [[pdf]](https://arxiv.org/pdf/2202.07183.pdf)
  + Jie Wang, Ghulam Mubashar Hassan, and Naveed Akhtar. arXiv, 2022.
* Threats to Pre-trained Language Models: Survey and Taxonomy. [[pdf]](https://arxiv.org/pdf/2202.06862.pdf)
  + Shangwei Guo, Chunlong Xie, Jiwei Li, Lingjuan Lyu, and Tianwei Zhang. arXiv, 2022.
* An Overview of Backdoor Attacks Against Deep Neural Networks and Possible Defences. [[pdf]](https://arxiv.org/pdf/2111.08429.pdf)
  + Wei Guo, Benedetta Tondi, and Mauro Barni. arXiv, 2021.
* Deep Learning Backdoors. [[pdf]](https://arxiv.org/pdf/2007.08273.pdf)
  + Shaofeng Li, Shiqing Ma, Minhui Xue, and Benjamin Zi Hao Zhao. arXiv, 2020.

**Toolbox**

* [BackdoorBox](https://github.com/THUYimingLi/BackdoorBox)
* [TrojanZoo](https://github.com/ain-soph/trojanzoo)
* [OpenBackdoor](https://github.com/thunlp/OpenBackdoor)
* [Backdoor Toolbox](https://github.com/vtu81/backdoor-toolbox)
* [BackdoorBench](https://github.com/SCLBD/BackdoorBench)
* [backdoors101](https://github.com/ebagdasa/backdoors101)
* [ART](https://github.com/Trusted-AI/adversarial-robustness-toolbox)

**Dissertation and Thesis**

* Defense of Backdoor Attacks against Deep Neural Network Classifiers. [[pdf]](https://etda.libraries.psu.edu/files/final_submissions/26656)
  + Zhen Xiang. *Ph.D. Dissertation at The Pennsylvania State University*, 2022.
* Towards Adversarial and Backdoor Robustness of Deep Learning. [[link]](http://rave.ohiolink.edu/etdc/view?acc_num=case1654872780807815)
  + Yifan Guo. *Ph.D. Dissertation at Case Western Reserve University*, 2022.
* Toward Robust and Communication Efficient Distributed Machine Learning. [[pdf]](https://www.proquest.com/docview/2572595657)
  + Hongyi Wang. *Ph.D. Dissertation at University of Wisconsin–Madison*, 2021.
* Towards Robust Image Classification with Deep Learning and Real-Time DNN Inference on Mobile. [[pdf]](https://www.proquest.com/docview/2572970976)
  + Pu Zhao. *Ph.D. Dissertation at Northeastern University*, 2021.
* Countermeasures Against Backdoor, Data Poisoning, and Adversarial Attacks. [[pdf]](https://www.proquest.com/docview/2572565404)
  + Henry Daniel. *Ph.D. Dissertation at University of Texas at San Antonio*, 2021.
* Understanding and Mitigating the Impact of Backdooring Attacks on Deep Neural Networks. [[pdf]](https://www.proquest.com/docview/2555308945?pq-origsite=gscholar&fromopenview=true)
  + Kang Liu. *Ph.D. Dissertation at New York University*, 2021.
* Un-fair trojan: Targeted Backdoor Attacks against Model Fairness. [[pdf]](https://digitalcommons.njit.edu/cgi/viewcontent.cgi?article=3010&context=theses)
  + Nicholas Furth. *Master Thesis at New Jersey Institute of Technology*, 2022.
* Check Your Other Door: Creating Backdoor Attacks in the Frequency Domain. [pdf]
  + Hasan Abed Al Kader Hammoud. *Master Thesis at King Abdullah University of Science and Technology*, 2022.
* Backdoor Attacks in Neural Networks. [[link]](https://repository.tudelft.nl/islandora/object/uuid:b830b4a2-b700-4c93-8d1d-88dc0191c468?collection=education)
  + Stefanos Koffas. *Master Thesis at Delft University of Technology*, 2021.
* Backdoor Defenses. [[pdf]](https://repositum.tuwien.at/bitstream/20.500.12708/18831/1/Milakovic%20Andrea%20-%202021%20-%20Backdoor%20defenses.pdf)
  + Andrea Milakovic. *Master Thesis at Technischen Universität Wien*, 2021.
* Geometric Properties of Backdoored Neural Networks. [[pdf]](https://www2.eecs.berkeley.edu/Pubs/TechRpts/2021/EECS-2021-78.pdf)
  + Dominic Carrano. *Master Thesis at University of California at Berkeley*, 2021.
* Detecting Backdoored Neural Networks with Structured Adversarial Attacks. [[pdf]](https://www2.eecs.berkeley.edu/Pubs/TechRpts/2021/EECS-2021-90.pdf)
  + Charles Yang. *Master Thesis at University of California at Berkeley*, 2021.
* Backdoor Attacks Against Deep Learning Systems in the Physical World. [[pdf]](https://newtraell.cs.uchicago.edu/files/ms_paper/ewillson.pdf)
  + Emily Willson. *Master Thesis at University of Chicago*, 2020.