# Recent Advances in Neural Network Verification

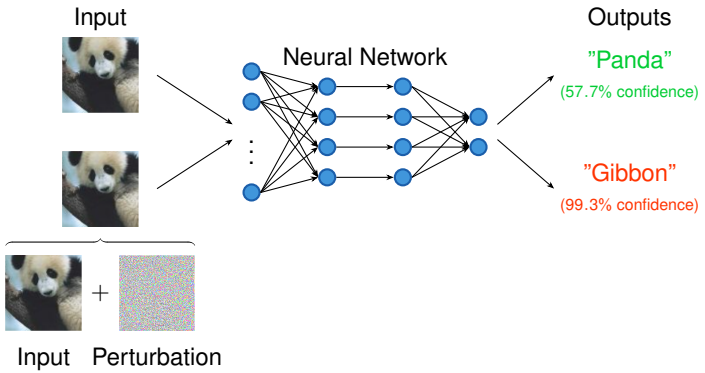
Lukas Koller

Prof. Dr.-Ing. Matthias Althoff
Cyber-Physical Systems Group
Technische Universität München

July 7, 2025

# Neural Networks and Adversarial Attacks

*Adversarial Attack:* Small carefully chosen input perturbation that leads to a misclassification[1].

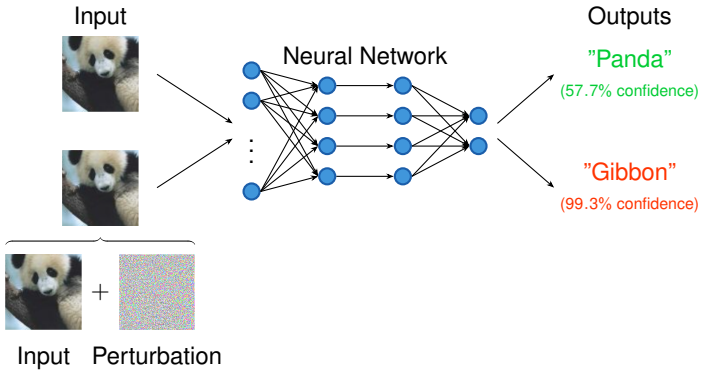



To avoid unexpected behaviour, we have to verify neural networks!

[1] Goodfellow et al., "Explaining and Harnessing Adversarial Examples"

# Neural Networks and Adversarial Attacks

*Adversarial Attack:* Small carefully chosen input perturbation that leads to a misclassification[1].




To avoid unexpected behaviour, we have to verify neural networks!

[1] Goodfellow et al., "Explaining and Harnessing Adversarial Examples"

# Neural Network Verification

**Your tasks:** Review and compare recent neural network verification approaches, e.g. [2], [3], [4], [5], [6].

**Interested? Contact me!**

Lukas Koller
lukas.koller@tum.de

[2] Zhou et al., “Scalable Neural Network Verification with Branch-and-bound Inferred Cutting Planes”

[3] Lopez et al., “NNV 2.0: the neural network verification tool”

[4] Wu et al., “Marabou 2.0: a versatile formal analyzer of neural networks”

[5] Duong et al., “Harnessing neuron stability to improve dnn verification”

[6] Lemesle et al., “Neural Network Verification with PyRAT”