

# ADVERSARIAL ROBUSTNESS OF DEEP NEURAL NETWORKS

**Márk Jelasity**

University of Szeged, Szeged, Hungary

Deep neural networks (DNNs) are the foundations of AI systems, yet they are known to have several vulnerabilities that are unsolved to this day. A particularly interesting vulnerability is that it is possible to construct inputs to these networks that are extremely close to natural inputs (eg. invisible perturbations to images) yet result in completely unexpected behavior. In the talk I will give several examples of this problem, and I will also present some of our own results in the area. I will briefly touch on the formal verification of neural networks from the point of view of adversarial robustness, and I will also discuss some applications like attacking semantic image segmentation networks and attacking entire ensembles of models.