

Model Manipulations: How to fool Neural Networks

Verena Heusser, *Student, KIT*

Abstract—

*Index Terms—*Explainability, Adversarial Methods, Neural Networks, White Models

I. INTRODUCTION