

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«МИРЭА – Российский технологический университет»**

**РТУ МИРЭА**

----------------------------------------------------------------------------------------------------

Институт кибербезопасности и цифровых технологий

Кафедра КБ-4 «Интеллектуальные системы информационной безопасности»

----------------------------------------------------------------------------------------------------

**Отчёт по лабораторной работе № 1**

По дисциплине

«Анализ защищенности систем искусственного интеллекта»

Студент Невретдинов Руслан

Группа ББМО-01-22

Вариант 34

Работу проверил

Спирин А.А.

Москва, 2023

**Цель лабораторной работы**

В данной лабораторной работе необходимо выявить закономерность или обнаружить отсутствие влияния параметра fgsm\_eps для сетей FC LeNet на датасете MNIST и NiN LeNet на датасете CIFAR.

**Результат эксперимента**

Для выполнения данной работы был использован язык Python. Реализация данного эксперимента представлена в файле AZSII\_lab\_1\_Nevretdinov\_RF. Ipynb.

Рассмотрим найденные закономерности для сети FC LeNet:

* При значениях fgsm\_eps равных 0.001 до 0.02 ошибка классификации остаётся низкой и не превышает 11%. Данное наблюдение свидетельствует о том, что при таких значениях сеть является относительно устойчивой к атакам.
* При значениях fgsm\_eps равных 0.5 до 0.9 ошибка классификации значительно увеличивается, что говорит о нарушении стойкости сети к атакам.
* При значении fgsm\_eps равным 10, происходит большое искажение входных данных, очевидно при этом ошибка классификации значительно высокая, из-за чего выполнение задачи классификации становится невозможным.

Рассмотрим найденные закономерности для сети NiN LeNet:

* При значениях fgsm\_eps равных 0.001 до 0.02 ошибка классификации остаётся достаточно низкой, хоть и значительно выше, чем у предыдущей сети.
* При значениях fgsm\_eps равных 0.5, 0.9 и 10 ошибка классификации значительно увеличивается.

**Заключение**

В результате выполнения лабораторной работы были выявлены закономерности влияния параметра fgsm\_eps для сетей FC LeNet на датасете MNIST и NiN LeNEt на датасете CIFAR. На основе данных закономерностей удалось сформировать выводы, что при небольших значениях fgsm\_eps сохраняется стойкость сетей к атакам и ошибки классификации остаются низкими, однако при увеличении значений fgsm\_eps, сети становятся более уязвимыми к атаке и возникает большое количество ошибок классификации.