# 1 ПРАКТИЧЕСКИЕ РАБОТЫ

Методические материалы по практическим работам

А.Е. Сулавко

Защищенное исполнение доверенного интеллекта

## Общие положения и методика оценивания результатов

В процессе обучения по дисциплине «Защищенное исполнение искусственного интеллекта» студент обязан выполнить 7 практических работ. Под выполнением практических работ подразумевается подготовка к работе, проведение экспериментальных исследований, подготовка отчета и его защита на практическом занятии. Выполнение практических работ студентами осуществляется в бригадах до 2 человек. Оформление отчета студентами осуществляется индивидуально в соответствии с принятыми в СПбГЭТУ правилами оформления студенческих работ. Отчет оформляется после выполнения экспериментальных исследований и представляется преподавателю на проверку. После проверки отчет либо возвращается (при наличии замечаний) на доработку, либо подписывается к защите. Практические работы защищаются студентами индивидуально. Каждый студент получает вопрос по теоретической и практической части, после чего ему предоставляется время для подготовки ответа. На защите практической работы студент должен показать: понимание и умение объяснять особенности применяемых методов, возможные области их применения и т.д., прогнозировать реакции исследуемого объекта на различные воздействия, навыки и умения, приобретенные при выполнении практических работы. При обсуждении ответа преподаватель может задать несколько уточняющих вопросов. В случае если студент демонстрирует достаточное знание вопроса, работа считается зачтенной. Текущий контроль включает в себя выполнение, сдачу в срок отчетов и их защиту по всем практическим работам, по результатам которой студент получает допуск к экзамену. Критерием оценки работы на практическом занятии является оценка, выставляемая по 5ти балльной шкале в соответствии со следующими критериями: оценка в 5 баллов выставляется при отличном выполнении задания, то есть при наличии полных (с детальными пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов и высококачественного оформления работы. Оценка в 4 балла выставляется при правильном выполнении задания, то есть при наличии полных (с пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов. Оценка в 3 балла выставляется при наличии отдельных неточностей в ответах (включая грамматические ошибки) или неточностях в решении задач непринципиального характера (описки и случайные ошибки арифметического характера). Оценка в 2 и ниже баллов выставляется в случаях, когда в ответах и в решениях задач имеются неточности и ошибки, свидетельствующие о недостаточном понимании вопросов и требующие дополнительного обращения к тематическим материалам.

## Содержание отчетов

* Титульный лист
* Цель работы
* Формулировка задания
* Краткое описание выполнения основных этапов работы
* Выводы по работе

## Практические работы

### Области применения защищенного режима исполнения нейросетевых алгоритмов. Обсуждение индивидуальных заданий

**Цель работы:** изучениеосновных областей применения защищенного режима исполнения нейросетевых алгоритмов с опорой на теоретический лекционный материал.

**Задание:**

Провести анализ индивидуального домашнего задания (ИДЗ), декомпозицию задач и построить план выполнения ИДЗ.

**Контрольные вопросы:**

1. Невозможность совершения каких действий подразумевает под собой «защищенное исполнение» нейросетевых алгоритмов?
2. Какие параметры необходимо шифровать с помощью криптографического ключа с целью защиты данных, хранящихся в памяти ИИ?
3. В чем заключается «атака одного бита» и каковы ее отличия от атаки «ключ под ковриком»?
4. Каким образом можно обеспечить архитектурную безопасность ИИ?
5. Перечислите основные требования к приложению искусственного интеллекта, реализующего защиту от состязательных атак.

### Оценка энтропии НПБК, обучаемого по ГОСТ Р 52633.5. Моделирование атаки извлечения знаний на НПБК

**Цель работы:** получение навыков обучения нейросетевого преобразователя биометрия-код (согласно ГОСТ Р 52633.5) и моделирования атаки извлечения знаний из него; изучение процесса оценки энтропии обученного НПБК.

**Задание:**

Провести эксперимент по моделированию атаки извлечения знаний на НПБК, построенном по ГОСТ Р 52633.5. Провести оценку энтропии выходных значений НПБК. Для выполнения задания может быть использован программный пакет AIC desktop.

**Контрольные вопросы:**

1. Сколько образов «Чужих» и «Свой» необходимо для обучения НПБК по ГОСТ Р 52633.5?
2. В чем заключается атака Маршалко? Каким образом необходимо настроить НПБК, чтобы избежать уязвимости в виде атаки Маршалко?
3. С какой целью при обучении НПБК используется активный параметр?
4. Каким образом связаны между собой энтропия выходных кодов НПБК и возможность проведения атаки извлечения знаний?
5. Какие выводы можно сделать исходя из анализа энтропии обученного НПБК?

### Защита нейросетевых контейнеров от извлечения знаний. Оценка энтропии НПБК в защищенном исполнении

**Цель работы:** получение навыков проектирования защищенных нейросетевых контейнеров НПБК и оценки их энтропии.

**Задание:**

Провести эксперимент по моделированию атаки извлечения знаний на НПБК, спроектированном в защищенном исполнении. Провести оценку энтропии выходных значений НПБК и сравнить полученные результаты со значениями энтропии выходных значений НПБК из практического задания №2. Для выполнения задания может быть использован программный пакет AIC desktop.

**Контрольные вопросы:**

1. Что такое режим защищенного нейросетевого контейнера (ЗНК)? Объясните принцип работы данного режима.
2. В чем заключается атака извлечения знаний, и каковы особенности реализации режима ЗНК от данной атаки?
3. Какие изменения в значении энтропии выходных значений НПБК, спроектированном в защищенном исполнении, наблюдаются в сравнении со значениями энтропии выходных значений обычного НПБК?
4. Что понимают под «симметрией» стабильности ответов образов «Чужих» относительно образа «Свой»?
5. Какой тип нейронных сетей используется для реализации режима ЗНК в НПБК, обученных по ГОСТ Р 52633.5?

### Защищенный режим исполнения сетей квадратичных нейронов

**Цель работы:** получение навыков проектирования и реализации защищенного режима исполнения сетей квадратичных нейронов.

**Задание:**

Провести эксперимент по моделированию защищенного исполнения НПБК на базе квадратичных нейронов. Провести оценку энтропии выходных значений НПБК и сравнить полученные результаты со значениями энтропии выходных значений НПБК из практического задания №2. Для выполнения задания может быть использован программный пакет AIC desktop.

**Контрольные вопросы:**

1. Что представляет собой процесс обучения квадратичного нейрона и чем данный процесс отличается от обучения линейных нейронов классического НПБК?
2. Каким образом осуществляется оценка вероятности появления ошибок первого рода при обучении квадратичных нейронов?
3. Каким образом осуществляется оценка вероятности появления ошибок второго рода при обучении квадратичных нейронов?
4. Как обучаются сети квадратичных нейронов?
5. В сокрытии каких параметров заключается защищенное исполнение сетей квадратичных нейронов?

### Защищенный режим исполнения сетей корреляционных нейронов

**Цель работы:** получение навыков проектирования и реализации защищенного режима исполнения сетей корреляционных нейронов.

**Задание:**

Провести эксперимент по моделированию защищенного исполнения НПБК на базе корреляционных нейронов. Провести оценку энтропии выходных значений НПБК и сравнить полученные результаты со значениями энтропии выходных значений НПБК из практического задания №2. Для выполнения задания может быть использован программный пакет AIC desktop.

**Контрольные вопросы:**

1. Что происходит с пространством признаков при наличии корреляционных связей между нейронами? Раскройте особенности данного процесса.
2. Для чего в сетях корреляционных нейронов вводятся нормирующие коэффициенты?
3. Каким образом обеспечивается дифференциальная конфиденциальность данных классов при проектировании сетей корреляционных нейронов?
4. Что понимается под «спрямляющим» мета-пространством и каким образом данное пространство отражено в процессе проектирования защищённого исполнения сетей корреляционных нейронов.
5. Перечислите основные принципы построения моделей корреляционных нейронов Байеса-Минковского для защищенного исполнения искусственного интеллекта.

### Защищенный режим исполнения гибридных нейронных сетей

**Цель работы:** получение навыков проектирования и реализации защищенного режима исполнения гибридных нейронных сетей.

**Задание:**

Провести эксперимент по моделированию защищенного исполнения гибридной нейронной сети. Для выполнения задания может быть использован программный пакет AIC desktop.

**Контрольные вопросы:**

1. В чем заключается теорема Кондорсе, и каким образом данный принцип используется при построении гибридный нейронных сетей?
2. В чем заключаются особенности построения гибридных нейронных сетей в защищенном исполнении?
3. Какая сеть может называться «гибкой» нейронной сетью?
4. Каким образом настраиваются нейроны, входящие в одну категорию внутри гибридной нейронной сети?
5. В чем заключается особенность построения многослойной гибридной нейронной сети, в том числе в защищенном исполнении?

### Защита индивидуальных домашних заданий

**Цель работы:** защита ИДЗ

**Задание:** ИДЗ связано с разработкой архитектуры искусственного интеллекта (ИИ), исполняемого в защищенном режиме. Каждое задание связано с определенной прикладной областью или сферой применения ИИ и выдается преподавателем. Студент должен проанализировать материалы лекций, основную и дополнительную литературу, интернет источники, составить расширенную постановку задачи и согласовать ее с преподавателем.

Отчет по выполненному ИДЗ должен содержать аннотированное описание проведенной работы, включая следующие элементы:

* Техническую и архитектурную документацию, объемом не менее 10 страниц, включая все функциональные схемы, графики, блок-схемы алгоритмов, формулы и описание модели ИИ, а также принципов защищенного исполнения нейросетевых алгоритмов. Дополнительным преимуществом, которое будет учитываться преподавателем при выставлении итоговой оценки, будет публикация научной статьи по теме ИДЗ;
* Расширенное описание назначения продукта (проблема и решение);
* Презентация доклада (на 510 слайдов).

Отчет выполняется в электронном виде.

**Контрольные вопросы:**

1. Каким угрозам информационной безопасности (ИБ) и рискам подвержен разрабатываемый ИИ?
2. Какие методы снижения рисков и защиты от актуальных угроз ИБ были выбраны при проектировании архитектуры ИИ?
3. Каким образом планируется реализация режима защищенного исполнения ИИ?
4. Какие программные продукты/фреймворки/библиотеки машинного обучения/платформы Вы собираетесь использовать?
5. Какие данные планируется использовать для обучения и тестирования ИИ, и на какие целевые показатели надежности Вы ориентируетесь?