

**09.04.04 Программная инженерия**  
**«Искусственный интеллект и инженерия данных»**

Программа ориентирована на получения фундаментальных знаний в области разработки систем искусственного интеллекта и обработки больших массивов данных. Выпускник получает навыки в сфере обработки естественного языка, компьютерного зрения и анализа временных рядов.

Региональный партнер: Челябинский государственный университет

IT-компания, выполнявшая экспертизу: ООО «ТРИДИВИ»

Экспертное заключение подписал: ведущий разработчик алгоритмов искусственного интеллекта Хайруллин Р.Г.

**Рабочие программы профильных дисциплин**

Разработка интеллектуальных систем на языке R	Тема 1. Принципы программирования на языке R	Операции над переменными: Простейшие операции. Логические операции. Математические функции. Тригонометрические функции. Операции над комплексными переменными. Условные операторы и циклы: операторы if, ifelse, for, while, repeat, break, 1 next, switch.	Лекции – 2 Практич. занятия – 4
	Тема 2. Виды данных в языке R, способы ввода и вывода данных	Классы, объекты, типы, структуры данных в языке R: Понятие набора и структуры данных. Вектор, способы задания. Символьные векторы и строки. Числовые и логические векторы. Задание имён элементам векторов. Векторы и индексы. Функция which(). Задание матрицы. Операции над матрицами и индексами. Многомерные массивы. Списки. Факторы и таблицы. Ввод, вывод и импорт данных: Способы получения данных. Клавиатурный ввод. Импорт данных из различных источников. Импорт из файлов CSV, Excel, XML-файлов. Извлечение данных из web-страниц. Импорт данных из баз данных. Функции scan( ), read.table(), read.csv(). Вывод данных. Функции write(), cat(), write.table(), write.csv().	Лекции – 2 Практич. занятия – 4
	Тема 3. Имитационное моделирование и профилирование кода в R	Имитационное моделирование: Функция sample(). Виды распределений и функции для имитации распределения (rnorm(), rpois() и и т.д.). Бутстрэппинг. Моделирование данных с заданными статистическими характеристиками. Применение функции replicate(). Профилирование кода в R: Использование графика пламени. Использование средства просмотра данных. Профилирование времени. Профилирование памяти.	Лекции – 2 Практич. занятия – 4
	Тема 4. Основы статистического анализа в R	Базовый статистический анализ данных: Функции mean(), median(), sd(), prop.table(), var(), IQR(), summary(). Базовые статистические тесты: Тесты Стьюдента, Уилкоксона, ШапироУилкса, ANOVA, хи-квадрат Пирсона, тесты бинарной классификации.	Лекции – 2 Практич. занятия – 4
	Тема 5. Регрессионный анализ данных	Простая парная линейная регрессия: Метод наименьших квадратов, проверка статистических гипотез. Определение параметров с помощью языка R, функции lm и summary, построение графиков для парной зависимости. Множественная линейная регрессия: Применение функции lm и summary. Проблема мультиколлинеарности, применение функции step(). Выявление значимых факторов с помощью функции vif().	Лекции – 2 Практич. занятия – 4

	Тема 6. Визуализация данных	Базовые возможности визуализации данных: Графическое представление данных, графические параметры. Символы и линии. Цвета. Характеристики текста. Размеры диаграмм и полей. Настройка параметров осей и условных обозначений. Опорные линии. Легенда. Аннотации. Объединение диаграмм. Виды диаграмм: Столбчатые диаграммы. Простые диаграммы, составные и диаграммы с группировкой. Диаграммы для средних значений. Оптимизация столбчатых диаграмм. Спинограммы. Круговые диаграммы. Гистограммы. Диаграммы оценки функции плотности. Диаграммы размахов. Точечные диаграммы.	Лекции – 2 Практич. занятия – 4
	Тема 7. Введение в машинное обучение	Метод главных компонент, классификация без обучения, кластерный анализ: Применение библиотеки RGL, пакетов scatterplot3d, rggobi, lattice для визуализации многомерных данных. Функции princomp(), biplot(), loadings(), пакеты ade4 и vegan. Функции daisy(), cmdscale(), hclust(), pvclust(), fanny(), smeans(). Метод опорных векторов (SVR) и деревья решения: Bagging. Случайные леса. Boosting. Важность переменной. Сортировка полей и поддержка векторного классификатора.	Лекции – 2 Практич. занятия – 4
Разработка систем искусственного интеллекта на языке Python	Тема 1. Введение в Python.	Основные конструкции и базовые типы. Интерпретатор Python, байткод. Структуры данных: списки, кортежи, множества, словари - особенности внутренней реализации. Функции. Виртуальное окружение (Virtualenv). Установка и запуск Jupyter Notebook. Работа в среде Jupyter Notebook. Написание функций на Python, работа со структурами данных.	Лекции – 2 Практич. занятия – 4
	Тема 2. ООП и паттерны проектирования на Python	Объектно-ориентированное программирование в Python. Классы в Python. Реализация принципов ООП в Python. Методы и декораторы. Наследование и композиция. Полиморфизм. Объектно-ориентированное проектирование, качество кода, рефакторинг, юнит-тестирование и паттерны в Python.	Лекции – 4 Практич. занятия – 8
	Тема 3. Python для анализа данных	Математические функции в Python. Библиотеки Numpy, Scipy. Обработка данных в библиотеке Pandas. Разведочный анализ данных (EDA). Визуализация данных: библиотеки Matplotlib, Seaborn.	Лекции – 4 Практич. занятия – 8
	Тема 4. Многопоточное и асинхронное программирование	Многопоточное и асинхронное программирование. Процессы и потоки. Синхронизация потоков. Сокеты, клиент-сервер. Библиотека asyncio.	Лекции – 2 Практич. занятия – 4
	Тема 5. Веб-разработка на Python	Python для сбора и обработки данных. Сбор данных со сторонних сайтов. Библиотека Beautiful Soup и работа с API. Слой доступа к данным. Работа с SQL/NoSQL базами данных в Python. Создание веб-интерфейсов с Django и Bootstrap. Знакомство с фреймворком Flask. Обзор веб-фреймворков для Python. Продвинутая инфраструктура веб-разработки: Git Flow, CI-инструменты, веб-сервер Nginx. Создание веб-сервисов с фреймворком fastapi. Реализация deploy-процесса для моделей машинного обучения в веб-сервис.	Лекции – 4 Практич. занятия – 24

Машинное обучение	Тема 1. Введение в машинное обучение.	Примеры задач машинного обучения с учителем и без. Одномерная линейная регрессия и метод максимального правдоподобия. Функция потерь, метод градиентного спуска. Множественная линейная регрессия. Нормализация признаков. Построение нелинейных моделей.	Лекции – 8 Практич. занятия – 6
	Тема 2. Задача классификации.	Задача бинарной классификации. Логистическая регрессия. Сигмоида и логлосс. Задача бинарной классификации. Проблема переобучения. Регуляризация. Гребневая регрессия. Лассо.	Лекции – 6 Практич. занятия – 8
	Тема 3. Метод kNN, деревья решений и ансамблевые методы.	Метод ближайших соседей kNN и его модификации. Деревья решений. Ансамбли деревьев решений.	Лекции – 6 Практич. занятия – 4
	Тема 4. Метод опорных векторов	Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Зазор между классами. Функции ядра (kernel trick). Нелинейный SVM.	Лекции – 4 Практич. занятия – 6
	Тема 5. Методы понижения размерности и визуализации данных	Методы понижения размерности данных - метод главных компонент (PCA) и метод t-SNE. Сжатие и визуализация данных.	Лекции – 4 Практич. занятия – 4
	Тема 6. Дополнительные вопросы обучения моделей машинного обучения.	Организация надежной валидации (dataset split, cross-validation), анализ learning curves. Метрики качества моделей, отбор признаков.	Лекции – 4 Практич. занятия – 4
Глубокие нейронные сети	Тема 1. Введение. Модель нейронной сети	Биологический нейрон человеческого мозга. Понятие искусственного нейрона и искусственной нейронной сети. Сферы применения искусственных нейронных сетей. Модель нейрона Мак-Каллока–Питса. Персептрон. Глубокая нейронная сеть прямого распространения. Вычислительная мощность персептрона.	Лекции – 4 Практич. занятия – 2
	Тема 2. Градиентный спуск	Возможность обучения нейронной сети. Трудности с обучением персептронных сетей. Сигмоидальный нейрон (сигмоид). Модель обучения нейронной сети. Обучающая и тестовые выборки. Функция потерь (стоимостная функция). Среднеквадратическая ошибка. Задача распознавания рукописных цифр. База данных MNIST. Обучение нейронной сети методом градиентного спуска. Переход к векторным операциям. Алгоритм стохастического градиентного спуска.	Лекции – 6 Практич. занятия – 4
	Тема 3. Метод обратного распространения ошибки	Метод обратного распространения ошибки. Функция потерь на основе перекрестной энтропии. Формулы обратного распространения ошибки для случая перекрестной энтропии.	Лекции – 4 Практич. занятия – 2
	Тема 4. Переобучение и регуляризация. Техники, улучшающие обучение нейронных сетей.	Переобучение. Валидационная выборка. Регуляризация L1 и L2. Прореживание (dropout). Искусственное увеличение обучающей выборки (аугментация данных). Инициализация весов и смещений на основе нормального распределения. Уменьшение скорости обучения. Градиентный спуск на основе импульса. Алгоритмы с адаптивной скоростью обучения: AdaGrad, RMSProp, Adam. ReLU и другие альтернативные функции активации. Использование функции softmax в качестве функции активации выходного слоя.	Лекции – 6 Практич. занятия – 2

	Тема 5. Сверточные нейронные сети	Сверточные нейронные сети (СНС). Слои свертки и пулинга. Построение карт признаков. Ядро свертки. Мах-пулинг и L2-пулинг. СНС для распознавания рукописных цифр. Преимущества СНС. Потенциальные проблемы, связанные с использованием СНС. Сферы применения СНС. Нейробиологические основания сверточных сетей.	Лекции – 4 Практич. занятия – 2
	Тема 6. Рекуррентные нейронные сети и автокодировщики.	Рекуррентные нейронные сети (РНС). Двухнаправленные РНС. Глубокие РНС. Вентильные РНС. Сети долгой краткосрочной памяти (LSTM). Вентильный рекуррентный модуль (GRU).	Лекции – 4 Практич. занятия – 2
	Тема 7. Автокодировщики.	Определение и общая структура автокодировщика. Области применения автокодировщиков. Понижающие автокодировщики. Регуляризованные автокодировщики. Разреженный автокодировщик. Шумоподавляющие автокодировщики. Глубокие автокодировщики.	Лекции – 2 Практич. занятия – 2
	Тема 7. История нейронных сетей и глубокого обучения.	Математическая модель нейрона Мак-Каллока и Питтса (1943). Обучение по Хеббу (1946). SNARC Марвина Минского (1951). Персептрон Розенблатта (1962). Тонкости машинного перевода Ноама Хомского (1954–1964). Критика персептрона (1969). Неокогнитрон (1980). Метод обратного распространения ошибки (1986). Автокодировщики (1986). Рекуррентные сети (1986). Сверточные сети (1989). Модель долгой краткосрочной памяти LSTM (1995). Парадокс Моравека (1988). Использование GPU (2009). Глубокая сверточная сеть распознает дорожные знаки на фотографиях лучше, чем человек (2011). Обучение с подкреплением (2013). Порождающие состязательные сети (2014). Нейронная сеть побеждает чемпиона мира по игре в го (2016). Нейронная сеть управляет автомобилем (2016). Предсказание пространственной структуры белков (2021).	Лекции – 2
Доверенный искусственный интеллект	Тема 1. Нормативно-правовые документы в области применения доверенного ИИ	Нормативно-правовые документы в области применения доверенного ИИ (Ethics guidelines for trustworthy AI – Европейская Комиссия; Trustworthy AI in health – Саудовская Аравия и др.)	Лекции – 2 Практич. занятия – 2
	Тема 2. Предвзятость в системах принятия решений	Методы оценки корректности принятия решений доверенным ИИ	Лекции – 2 Практич. занятия – 2
	Тема 3. Вопросы доверия к большим данным	Алгоритмы оценки корректности больших данных. Методы защиты больших данных.	Лекции – 2 Практич. занятия – 4
	Тема 4. Вопросы доверия к автономным системам ИИ	Основные принципы разработки автономных систем ИИ в различных областях (беспилотный транспорт, робототехника, медицина и др.)	Лекции – 2 Практич. занятия – 2
	Тема 5. Стандартизация в области разработки доверенного ИИ	Стандартизация разработки доверенного ИИ (ISO/IEC AWI 5339; ISO/IEC DTS 4213 и др.)	Лекции – 2 Практич. занятия – 2
	Тема 6. Принципы разработки доверенного ИИ	Основные этапы проектирования, разработки и внедрения доверенного ИИ на реальных примерах	Лекции – 2 Практич. занятия – 2
	Тема 7. Проблемы внедрения и использования систем с использованием доверенного ИИ	Законодательное регулирование и риски внедрения и использования доверенного ИИ	Лекции – 2 Практич. занятия – 2

Анализ естественного языка методами искусственного интеллекта	Тема 1. Введение в обработку естественного языка	Введение в обработку естественного языка (NLP). Основные задачи NLP. Представления текстовых данных. Предобработка текста, лемматизация, стемминг.	Лекции – 2 Практич. занятия – 4
	Тема 2. Классификации текстовых документов на основе частотных мер	Методы машинного обучения для классификации текстовых документов на основе частотных мер. Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов. Мера TF-IDF.	Лекции – 2 Практич. занятия – 4
	Тема 3. Классификация текстов методами машинного обучения и глубоких нейронных сетей	Языковые модели. Нейросетевые модели языка - word2vec, fasttext. Мера семантической близости. Классификация текстов методами машинного обучения и глубокого обучения на основе нейросетевых моделей языка.	Лекции – 2 Практич. занятия – 4
	Тема 4. Кластеризация текстовых документов	Задача кластеризации текстовых документов. Тематическое моделирование. Методы LSA, pLSA. Аддитивная регуляризация тематических моделей в BigARTM.	Лекции – 2 Практич. занятия – 4
	Тема 5. Трансформеры	Механизм внимания. Архитектуры encoder-decoder-attention. Transfer learning в задачах анализа текстов. Архитектуры трансформеров BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерация текстов. Fine-tuning трансформеров.	Лекции – 4 Практич. занятия – 6
	Тема 6. Генеративные нейронные сети в задачах обработки естественного языка	Задачи обработки последовательностей: машинный перевод, автоматическое реферирование, вопросно-ответные системы.	Лекции – 2 Практич. занятия – 6
	Тема 9. Построение диалоговых систем	Построение диалоговых систем. Архитектура диалоговых систем. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM). Сложности построения диалоговых систем. Проектирование UX/UI диалоговых ассистентов в чатах и голосе. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform.	Лекции – 2 Практич. занятия – 4
Компьютерное зрение	Тема 1. Сверточные нейронные сети в задачах компьютерного зрения.	Сверточные нейронные сети в задачах компьютерного зрения. Классификация изображений. Разбор архитектур нейронных сетей VGG, Inception, ResNet, EfficientNet. Transfer learning.	Лекции – 2 Практич. занятия – 4
	Тема 2. Применение библиотек OpenCV и Pillow	Знакомство с OpenCV. Обработка изображений. Знакомство с Pillow.	Практич. занятия – 6
	Тема 3. Сегментация изображений и детекция объектов	Сегментация изображений и детекция объектов. Постановка задач сегментации и детекции. Разбор архитектур FCN и Unet в задачах сегментации. Изучение функций потерь для задачи сегментации. Разбираются одностадийные (SSD, YOLO) и двухстадийные (FASTER RCNN, Mask R-CNN) детекторы в задачах детекции. Изучения функций потерь в задаче детекции.	Лекции – 3 Практич. занятия – 4
	Тема 4. Основы создания наборов данных	Основы создания наборов данных изображений. Предобработка изображений. Инструменты разметки датасета изображений. Аугментация данных.	Лекции – 1 Практич. занятия – 2
	Тема 5. Поиск изображений	Поиск изображений. Постановка задачи обучения по метрике для сравнения схожести двух изображений и поиска изображений по содержанию. Разбор функций потерь contrastive loss, triplet loss, arcface.	Лекции – 2 Практич. занятия – 4

	Тема 6. Распознавание текста на изображении	Постановка задачи распознавания текста на изображении. Разбор алгоритмов CRNN, Attention OCR и Transformer OCR	Лекции – 2 Практич. занятия – 4
	Тема 7. Обнаружение и трекинг объектов в видеопотоке	Обнаружение и трекинг объектов в видеопотоке. Разбор алгоритмов SORT, DeepSORT.	Лекции – 3 Практич. занятия – 4
	Тема 8. Компрессия моделей компьютерного зрения	Компрессия моделей компьютерного зрения Изучения различных способов оптимизации работы моделей: прунинг, перенос знаний, квантизация.	Лекции – 2 Практич. занятия – 2
Нейробайесовские методы в машинном обучении	Тема 1. Стохастический вариационный вывод	Стохастический вариационный вывод. Применение SVI на примере масштабируемых тематических моделей. Дважды стохастический вариационный вывод. Дисперсия стохастических градиентов в примерах	Лекции – 3 Практич. занятия – 4
	Тема 2. Вариационный автокодировщик	Вариационный автокодировщик, нормализующие потоки для вариационного вывода. Репараметризация, IWAE. Оценка отношения плотностей распределений, применение на примере $\alpha$ -GAN.	Лекции – 4 Практич. занятия – 2
	Тема 3. Байесовские нейронные сети	Байесовские нейронные сети Байесовское сжатие нейронных сетей. Тернарные сети, структурный байесовский прунинг. Методы снижения дисперсии в моделях со скрытыми переменными.	Лекции – 4 Практич. занятия – 4
	Тема 4. Генеративно-состязательные сети	Дискриминационные алгоритмы. Генеративные алгоритмы. Проблемы обучения GAN. Реализация сверточной GAN с помощью PyTorch. Модификации GAN.	Лекции – 4 Практич. занятия – 4
Нейросетевые технологии в задачах синтетических медиа	Тема 1. Синтетические медиа	Синтетические медиа: кейсы, тренды и прорывы, проблемы и задачи. Введение в генеративное глубокое обучение. Вариационные автокодировщики. Генеративно-состязательные сети.	Лекции – 4 Практич. занятия – 8
	Тема 2. Анализ аудио. Распознавание и синтез речи	Анализ аудио. Преобразование Фурье. Распознавание и синтез речи. Классификация систем распознавания речи. Архитектуры систем распознавания и синтеза речи (ASR / TTS). Понятия разборчивости и качества речи. Вокодеры. Энкодеры для верификации голоса спикера.	Лекции – 4 Практич. занятия – 4
	Тема 3. Анализ и синтез изображений и видео по содержанию	Анализ и синтез изображений и видео по содержанию. Синтез изображений по заданным параметрам. Генерация изображений.	Лекции – 2 Практич. занятия – 4
	Тема 4. Синтез лица	Синтез лица, создание deepfake. Нейросетевые технологии замены лиц – face swap. Нейросетевые технологии реконструкции лиц – reenactment. Нейросетевые технологии синхронизации лица спикера и речи – lip-sync.	Лекции – 4 Практич. занятия – 8
	Тема 5. Генерация текста	Генерация текста с использованием GAN. Архитектуры генеративных сетей для текста: GPT-2 и GPT-3. Особенности обучения.	Лекции – 2 Практич. занятия – 8
Анализ и прогнозирование временных рядов методами искусственного интеллекта	Тема 1. Поиск подпоследовательностей по образцу	Формальное определение и обзор подходов к решению задачи поиска подпоследовательностей по образцу. Мера динамической трансформации времени (DTW, Dynamic Time Warping). Z-нормализация подпоследовательностей. Техника нижних границ: границы LBKim, LBKeogh, LBKeoghEC, каскадное применение нижних границ.	Лекции – 2 Практич. занятия – 4

	Тема 2. Поиск аномалий во временных рядах	Понятия выброса и диссонанса. Статистические методы поиска выбросов. Кусочно-агрегатное сжатие (PAA, Piecewise Aggregate Approximation), Символьно-агрегатное кодирование (SAX, Symbolic Aggregate Approximation). Алгоритм HotSAX. Понятие диапазонного диссонанса. Алгоритмы DRAG и MERLIN	Лекции – 4 Практич. занятия – 8
	Тема 3. Матричный профиль временного ряда и примитивы анализа данных на его основе	Понятие матричного профиля. Алгоритм SCAMP вычисления матричного профиля. Поиск мотивов. Мера схожести MPdist. Поиск сниппетов (типичных подпоследовательностей). Поиск цепочек (эволюционирующих шаблонов).	Лекции – 4 Практич. занятия – 10
	Тема 4. Восстановление пропусков и прогноз значений временного ряда	Аналитические алгоритмы HotDeck, kNN, REBOM, ТКСМ. Модель ARIMA. Нейросетевые методы прогнозирования временных рядов: сеть с долговременной и кратковременной памятью (Long short term memory, LSTM), сеть с управляемыми рекуррентными блоками (Gated Recurrent Units, GRU)	Лекции – 4 Практич. занятия – 10
Технологии распределенной обработки данных	Тема 1. Введение в большие данные	Понятие больших данных. Проблема больших данных. Система хранения больших данных. Распределенная обработка больших данных. Введение в репликацию и секционирование данных. Теорема CAP. SQL, NoSQL и NewSQL-решения. Структуры для хранения больших данных. Хеш-индексы. SS-таблицы. LSM-деревья. B-деревья.	Лекции – 4
	Тема 2. Секционирование и репликация данных	Секционирование. Виды секционирования: по диапазонам значений ключа, по хешу ключа. Добавление/удаление секций, методы перебалансировки. Репликация. Виды репликации: синхронная, асинхронная, полусинхронная репликация. Репликация с одним ведущим узлом. Добавление узлов в систему. Обработка сбоев узлов. Журнал репликации. Задержка репликации. Репликация с несколькими ведущими узлами. Репликация без ведущего узла. Чтение и запись по кворуму. Алгоритм «векторные часы». Согласованность. Виды согласованности: строгая (Strong Consistency), конечная (Eventual Consistency), согласованное префиксное чтение (Consistent Prefix), с ограниченным устареванием (Bounded Staleness), монотонные чтения (Monotonic Reads, Session guarantee), чтение своих записей (Read My Writes).	Лекции – 8
	Тема 3. Введение в Hadoop	Введение в платформу Hadoop. Примеры систем на базе Hadoop. Распределенная файловая система Hadoop (HDFS). Технология MapReduce.	Лекции - 6 часов Практич. занятия - 8 часов
	Тема 4. Экосистема Hadoop	Экосистема Hadoop. Обработка данных в Hadoop с помощью Pig и СУБД Apache Hive. Архитектура MapReduce 2.0. Планирование и управление ресурсами с помощью Hadoop YARN. Координация распределенных приложений с Zookeeper. Брокер сообщений Apache Kafka.	Лекции - 7 часов Практич. занятия - 4 часа
	Тема 5. Анализ данных в Hadoop	Apache Spark. Библиотека машинного обучения MLlib.	Лекции - 3 часа Практич. занятия - 4 часа

	Тема 6. NoSQL-решения	Классификация NoSQL-решений. Хранилища "ключ-значения". Документоориентированные хранилища. Хранение в виде семейства столбцов. Графовые СУБД. Модель графа свойств (property graph). Модель хранилища тройных кортежей (triple-store). Язык запросов Cypher. Фреймворк Pregel. Введение в СУБД HBase.	Лекции – 4
--	-----------------------	--	------------