Федеральное государственное бюджетное учреждение науки
Институт машиноведения им. А.А. Благонравова
Российской академии наук
(ИМАШ РАН)

Рабочая программа дисциплины

«Современные архитектуры нейронных сетей»

Направление подготовки:	02.04.03 Математическое обеспечение и администрирование информационных систем
Профиль подготовки:	Искусственный интеллект и интеллектуальная обработка данных
Квалификация:	Магистр
Форма обучения	Очная

Москва, 2026 г.

1. Аннотация к дисциплине

1.1.	Составлена в соответствии с ФГОС ВО - магистратура по направлению подготовки 02.04.03 Математическое обеспечение и администрирование информационных систем, утвержденным приказом Минобрнауки от 23.08.2017г. № 812.
1.2.	Дисциплина посвящена изучению эволюции и актуальных архитектур глубокого обучения, используемых для обработки структурированных и неструктурированных данных (изображения, текст, графовые данные, временные ряды). В курсе рассматриваются: полносвязные (MLP), сверточные (CNN), рекуррентные (RNN, LSTM, GRU), трансформеры (BERT, GPT), генеративно-состязательные сети (GAN), вариационные автокодировщики (VAE) и графовые нейронные сети (GNN). Особое внимание уделяется вопросам масштабирования, оптимизации (смешанная точность, распределенное обучение) и промышленного внедрения моделей (ONNX, TensorRT, MLOps).
1.3.	Курс ориентирован на подготовку магистров к решению прикладных задач анализа больших данных и разработки интеллектуальных систем на основе глубокого обучения.

2. Место дисциплины в структуре ООП

2.1.	Цикл (раздел) ООП: Б1.В.ОД.3
2.2.	1 курс, 1 семестр

3. Цель освоения дисциплин

3.1.	Формирование у обучающихся необходимых компетенций для успешного освоения образовательной программы

4. Задачи освоения дисциплины

4.1.	- сформировать знания о теоретических основах глубокого обучения (градиентный спуск, проблема взрыва/исчезновения градиента, функции активации); - сформировать знания об архитектурах CNN, RNN, Transformer, GAN, VAE, GNN и принципах их функционирования; - сформировать знания о механизме внимания (self-attention, cross-attention); - сформировать знания о методах ускорения обучения (BatchNorm, оптимизаторы AdamW, mixed precision); - сформировать знания о подходах к распределенному обучению (DataParallel, DistributedDataParallel); - сформировать знания о методах сжатия моделей (Quantization, Pruning) и инференс-оптимизациях (ONNX, TensorRT); - сформировать знания о принципах MLOps (версионирование, мониторинг, развертывание); - сформировать навыки реализации и обучения современных нейросетевых архитектур на PyTorch/TensorFlow; - сформировать навыки тонкой настройки (fine-tuning) предобученных моделей (BERT, ResNet) под конкретные задачи; - сформировать навыки применения GNN для анализа социальных и графовых данных; - сформировать навыки построения генеративных моделей (GAN/VAE) для синтеза данных; - сформировать навыки оптимизации инференса с помощью ONNX Runtime и TensorRT; - сформировать навыки развертывания моделей в виде REST API с использованием Docker и FastAPI; - сформировать навыки диагностики проблем обучения (мод коллапс GAN, переобучение, градиентный взрыв). - сформировать умение выбирать архитектуру нейронной сети в зависимости от типа данных и бизнес-задачи; - сформировать умение обосновывать выбор гиперпараметров (число слоев, фильтров, скорость обучения);

	- сформировать умение оценивать вычислительные затраты и масштабировать обучение на несколько GPU; - сформировать умение внедрять обученные модели в промышленную эксплуатацию с мониторингом качества; - сформировать умение адаптировать современные архитектуры для анализа больших данных в своей предметной области.

5. Перечень планируемых результатов обучения, соотнесенных с планируемыми результатами освоения образовательной программы

ПК-1	**Создание информационных технологий нового поколения**
ПК-1.2	Управление разработкой продуктов, услуг и решений на основе больших данных
ПК-1.3	Разработка и внедрение новых методов и технологий исследования больших данных

Код индикатора компетенции	Планируемые результаты обучения по дисциплине
ПК-1.2	*Знать:* - современные нейросетевые архитектуры (CNN, Transformer, GNN, GAN) и критерии их выбора в зависимости от типа больших данных (изображения, текст, графы, временные ряды); - методы оценки качества и вычислительной эффективности моделей при работе с большими объемами данных; - принципы построения пайплайнов обработки больших данных с использованием глубокого обучения (ETL, батчирование, шардирование). *Уметь:* - обосновывать выбор конкретной архитектуры для бизнес-задачи на основе анализа данных и требований к продукту; - оценивать временные и вычислительные затраты обучения и инференса на реальных кластерах данных; - организовывать эксперименты по сравнению нескольких архитектур для выбора лучшего решения. *Иметь практический опыт:* - управления процессом разработки: от постановки задачи до внедрения обученной модели в пилотный продукт; - настройки гиперпараметров и распределенного обучения на нескольких GPU для обработки датасетов терабайтного размера; - применения MLOps-практик (версионирование данных, модели, экспериментов) при разработке продуктов на больших данных.
ПК-1.3	*Знать:* - передовые архитектуры нейронных сетей (Transformer с разреженным вниманием, Graph Transformer, диффузионные модели, GAN с усилением); - методы сжатия и ускорения моделей (Quantization-Aware Training, Pruning, Knowledge Distillation) для внедрения на граничных устройствах; - принципы построения гибридных систем (например, нейро-символьный подход) для анализа больших слабоструктурированных данных. *Уметь:* - адаптировать и дорабатывать существующие нейросетевые архитектуры под специфику новых типов больших данных (например, пространственно-временные, мультимодальные); - проводить сравнительный анализ существующих методов и предлагать улучшения (модификация функции потерь, архитектуры, метода оптимизации); - развертывать исследовательские прототипы в экспериментальной среде и оценивать их масштабируемость на реальных больших данных. *Иметь практический опыт:* - разработки и реализации новой нейросетевой архитектуры (или значительной

	модификации существующей) для нестандартной задачи анализа больших данных; - создания пайплайна обучения и оценки с поддержкой mixed precision и распределенных вычислений на кластере; - оптимизации инференса для промышленного внедрения (TensorRT, ONNX, аппаратные ускорители) с обеспечением SLAs по latency и throughput.

6. Тематический план

Наименование темы/раздела	Виды занятий и трудоемкость (в часах)					Формы контроля Форма ПА
	Всего	Л	ПЗ	ЛЗ	СР	
Тема 1. Эволюция архитектур и основы глубокого обучения	**24**	4	2	2	16	Отчет по лабораторному практикуму/15 Отчет по практикуму по решению задач /10
Тема 2. Современные архитектуры для обработки последовательностей и внимания	**2**	4	4	2	16	Отчет по лабораторному практикуму/15 Отчет по практикуму по решению задач /10
Тема 3. Генеративные модели и работа с неструктурированными данными	**30**	4	4	6	16	Отчет по лабораторному практикуму/15 Отчет по практикуму по решению задач /10
Тема 4. Масштабирование, оптимизация и внедрение моделей	**28**	2	4	4	18	Отчет по лабораторному практикуму/15 Отчет по практикуму по решению задач /10
Всего часов	**108**	**14**	**14**	**14**	**66**	100 ТКУ
Контроль, час	**-**					Зачет
Объем дисциплины (в академических часах)	108					
Объем дисциплины (в зачетных единицах)	3					

7. Содержание дисциплины

Наименование темы/раздела	Краткое содержание	Объем часов
Тема 1. Эволюция архитектур и основы глубокого обучения	Переход от классического машинного обучения к глубокому. Проблема градиентного взрыва/исчезновения. Архитектуры прямого распространения (MLP). Введение в сверточные нейронные сети (CNN): свертка, пулинг, рецептивное поле. Рекуррентные нейронные сети (RNN) и их модификации (LSTM, GRU).	4
Тема 2. Современные архитектуры для обработки последовательностей и внимания	Механизм внимания (Attention) как решение проблем RNN. Архитектура Transformer (энкодер, декодер, self-attention). BERT, GPT – принципы предобучения и тонкой настройки.	4
Тема 3. Генеративные модели и работа с неструктурированными данными	GAN (генератор, дискриминатор, проблемы обучения, мод коллапс). VAE. Графовые нейронные сети (GNN, GCN, GraphSAGE) для анализа связей.	4

Тема 4. Масштабирование, оптимизация и внедрение моделей	Пакетная нормализация, AdamW, mixed precision. Распределенное обучение на нескольких GPU/узлах. Quantization, Pruning, TensorRT, ONNX Runtime. MLOps: версионирование (DVC), мониторинг, развертывание.	2

8. Самостоятельная работа студентов

Наименование темы/раздела	Виды самостоятельной работы	Трудоемкость в часах
Тема 1. Эволюция архитектур и основы глубокого обучения	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	16
Тема 2. Современные архитектуры для обработки последовательностей и внимания	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	16
Тема 3. Генеративные модели и работа с неструктурированными данными	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	16
Тема 4. Масштабирование, оптимизация и внедрение моделей	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	18

9. Тематика практических и лабораторных занятий

Наименование темы/раздела	Вид занятия	Краткое содержание занятия	Объем часов
Тема 1. Эволюция архитектур и основы глубокого обучения	практикум по решению задач	Решение задач	2
Тема 1. Эволюция архитектур и основы глубокого обучения	лабораторный практикум	Выполнение заданий в лабораторных условиях	2
Тема 2. Современные архитектуры для обработки последовательностей и внимания	практикум по решению задач	Решение задач	4
Тема 2. Современные архитектуры для обработки последовательностей и внимания	лабораторный практикум	Выполнение заданий в лабораторных условиях	2
Тема 3. Генеративные модели и работа с неструктурированными данными	практикум по решению задач	Решение задач	4
Тема 3. Генеративные модели и работа с неструктурированными данными	лабораторный практикум	Выполнение заданий в лабораторных условиях	6
Тема 4. Масштабирование, оптимизация и внедрение моделей	практикум по решению задач	Решение задач	4
Тема 4. Масштабирование, оптимизация и внедрение моделей	лабораторный практикум	Выполнение заданий в лабораторных условиях	4

10. Фонд оценочных средств

Оценочные материалы для проведения текущего контроля успеваемости, промежуточной аттестации, а также порядок начисления баллов и критерии оценивания представлены в комплекте оценочных материалов (КОМ) по дисциплине.

11. Учебно-методическое и информационное обеспечение дисциплины

9.1. Основная литература

Авторы	Заглавие	Издательство, год	Эл. адрес
Воронов, М. В.	Системы искусственного интеллекта: учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. - 2-е изд., перераб. и доп.	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/588642
Зыков, С. В.	Архитектура информационных систем. Основы проектирования: учебник для вузов	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/590259
Толстобров, А. П.	Архитектура ЭВМ: учебник для вузов	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/583536

9.2. Дополнительная литература

Авторы	Заглавие	Издательство, год	Эл. адрес
Федоров, Д. Ю.	Программирование на Python: учебник для вузов	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/600874
Черпаков, И. В.	Алгоритмизация и программирование на Python: учебник для вузов	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/582412

12. Материально-техническое обеспечение дисциплины

13. Наименование помещения	Перечень основного оборудования
Учебная аудитория для проведения лекционных и практических занятий, консультаций, текущего контроля, промежуточной аттестации.	Мультимедийное оборудование (проектор, экран), маршрутизатор/роутер, мебель
Компьютерный класс для проведения лабораторных занятий, консультаций, текущего контроля, промежуточной аттестации.	Компьютерная техника с возможностью подключения к Интернету, маршрутизатор/роутер, мультимедийное оборудование (проектор, экран), мебель.
Помещения для самостоятельной работы обучающихся	Мебель, компьютерная техника с возможностью подключения к сети Интернет и обеспечением доступа в электронную информационно-образовательную среду организации, маршрутизатор/роутер.