



Бухановский Александр Валерьевич

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРАКТИКО-
ОРИЕНТИРОВАННОЙ РЕАЛИЗАЦИИ ПРОФЕССИОНАЛЬНЫХ
ДИСЦИПЛИН МАГИСТЕРСКИХ ПРОГРАММ В ОБЛАСТИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

Санкт-Петербург

2021



РЕФЕРАТ

Методические указания разработаны Университетом ИТМО в рамках Соглашения от 29.09.2021 №075-15-2021-1046 о предоставлении из федерального бюджета грантов в форме субсидий на разработку программ бакалавриата и программ магистратуры по профилю "искусственный интеллект", а также на повышение квалификации педагогических работников образовательных организаций высшего образования в сфере искусственного интеллекта.

Методические указания определяют подходы и механизмы практико-ориентированной реализации (выполнения практических заданий) профессиональных дисциплин в рамках программ магистратуры в области искусственного интеллекта (ИИ) в соответствии с траекторией компетенций – разработка систем ИИ.

Методические указания применимы к программам магистратуры, реализуемым в Университете ИТМО, а также в вузах - региональных партнерах: Тюменский индустриальный университет, Оренбургский государственный университет, Ивановский государственный энергетический университет, Тихоокеанский государственный университет, Астраханский государственный университет, Марийский государственный университет, а также в иных вузах Российской Федерации, ведущих образовательную деятельность на основе ФГОС 3++.



СОДЕРЖАНИЕ

1. Общие положения.....	4
2. Постановка задания на исследование / разработку.....	5
3. Порядок подготовки наборов данных.....	6
4. Жизненный цикл выполнения исследований / разработок.....	7
5. Представление результатов исследований / разработок.....	8
6. Приемка и оценивание результатов исследований / разработок.....	11

1. ОБЩИЕ ПОЛОЖЕНИЯ

1.1 Данные методические указания определяют подходы и механизмы практико-ориентированной реализации профессиональных дисциплин в рамках программ магистратуры в области ИИ в соответствии с траекторией компетенций – разработка систем ИИ.


1.2 Разработка систем ИИ является практико-ориентированным направлением информационных технологий; потому подготовка по данной траектории требует обязательного наличия в составе профессиональных дисциплин блока практических занятий, ориентированных на самостоятельную разработку (адаптацию) и программную реализацию алгоритмов машинного обучения и моделей ИИ с использованием современного математического и программного обеспечения.

1.3 Формы реализации практических занятий в области разработки систем ИИ:

- мастер-класс (демонстрация практических навыков преподавателем с последующим их усвоением обучаемыми путем повторения на компьютере, с элементами самостоятельной работы);
- аудиторная практическая работа (работа на компьютере в учебном классе по установленному заданию, одиночная или в малых группах, с консультацией преподавателя);
- самостоятельная работа (работа на компьютере в учебном классе или иных условиях по установленному заданию, одиночная или в малых группах, с предъявлением преподавателю окончательного цифрового результата для последующей приемки);
- научно-практический семинар (публичное представление обучаемым результатов аудиторной практической или самостоятельной работы с возможностью коллективного обсуждения и последующей приемки);
- смешанная форма реализации (может использовать все перечисленные выше формы).

1.4 В силу разнообразия и быстрого развития технологий прикладного ИИ и средств их разработки, для расширения содержательного охвата образовательных программ рекомендуется использовать «перевернутую» форму обучения (flipped classes), в рамках которой обучаемые публично представляют, обсуждают и валидируют результаты самостоятельного решения различных задач ИИ в соответствии с учебно-тематическим планом при модерации этого процесса преподавателем.

1.5 Технологии ИИ предназначены для воспроизведения когнитивной деятельности человека в условиях отсутствия регламентированного алгоритма действий, неопределенности и неполноты исходных данных. Потому ключевым элементом успешной



практико-ориентированной реализации, регламентируемым данными методическими указаниями, является обеспечение:

- постановки задания на исследование / разработку систем с элементами ИИ;
- подготовки массивов данных для исследования / разработки систем с элементами ИИ;
- формирования корректного жизненного цикла исследования / разработки систем с элементами ИИ;
- обеспечение воспроизводимости систем с элементами ИИ с целью их объективной оценки и анализа преподавателем;
- оценки качества результатов, полученных в ходе разработки систем с элементами ИИ.


2. ПОСТАНОВКА ЗАДАНИЯ НА ИССЛЕДОВАНИЕ / РАЗРАБОТКУ

2.1 Задание на исследование / разработку в рамках практических занятий может относиться к следующим категориям (определяемым спецификой дисциплины):

- фундаментальные исследования, направленные на улучшение существующих, а также разработку принципиально новых подходов ИИ;
- поисковые исследования, выполняемые для изучения возможных способов применения в прикладных исследованиях результатов фундаментальных исследований ИИ;
- прикладные исследования, выполняемые для достижения цели или решения конкретной прикладной задачи с использованием известных алгоритмов ИИ;
- практическая разработка, необходимая для отработки и закрепления умений и навыков, связанных с решением типовых прикладных задач с помощью существующих алгоритмов ИИ и средств их реализации.

2.2 Практическим результатом выполнения задания может быть:

- программная система (в форме исходных кодов), демонстрирующая работу реализованного метода ИИ или решение поставленной задачи на заданном наборе данных;
- программный сервис (доступный в форме SaaS), демонстрирующий работу реализованного метода ИИ или решение поставленной задачи на произвольных наборах данных в соответствии с заданными условиями эксплуатации;
- новые знания (соотношения на данных), основанные на воспроизводимых экспериментах с помощью алгоритмов ИИ (в форме исходных кодов, скриптов или последовательности операций на специализированных платформах);
- новые синтетические массивы данных, полученные с помощью алгоритмов ИИ.



2.3 Основанием для выполнения практических занятий является задание на исследование / разработку, формируемое преподавателем, исходя из специфики рабочей программы дисциплины. Оно должно содержать требования:

- по обеспечению и согласованию данных, необходимых для решения задачи, их объему и источникам;
- по классам алгоритмов ИИ, применимых для решения задачи;
- по выбранным метрикам оценки качества, включая (по усмотрению преподавателя): метрики классификации (точность, логарифмическая потеря, площадь под кривой ROC и др.); метрики для прогнозирования (матрица путаницы, классификационный отчет и др.), показатели регрессии (средняя абсолютная ошибка, средняя квадратичная ошибка, R^2 и др.), а также метрики для иных классов моделей ИИ;
- по способу представления результатов для проверки преподавателем.

2.4 В случае выполнения прикладных исследований, в задании может содержаться требование обосновать необходимость использования методов ИИ и невозможность эффективного решения поставленной прикладной задачи какими-либо иным (точным или эвристическим) методом.

3. ПОРЯДОК ПОДГОТОВКИ НАБОРОВ ДАННЫХ

3.1 Данные в исследованиях и разработках в области ИИ играют основополагающую роль. Для выполнения практических заданий преподавателем должны быть подготовлены следующие наборы данных:

- набор данных для обучения модели ИИ (ДО). Модель ИИ, в результате, будет обладать знаниями на базе этих данных, приобретенными в процессе обучения;
- проверочный набор данных для модели ИИ (ДП): набор данных, используемый для объективной эмпирической оценки соответствия модели набору ДО;
- тестовый набор данных (ДТ): набор данных, используемых для объективной оценки окончательной модели ИИ, обученной на ДО.

3.2 Наборы данных ДО и ДП должны быть предоставлены обучаемому в рамках задания на исследования / разработку. Набор данных ДТ обучаемому не раскрывается и используется для оценки успешности выполнения задания.

3.3 В ряде случаев, направленных на отработку практических навыков обучаемых, набор ДТ не требуется, поскольку степень решения задачи может быть с высокой вероятностью оценена по набору ДП.

3.4 Наборы ДО, ДП, ДТ должны быть либо размеченными данными, либо условно размеченными, когда известно как входные экземпляры данных соответствуют ожидаемым выходным результатам.

3.5 Преподаватель может формировать наборы данных таким образом, что они будут содержать элементы неоднородности, выбросы и исключения, ошибки и противоречия. Целесообразность этого определяется спецификой преподаваемой дисциплины. При этом о таких особенностях данных должно быть указано в задании на исследование / разработку.

3.6 Преподаватель может использовать для составления ДО, ДП и ДТ данные в различных форматах, в т.ч. из различных источников (по ссылкам в Интернет). В этом случае обучаемый должен провести процедуру сбора, унификации и форматирования данных самостоятельно, о чем должно быть указано в задании на исследование / разработку.

4. ЖИЗНЕННЫЙ ЦИКЛ ВЫПОЛНЕНИЯ ИССЛЕДОВАНИЙ / РАЗРАБОТОК

4.1 Жизненный цикл исследований / разработок систем с элементами ИИ в рамках реализации практических занятий состоит из следующих этапов:

- получение данных, предобработка и формирование единого представления, конструирование признаков модели, разделение данных на категории. Его выполняют с целью дальнейшей фокусировки на подходящих классах моделей и алгоритмов ИИ;
- выбор пути реализации исследования / разработки на основе первичного анализа данных; проводят с целью определения оптимального варианта направления исследований на основе анализа состояния исследуемой проблемы;
- итеративные теоретические и экспериментальные (посредством разработки программных реализаций) исследования выбранных классов моделей и алгоритмов ИИ; проводят с целью получения достаточных теоретических и достоверных экспериментальных (количественных) результатов решения задачи;
- оценка качества результатов исследований / разработок; проводят с целью доказательства работоспособности разработанного ПО и состоятельности предложенного решения задачи;
- обобщение и оценка результатов исследований / разработок, в том числе, подготовка технической документации по разработке, а также упаковка результатов в форму, пригодную для проверки преподавателем, в т.ч., с помощью автоматизированных средств;
- предъявление результатов к проверке преподавателем.

4.2 Для обеспечения своевременного выполнения исследований / разработок, осуществления оперативного контроля за выполнением работ и оценки промежуточных результатов обучаемый (при необходимости) разрабатывает и согласовывает с преподавателем план работ (план-график), содержащий последовательность и сроки выполнения отдельных операций, состав исполнителей (при групповой работе), сроки и форму предоставления промежуточных результатов.

4.3 В процессе выполнения работ обучаемый по согласованию с преподавателем может уточнять и корректировать план-график работ в пределах условий задания на выполнение исследований / разработок.


4.4 Для реализации процессов жизненного цикла исследований / разработок в учебных целях могут применяться цифровые инструменты, в т.ч. платформы DataMall и SMILE, разработанные в Университете ИТМО, или их аналоги.

5. ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ / РАЗРАБОТОК

5.1 Ключевой проблемой развития технологий ИИ является их воспроизводимость. Под воспроизводимостью понимается качественное свойство, определяющее достаточность информации об алгоритме обучения и модели ИИ, а также наборе обучающих данных и технологических условиях проведения оригинального эксперимента для повторения эксперимента в независимых условиях и получения результатов машинного обучения, близких к полученным в оригинальном эксперименте. Близость оценивается в соответствии с выбранной мерой. По данным 2020 г., более 50% научных статей в области ИИ не содержат данных, позволяющих адекватно воспроизвести указанные в них результаты. Потому обеспечение воспроизводимости является принципиальным для оценки результатов практической работы обучаемых, что требует выполнения ими ряда требований к описанию (упаковке) результатов.

5.1 Описание реализации модели ИИ и методов обучения должно соответствовать следующим требованиям:

- должно быть представлено текстовое описание того, какой метод обучения и модель ИИ были использованы, их конфигурация (например, значения гипер-параметров моделей);
- должна быть указана версия используемого языка программирования, а также названия внешних библиотек или математических пакетов, из которых был произведен импорт реализаций методов обучения или моделей ИИ;
- должны быть указаны конкретные версии библиотек, а также тип и версии операционной системы, на которой данную реализацию возможно воспроизвести. В случае, если возникает ситуация конфликта между версиями библиотек, и



автоматическая установка всех библиотек невозможна, следует дополнительно указать в описании, какие из версий не вызывают данного конфликта;

— сведения, перечисленные выше, должны быть зафиксированы в файле текстового формата. Формат файла должен быть наглядным для визуального восприятия (например, CSV, YAML, XML). В данный файл заносится следующая информация (при выполнении части требований сохраняются только соответствующие им данные):


- названия моделей и методов МО, которые были использованы;
- название и версия языка программирования;
- список использованных библиотек и математических пакетов, их версии;
- тип и версия операционной системы, на которых обеспечивается воспроизводимость результатов;
- список названий и версий библиотек и математических пакетов, которые использовались для импорта реализаций методов обучения;
- список названий и версий вспомогательных и системных библиотек;
- путь или ссылка на образ виртуальной машины или файловой системы, на которой развернуто необходимое ПО для воспроизведения результатов.


5.2 Описание аппаратного окружения, в котором выполняется обучение модели, должно соответствовать следующим требованиям:

- для фиксации аппаратного окружения, на котором возможно обеспечить воспроизводимость результатов, должны быть использованы публичные и открытые сервисы и средства виртуализации вычислений (предпочтительно, в собственном корпоративном облаке вуза или в публичном облаке);
- при невозможности использования сервисов виртуализации необходимо предоставить детальное описание аппаратного окружения, на котором обеспечивается воспроизводимость: (а) тип, частота и количество ядер CPU, (б) объем оперативной памяти и жесткого диска, тип и количество ядер GPU.

5.3 Описание входных и выходных данных должно соответствовать следующим требованиям:

- Общие требования к хранению данных должны быть выполнены в соответствии со стандартом [ISO/IEC TR 20547-2:2018 Information technology — Big data reference architecture — Part 2: Use cases and derived requirements].
- Для всех типов данных должны соблюдаться следующие условия:

- 
- сохранение данных должно выполняться как для набора ДО, использованного для построения модели ИИ, так и для набора ДП, используемого для оценки её качества;
 - принадлежность файлов к различным выборкам должна быть явно описана;
 - количество, порядок, названия, семантика, метаданные объектов должны соответствовать использованному при обучении модели ИИ представлению.
- Данные должны быть доступны в корпоративном облаке вуза (преподавателю предоставляются права доступа).
 - Для данных, хранящихся в текстовом табличном формате (CSV), дополнительно необходимо соблюдение следующих требований:
 - количество, порядок и название столбцов в таблицах должно соответствовать использованному при обучении модели представлению;
 - способ кодирования пропущенных значений должен быть явно обозначен в сопроводительной документации или начальных строках файла, предназначенных для хранения метаданных (обозначенных при этом специальным символом);
 - число записей в таблице должно соответствовать размеру выборки, использованной в рамках жизненного цикла создания модели ИИ;
 - точность сохранения значений переменных с плавающей запятой должна соответствовать таковой в среде, использованной для обучения и валидации модели.
 - Для данных, хранящихся в бинарном формате, дополнительно необходимо соблюдение следующих требований:
 - необходимо предоставить описание формата и версии открытого ПО, использованного для работы с файлом при обучении модели;
 - следует приложить к файлам описание семантики содержащихся в них данных в свободной форме.
 - Для данных, хранящихся в графических форматах, необходимо соблюдение следующих требований:
 - разрешение изображений должно совпадать с использованным при обучении и валидации модели;
 - сжатие изображений допустимо только с использованием алгоритмов сжатия без потерь, при условии указания конкретного ПО, использованного для сжатия.
 - Для данных, хранящихся в полнотекстовых форматах, необходимо соблюдение следующих требований:
 - указана кодировка текста;

- 
- в случае невозможности предоставления исходного текста, необходимо предоставлять его векторизованное описание с приложением текста программы, использованной для векторизации.

5.4 Требования к представлению моделей и данных могут быть использованы при создании автоматизированных средств оценки выполнения практических заданий.

6. ПРИЕМКА И ОЦЕНИВАНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ / РАЗРАБОТОК

6.1 К приемке предъявляют: результат разработки – цифровой продукт (код и данные), оформленные в соответствии с п. 5, комплект технической документации (если определено заданием), иные материалы (если определено заданием).

6.2 Приемка заключается в рассмотрении (преподавателем или иными обучаемыми в рамках модели flipped classes) результатов выполненных работ с целью оценки освоения обучаемым содержания практических занятий, в том числе:

- определения соответствия установленным в задании интервалам метрик качества моделей ИИ;
- оценки степени проработки альтернативных решений и корректности использования алгоритма обучения для поиска параметров и гиперпараметров модели ИИ;
- оценки устойчивости модели ИИ на синтетических наборах данных, не использованных в наборах ДО и ДП;
- оценки эффективности модели ИИ в сравнении с базовым решением (baseline), полученным независимо (например, используя функциональности платформы SMILE Университета ИТМО, или аналогов).

6.3 Оценивание результатов работ выходит за рамки данной методики в силу необходимости определения весов критериев оценки в соответствии с компетентностной моделью конкретной образовательной программы.