Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

УТВЕРЖДАЮ
декан факультета вычислительной
математики и кибернетики
_____/И.А. Соколов /
«14» октября 2021г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Атаки на системы искусственного интеллекта»

Уровень высшего образования:

магистратура

Направление подготовки / специальность:

01.04.02 "Прикладная математика и информатика" (3++)

Направленность (профиль) ОПОП:

Искусственный интеллект в кибербезопасности

Форма обучения:

очная

Рабочая программа рассмотрена и утверждена
на заседании Ученого совета факультета ВМК
(протокол №7, от 29 сентября 2021 года)

Москва 2021

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 01.04.02 "Прикладная математика и информатика" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. Место дисциплины (модуля) в структуре ОПОП ВО:
Дисциплина (модуль) относится к обязательной части основной профессиональной образовательной программы.

2. Входные требования для освоения дисциплины (модуля), предварительные условия:
Изучение дисциплины базируется на освоении дисциплины «Робастные модели в машинном обучении»

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	**Индикатор (показатель) достижения компетенции**	**Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций**
ОПК-4. Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	ОПК-4.1. Разрабатывает программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности	ОПК-4.1. З-1. Знает новые научные принципы и методы разработки программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач в различных предметных областях. ОПК-4.1. У-1. Умеет разрабатывать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности для решения профессиональных задач в различных предметных областях.
ПК-2. Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования	ПК-2.1. Выбирает и разрабатывает программные компоненты систем искусственного интеллекта ПК-2.2. Проводит экспериментальную проверку работоспособности систем искусственного интеллекта	ПК-2.1. З-1. Знает основные критерии эффективности и качества функционирования системы искусственного интеллекта: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем искусственного интеллекта ПК-2.1. З-2. Знает методы, языки и программные средства разработки программных компонентов систем искусственного интел-

		лекта ПК-2.1. У-1. Умеет выбирать, адаптировать, разрабатывать и интегрировать программные компоненты систем искусственного интеллекта с учетом основных критериев эффективности и качества функционирования ПК-2.2. З-1. Знает методы постановки задач, проведения и анализа тестовых и экспериментальных испытаний работоспособности систем искусственного интеллекта ПК-2.2. У-1. Умеет ставить задачи и проводить тестовые и экспериментальные испытания работоспособности систем искусственного интеллекта анализировать результаты и вносить изменения

4. Объем дисциплины (модуля) составляет 5 з.е., в том числе 72 академических часа, отведенных на контактную работу обучающихся с преподавателем, 108 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

Целью курса является обучение слушателей планированию и проведению атак на модели машинного обучения. Для достижения чего необходимо решить следующие задачи:

1. подробно рассмотреть существующие методы защиты моделей машинного обучения от внешних воздействий;
2. подробно рассмотреть существующие методы и средства для проведения атак;
3. представить возможность слушателям атаковать существующие модели машинного обучения.

Наименование и крат-	**Всего**	В том числе

кое содержание разделов и тем дисциплины (модуля), **форма промежуточной аттестации по дисциплине (модулю)**	**(часы)**	**Контактная работа (работа во взаимодействии с преподавателем), часы** из них						**Самостоятельная работа обучающегося, часы** из них		
		Занятия лекционного типа	Практические занятия	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости (коллоквиумы, практические контрольные занятия и др)	**Всего**	Выполнение домашних заданий	Подготовка рефератов и т.п.	**Всего**
Тема 1. Введение в тему атак на модели машинного обучения	2	2	-	-	-	-	2	-	-	-
Тема 2. Схемы атак	8	6	6	-	-	-	12	24	-	24
Тема 3. Атаки отравлением	10	6	6	-	-	-	12	24	-	24
Тема 4. Атаки уклонением	20	6	6				12	24	-	24
Тема 5. Атаки извлечением	16	6	6	-	-	-	12	24	-	24
Тема 6. Атаки с применением порождающих алгоритмов	8	6	6	-	-	-	12	24	-	24

Тема 7. Автоматизация процесса атак	10	4	6				10	24	-	24
8. Промежуточная аттестация экзамен										
Итого		36	36				72			108

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания

Примерные практические задания

- **Задание 1.** Разработка порождающей модели МО для генерации изображения лица целевой персоны, позволяющий нарушить работу биометрического классификатора пользователей по лицу. Биометрический классификатор будет предоставлен.
- **Задание 2.** Разработка порождающей модели МО для генерации голоса целевого диктора, позволяющий нарушить работу биометрического классификатора дикторов по голосу. Биометрический классификатор будет предоставлен.
- **Задание 3**. Реализация пула состязательных атак, позволяющих нарушить работу биометрического классификатора пользователей по лицу. Биометрический классификатор будет предоставлен.
- **Задание 4.** Реализация бинарного классификатора синтетических данных, позволяющий идентифицировать такого сорта данные и тем самым защитить модель биометрической классификации лиц. Защита должна эффективно работать от атак, разработанных командой в рамках задания 1. Биометрический классификатор будет предоставлен.
- **Задание 5.** Реализация механизма состязательного обучения, позволяющего защитить модель биометрической классификации лиц. Защита должна эффективно работать от атак, разработанных командой в рамках задания 3. Биометрический классификатор будет предоставлен.

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине, критерии и шкалы оценивания

Список вопросов для экзамена.

1. Подходы к созданию состязательных примеров.
2. Атаки отравление.
3. Атаки уклонением.
4. Атаки извлечением.
5. Атаки с применением порождающих моделей.
6. Подходы к автоматизации процесса атак.

Методические материалы для проведения процедур оценивания результатов обучения

Особенности организации процесса обучения

Для эффективного освоения курса рекомендуется перед каждым занятие привести в порядок конспекты лекций. После каждого занятия рекомендуется найти и прочитать дополнительную литературу по теме лекции и прочитать свои конспекты.

Система контроля и оценивания

За каждую домашнюю выставляются баллы (максимум 40 баллов). Пусть M – максимальное число баллов, которое может набрать студент. В конце семестра баллы конвертируются в оценку O1 следующим образом:

меньше M/2 баллов: O1=2;

больше или равно M/2 баллов, но меньше 2M/3: O1=3;

больше или равно 2M/3 баллов, но меньше 5M/6: O1=4;

больше или равно 5M/6 баллов: O1=5.

На экзамене оценка O1 является стартовой. Окончательная оценка определяется исходя из оценки устного ответа студента, при этом она не может отличаться от стартовой оценки более чем на 1 балл.

Структура и график контрольных мероприятий

Устная сдача домашних заданий в конце каждой недели, устный экзамен в конце семестра.

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)				
Оценка	2	3	4	5

РО и соответствующие виды оценочных средств				
Знания *Экзамен*	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения *Практические задания*	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) *Экзамен, практические занятия*	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Основная литература:

1. Шакла,Нишант Машинное обучение & TensorFlow : [пер. с англ.]. / Нишант Шакла при участии Кена Фрикласа. - СПб. [и др.] : Питер, 2019. - 331, [1] с.; 24 см - (Библиотека программиста).
2. Шолле,Франсуа Глубокое обучение на Python / Франсуа Шолле ; [пер. с англ. А. Киселева]. - СПб. [и др.] : Питер, 2020. - 397, [1] с.; 24 см - (Библиотека программиста).

Дополнительная литература:

1. Bishop C. M. Pattern recognition and machine learning. – Springer, 2006
2. .Коэльо Л. П., Ричарт В. Построение систем машинного обучения на языке Python. – М: ДМК Пресс. – 2016. (Coelho L. P., Richert W. Building machine learning systems with Python. — 2nd ed. — Packt Publishing Ltd, 2015.)
3. Max Kuhn, Kjell Johnson. Applied Predictive Modeling. — Springer, 2013.
4. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p. — ISBN 978-0-387-84857-0.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства

При реализации дисциплины может быть использовано следующее программное обеспечение:

1. Программное обеспечение для подготовки слайдов лекций MS PowerPoint.
2. Программное обеспечение для создания и просмотра pdf-документов Adobe Reader.
3. Издательская система LaTeX.

7.3. Перечень профессиональных баз данных и информационных справочных систем

1. http://www.edu.ru – портал Министерства образования и науки РФ
2. http://www.ict.edu.ru – система федеральных образовательных порталов «ИКТ в образовании»
3. http://www.openet.ru - Российский портал открытого образования
4. http://www.mon.gov.ru - Министерство образования и науки Российской Федерации
5. http://www.fasi.gov.ru - Федеральное агентство по науке и инновациям

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Math-Net.Ru [Электронный ресурс] : общероссийский математический портал / Математический институт им. В. А. Стеклова РАН ; Российская академия наук, Отделение математических наук. - М. : [б. и.], 2010. - Загл. с титул. экрана. - Б. ц.
URL: http://www.mathnet.ru
2. Университетская библиотека Online [Электронный ресурс] : электронная библиотечная система / ООО "Директ-Медиа" . - М. : [б. и.], 2001. - Загл. с титул. экрана. - Б. ц. URL: www.biblioclub.ru
3. Универсальные базы данных East View [Электронный ресурс] : информационный ресурс / East View Information Services. - М. : [б. и.], 2012. - Загл. с титул. экрана. - Б. ц.
URL: www.ebiblioteka.ru
4. Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] : информационный портал / ООО "РУНЭБ" ; Санкт-Петербургский государственный университет. - М. : [б. и.], 2005. - Загл. с титул. экрана. - Б. ц.
URL: www.eLibrary.ru

7.5. Описание материально-технического обеспечения.

Факультет, ответственный за реализацию данной Программы, располагает соответствующей материально-технической базой, включая современную вычислительную технику, объединенную в локальную вычислительную сеть, имеющую выход в Интернет. Используются специализированные компьютерные классы, оснащенные современным оборудованием. Материальная база факультета соответствует действую-

щим санитарно-техническим нормам и обеспечивает проведение всех видов занятий (лабораторной, практической, дисциплинарной и междисциплинарной подготовки) и научно-исследовательской работы обучающихся, предусмотренных учебным планом.

8. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

9. Разработчик (разработчики) программы.

Малоян Нарек Гагикович, Саада Даниель Фирасович, Ильюшин Евгений Альбинович, Намиот Дмитрий Евгеньевич.

10. Язык преподавания - русский.