Приложение к ОПОП «Безопасность и этика искусственного интеллекта»

СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования
«Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)

РАБОЧАЯ ПРОГРАММА

дисциплины

«АНАЛИТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ В ЗАДАЧАХ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ»

для подготовки магистров

по направлению

09.04.01 «Информатика и вычислительная техника»

по программе

«Безопасность и этика искусственного интеллекта»

Санкт-Петербург

2023

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:
преподаватель Смирнов Г.Е.

Рабочая программа рассмотрена и одобрена на заседании кафедры ВТ
02.09.2021, протокол № 6

Рабочая программа рассмотрена и одобрена учебно-методической комиссией ФКТИ, 16.09.2021, протокол № 6

Согласовано в ИС ИОТ
Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ИБ
Общая трудоемкость (ЗЕТ)	4
Курс	1
Семестр	2
Виды занятий	
Лекции (академ. часов)	17
Практические занятия (академ. часов)	17
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	35
Самостоятельная работа, включая часы на контроль (академ. часов)	109
Всего (академ. часов)	144
Вид промежуточной аттестации	
Дифф. зачет (курс)	1

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«АНАЛИТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ В ЗАДАЧАХ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ»

Дисциплина посвящена изучению безопасности личности в цифровом пространстве, разведке в информационном пространстве, информационных войнах и применению больших объемов данных в задачах информационной безопасности. В рамках данной дисциплины рассматриваются основные подходы к сбору больших объемов информации из открытых источников, их накоплению, обработке и анализу. Дисциплина формирует знания студентов о современных технологиях анализа цифрового следа личности, дает понимание природы утечек информации и применения больших данных в SIEM (также необходимости самих SIEM). Также дисциплина формирует умения и навыки работы с самыми современными технологиями контейнеризации и их аспектами безопасности, нереляционными и графовыми базами данных, стеком Apache Hadoop и Apache NiFi (для хранения и обработки больших объемов данных), технологиями машинного обучения и графовыми нейронными сетями, а также применения данных технологий в информационной безопасности.

SUBJECT SUMMARY

«ANALYTICAL DATA PROCESSING IN INFORMATION SECURITY TASKS»

The discipline is devoted to the study of personal security in the digital space, intelligence in the information space, information wars and the use of large amounts of data in information security tasks. Within the framework of this discipline, the main approaches to the collection of large amounts of information from open sources, their accumulation, processing and analysis are considered. The discipline forms students' knowledge about modern technologies for analyzing the digital fingerprint of a person, gives an understanding of the nature of information leaks and the use

of big data in SIEM (also the need for SIEM itself). The discipline also forms the skills and abilities of working with the most modern containerization technologies and their security aspects, non-relational and graph databases, the Apache Hadoop and Apache NiFi stack (for storing and processing large amounts of data), machine learning technologies and graph neural networks and the use of these technologies in information security.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Предметом изучения являются основные подходы к обработке, накоплению и анализу данных.

2. Задачей дисциплины является рассмотрение основных подходов к сбору больших объемов информации из открытых источников, их накоплению, обработке и анализу в интересах информационной безопасности.

3. Дисциплина формирует знания студентов о современных технологиях анализа цифрового следа личности, дает понимание природы утечек информации и применения больших данных.

4. Дисциплина формирует умения и навыки работы с самыми современными технологиями контейнеризации и их аспектами безопасности, нереляционными и графовыми базами данных, технологиями машинного обучения и графовыми нейронными сетями, а также применения данных технологий в информационной безопасности

5. Результатом освоения дисциплины является приобретение практических навыков в анализе существующих методов и средств, применяемых для контроля и защиты информации, разработке математических моделей, реализуемых в средствах защиты информации и выполнении анализа защищенности сетевых сервисов с использованием средств автоматического реагирования на попытки несанкционированного доступа к ресурсам компьютерных систем и сетей, разрабатывать программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях, модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом тре-

бований информационной безопасности в различных предметных областях.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Архитектура параллельных вычислительных систем»

2. «Введение в нейронные сети»

3. «Машинное обучение»

и обеспечивает изучение последующих дисциплин:

1. «Производственная практика (технологическая (проектно-технологическая) практика)»

2. «Анализ данных в искусственном интеллекте»

3. «Аппаратно-программные средства защиты информации в компьютерных системах»

4. «Основы построения защищенных компьютерных сетей»

5. «Теория информации и теория кодирования»

6. «Производственная практика (преддипломная практика)»

3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

Код компетенции/ индикатора компетенции	**Наименование компетенции/индикатора компетенции**
ПК-30	Способен разрабатывать и модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности в различных предметных областях
ПК-30.1	*Разрабатывает программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях*
ПК-30.2	*Модернизирует программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях*

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Содержание разделов дисциплины

4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	ИКР, ач	СР, ач
1	Введение	1			2
2	Влияние социальных сетей, медиа и всеобщего проникновения Интернет на жизнь современного человека. Проблемы безопасности личности в цифровом пространстве. Цифровой след личности в медиапространстве.	2			15
3	Проблемы классических подходов к обработке, накоплению и анализу данных, разработка новых подходов. Изменчивость информационных систем.	2			15
4	Хранение больших объемов данных. Стек технологий Apache Hadoop. Файловая система HDFS. Потоковая обработка данных с помощью Apache Ni-Fi. Архитектурные решения хранения больших объемов данных, примененные в Apache Hadoop.	2	6		15
5	Вычисления в памяти как единственный способ обработки больших данных в реальном времени. Современные технологии вычислений в памяти.	2	5		15
6	Модель анализа текстов BERT. Модель анализа текстов CatBoost	2			16
7	Графовые нейронные сети.	2	6		16
8	Применение изученных подходов для хранения и анализа событий информационной безопасности.	3			15
9	Заключение	1		1	0
	Итого, ач	17	17	1	109
	Из них ач на контроль	0	0	0	0
	Общая трудоемкость освоения, ач/зе	144/4			

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Структура рабочей программы курса. Цели информационной безопасности. Место анализа данных и машинного обучения в общей комплексной системе защиты информации. Цели анализа данных в задачах информационной безопасности.

№ п/п	Наименование темы дисциплины	Содержание
2	Влияние социальных сетей, медиа и всеобщего проникновения Интернет на жизнь современного человека. Проблемы безопасности личности в цифровом пространстве. Цифровой след личности в медиапространстве.	Влияние социальных сетей, медиа и всеобщего проникновения ГИС Интернет на жизнь современного человека. Является ли виртуальная свобода несвободой, зависимостью? Ощущают ли студенты, школьники и взрослые люди манипулятивное влияние социальных сетей. Проблемы безопасности личности в цифровом пространстве. Дается понятие ”Цифрового следа” личности человека. Приводятся примеры угроз безопасности личности посредством анализа ее цифрового следа.
3	Проблемы классических подходов к обработке, накоплению и анализу данных, разработка новых подходов. Изменчивость информационных систем.	Освещение основных классических подходов к обработке, накоплению и анализу данных. Проблемы классических подходов. Пути решения проблем классических подходов на примере новых подходов. Эффективность новых подходов в условиях изменчивости информационных систем.
4	Хранение больших объемов данных. Стек технологий Apache Hadoop. Файловая система HDFS. Потоковая обработка данных с помощью Apache Ni-Fi. Архитектурные решения хранения больших объемов данных, примененные в Apache Hadoop.	Дается понятие больших данных. Проблемы хранения больших объемов данных. Рассматривается стек технологий Apache Hadoop как решение с открытым исходным кодом. Уделяется внимание файловой системе HDFS, потоковой обработке больших данных с помощью системы Apache NiFi с последующим сохранением в HDFS. Приводятся примеры архитектурных решений для хранения больших данных, примененные в Apache Hadoop.
5	Вычисления в памяти как единственный способ обработки больших данных в реальном времени. Современные технологии вычислений в памяти.	Рассматриваются проблемы обработки больших данных в реальном времени. Приводятся примеры решений и современные технологии вычислений в памяти.
6	Модель анализа текстов BERT. Модель анализа текстов CatBoost	Раскрывается необходимость применения машинного обучения в анализе данных, подходы к анализу текстов в нейронных сетях.
7	Графовые нейронные сети.	Приводятся примеры графовых нейронных сетей. Раскрывается необходимость использования машинного обучения в задачах информационной безопасности.
8	Применение изученных подходов для хранения и анализа событий информационной безопасности.	Производится соединение воедино изученных подходов хранения, обработки и анализа данных. Раскрывается применение данных подходов в задачах информационной безопасности.
9	Заключение	Подведение итогов. Уяснение изученного материала.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Особенности хранения, обработки и анализа больших объемов данных	6
2. Вычисления с помощью Apache Spark	5
3. Особенности применения графовых нейронных сетей	6
Итого	17

4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

4.5 Реферат

Реферат не предусмотрен.

4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.

4.7 Доклад

Доклад не предусмотрен.

4.8 Кейс

Кейс не предусмотрен.

4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь

период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

Текущая СРС	**Примерная трудоемкость, ач**
Работа с лекционным материалом, с учебной литературой	30
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	3
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	44
Подготовка к контрольным работам, коллоквиумам	14
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	5
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	3
Подготовка к зачету, дифференцированному зачету, экзамену	10
ИТОГО СРС	109

5 Учебно-методическое обеспечение дисциплины

5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Элбон, Крис. Машинное обучение с использованием Python. Сборник рецептов [Текст] : [пер. с англ.] / К. Элбон, 2020. -369 с.	20
2	Цехановский, Владислав Владимирович . Интеллектуальный анализ данных [Текст] : учеб. пособие / В. В. Цехановский, В. Д. Чертовской, 2019. -55 с.	23
3	Мхитарян, Владимир Сергеевич. Анализ данных [Электронный ресурс] : Учебник для вузов / под ред. Мхитаряна В.С., 2020. -490 с	неогр.
4	Карау Х. Изучаем Spark: молниеносный анализ данных [Электронный ресурс], 2015. -304 с.	неогр.
5	Маккинни У. Python и анализ данных [Электронный ресурс] : научное издание / У. Маккинни, 2020. -540 с.	неогр.
6	Чио К. Машинное обучение и безопасность [Электронный ресурс] : руководство / К. Чио, Д. Фримэн, 2020. -388 с.	неогр.
Дополнительная литература		
1	Цехановский, Владислав Владимирович . Интеллектуальный анализ данных [Текст] : учеб. пособие / В. В. Цехановский, В. Д. Чертовской, 2019. -55 с.	23
2	Бенджамин Бенгфорт Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка [Электронный ресурс] / Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони, 2021. -368 с.	неогр.

5.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	Машинное обучение.https://stepik.org/course/8057/promo?search=756628146
2	Hadoop. Система для обработки больших объемов данных.https://stepik.org/course/150/syllabus

5.3 Адрес сайта курса

Адрес сайта курса: https://vec.etu.ru/moodle/course/view.php?id=7574

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Аналитическая обработка данных в задачах информационной безопасности» предусмотрены следующие формы промежуточной аттестации: зачет с оценкой.

Зачет с оценкой

Оценка	Описание
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок теорем
Хорошо	Студент овладел курсом, но в отдельных вопросах испытывает затруднения. Умеет решать задачи
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач.

Особенности допуска

Основным критерием допуска к зачету с оценкой является успешная защита всех отчетов по практическим занятиям, а также достижение нормы значением в 80% по посещаемости лекционных и практических занятий. При явке на дифф. зачет и при получении зачета студент обязан иметь при себе зачетную книжку, которую он предъявляет преподавателю при получении зачета или в начале экзамена. Приём зачёта с оценкой без зачётной книжки не разрешается. Зачет с оценкой проводится в письменной форме, по билетам, составленным в соответствии с программой курса и утвержденным заведующим кафедрой и деканом факультета. При проведении зачетов с оценкой могут быть использованы технические средства. Преподавателю предоставляется право задавать студентам вопросы сверх билета, в соответствии с учебной программой.

6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

Вопросы к дифф.зачету

№ п/п	Описание
1	Влияние соцмедиа и всеобщего проникновения Интернет в жизнь современного человека.
2	Проблемы безопасности личности в цифровом пространстве.
3	Цифровой след пользователя соцмедиа.
4	Российское и международное законодательство в области персональных данных . ФЗ-152. GDPR.
5	Контроль утечек информации в открытых источниках.
6	Конкурентная разведка. Понятие, методы и средства. Защита от конкурентной разведки.
7	Информационные войны. Понятие, методы и средства. Влияние на общество.
8	Проблемы классических подходов обработки, накопления и анализа данных.
9	Изменчивость информационных систем. Обработка информации в условиях изменчивости информационных систем.
10	Технология контейнеризации. Преимущества. Особенности.
11	Аспекты безопасности технологии контейнеризации.
12	Нереляционные базы данных. Достоинства и недостатки.
13	Архитектурные особенности нереляционных баз данных.
14	Применение нереляционных баз данных в информационной безопасности

15	Стек Apache Hadoop. Обзор возможностей.
16	Файловая система HDFS. Достоинства и недостатки. Применение.
17	Парадигма MapReduce. Достоинства и недостатки. Применение.
18	Система потоковой обработки информации Apache Ni-Fi. Достоинства и недостатки. Применение.
19	Архитектурные решения хранения больших объемов данных примененные в Hadoop.
20	СУБД Apache HBase. Особенности. Достоинства и недостатки. Применение.
21	Платформа Apache Pig. Особенности. Достоинства и недостатки. Применение.
22	СУБД Apache Hive. Особенности. Достоинства и недостатки. Применение.
23	Виды параллельных вычислений. Вычисления в памяти. Методы и средства.
24	Фреймворк Apache Spark. Особенности. Достоинства и недостатки. Применение.
25	Визуализация больших данных. Особенности. Достоинства и недостатки. Применение.
26	Графовые базы данных. Применение в информационной безопасности.
27	СУБД JanusGraph. Особенности. Достоинства и недостатки. Применение в современных решениях.
28	Аспекты безопасности в Apache Hadoop. Контроль доступа.
29	Аспекты безопасности в Apache Hadoop. Контроль целостности.
30	Особенности открытого программного обеспечения с точки зрения ИБ.
31	Статистический анализ кода. Методы и средства.
32	Уязвимости открытого ПО.
33	Бэкдоры в открытом ПО.
34	Особенности сборки отрытого ПО.
35	Машинное обучение. Развитие. Применение в ИБ.
36	Модель анализа текстов BERT.
37	Модель анализа текстов CatBoost.
38	Графовые нейронные сети GNN.
39	SIEM. Определение. Архитектурные решения. Достоинства и недостатки.
40	Применение изученных подходов для хранения и анализа событий информационной безопасности в SIEM.

Форма билета

Министерство науки и высшего образования Российской Федерации

ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» имени В.И. Ульянова (Ленина)»

_____

БИЛЕТ № 1

Дисциплина Аналитическая обработка данных

в задачах информационной безопасности ФКТИ

1. Влияние соцмедиа и всеобщего проникновения Интернет в жизнь современного человека.

2. Применение изученных подходов для хранения и анализа событий информационной безопасности в SIEM.

УТВЕРЖДАЮ

Заведующий кафедрой

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
6	Хранение больших объемов данных. Стек технологий Apache Hadoop. Файловая система HDFS. Потоковая обработка данных с помощью Apache Ni-Fi. Архитектурные решения хранения больших объемов данных, примененные в Apache Hadoop.	
7		
8		Коллоквиум
9	Вычисления в памяти как единственный способ обработки больших данных в реальном времени. Современные технологии вычислений в памяти.	
10		Коллоквиум
14	Графовые нейронные сети.	
15		Коллоквиум

6.4 Методика текущего контроля

на лекционных занятиях

Текущий контроль включает в себя контроль посещаемости (не менее ***80*** % занятий), по результатам которого студент получает допуск на экзамен.

на практических (семинарских) занятиях

Текущий контроль включает в себя контроль посещаемости (не менее ***80*** % занятий), а также выполнение, сдачу в срок отчетов и их защиту по всем практическим работам на коллоквиумах, по результатам которой студент получает допуск на экзамен.

В ходе проведения семинарских и практических занятий целесообразно привлечение студентов к как можно более активному участию в дискуссиях, решении задач, обсуждениях и т. д. При этом активность студентов также может учитываться преподавателем, как один из способов текущего контроля на практических занятиях.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на лекционных и практических занятиях студентов по методикам, описанным выше.

7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, компьютер или ноутбук, проектор, экран, маркерная доска.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше, СДО ”Moodle”
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя. Количество ЭВМ -в соответствии с контингентом.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Самостоятельная работа	Помещение для самостоятельной работа	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Ubuntu 18.04 и выше

8 Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Дата	Изменение	Дата и номер протокола заседания УМК	Автор	Начальник ОМОЛА
1	14.02.2023	Программа актуальна, изменения не требуются.	14.02.2023 г., протокол заседания УМК № 2	преподаватель, Г.Е. Смирнов	