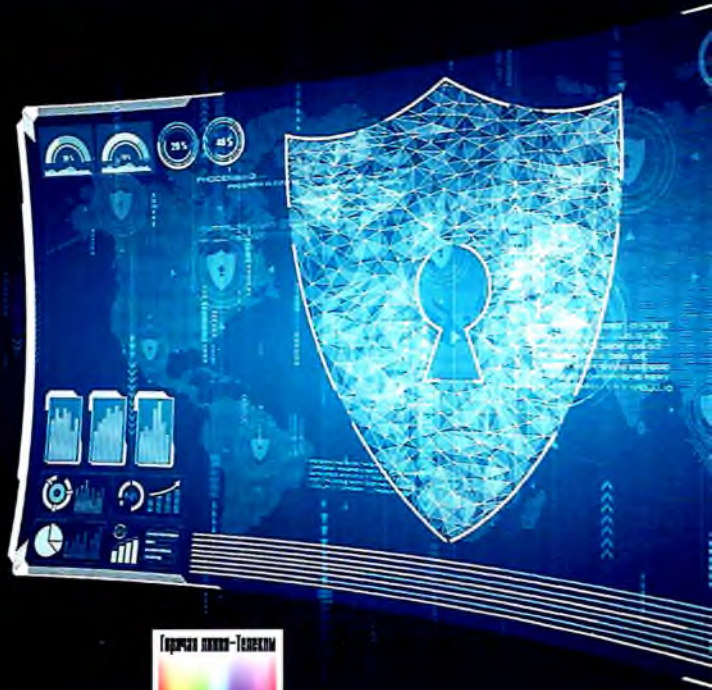


В. А. Частикова, В. Г. Гуляй, С. А. Жерлицын

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КИБЕРБЕЗОПАСНОСТИ



Горюхины-Тележкины



Оглавление

Введение	3
Раздел I. МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	6
Глава 1. Классические нейронные сети	9
1.1. Персептрон	11
1.2. Алгоритм обратного распространения ошибки	14
1.3. Сеть Хэмминга	16
Контрольные вопросы	18
Практические задания	19
Глава 2. Глубокое обучение	20
2.1. Нейросетевые эмбединги	21
2.2. Сверточные нейронные сети	22
2.2.1. Базовые сверточные нейронные сети	22
2.2.2. Быстрые сверточные нейронные сети	27
2.2.3. Региональные сверточные нейронные сети	29
2.2.4. Развертываемые нейронные сети	32
2.2.5. Применение CNN в задачах компьютерного зрения	32
2.3. Рекуррентные нейронные сети	34
2.3.1. Базовые рекуррентные нейронные сети	34
2.3.2. Рекуррентные нейронные сети с долгой кратковременной памятью	36
2.3.3. Управляемые рекуррентные нейроны	38
2.3.4. Двухнаправленные рекуррентные нейронные сети	39
2.3.4. Механизм самовнимания	41
2.3.5. Механизм многоголового внимания	44
2.3.6. Архитектура «Трансформер»	45
Контрольные вопросы	47

Практические задания	48
Глава 3. Обучение нейронных сетей	51
3.1. Функции потерь и перекрестная энтропия	51
3.2. Функция потерь	52
3.3. Градиентный спуск	53
3.3.1. Стохастический градиентный спуск	54
3.3.2. Стохастический градиентный спуск с импульсом Нестерова	55
3.3.3. Алгоритм Adagrad	56
3.3.4. Алгоритм Adadelta	57
3.3.5. Адаптивный алгоритм оптимизации RMSProp ..	57
3.3.6. Адаптивный алгоритм оптимизации Adam	58
3.3.7. Алгоритм AdaMax	60
3.4. Способы борьбы с проблемой переобучения	60
3.5. Метаэвристические алгоритмы оптимизации	62
3.5.1. Генетический алгоритм	62
3.5.2. Метод дуэлей	66
3.5.3. Модифицированный генетический алгоритм ду- элей	68
Контрольные вопросы	69
Практические задания	70
Глава 4. Алгоритмы анализа естественного языка ..	74
4.1. Символические языковые модели	77
4.2. Базовые методы представления лингвистических дан- ных в векторном виде	79
4.3. Частотные языковые модели	81
4.4. Вероятностные языковые модели	83
4.5. Матричные языковые модели	84
4.6. Векторные языковые модели в многомерном прост- ранстве	89
4.7. Контекстуальные представления на основе RNN/ LSTM	98
4.8. Языковые модели на базе архитектуры «Трансфор- мер»	101
4.8.1. Encoder-only-модели и многозадачные архитек- туры	104
4.8.2. Decoder-only-модели и большие языковые мо- дели	108

4.8.3. Гибридные Encoder-Decoder модели	115
Контрольные вопросы	121
Практические задания	121
Глава 5. Генеративный искусственный интеллект ...	125
5.1. Основные понятия, архитектуры и задачи генеративного ИИ	125
5.2. Вариационные автоэнкодеры	127
5.2.1. Байесовский подход и архитектура модели	127
5.2.2. Обучение VAE и максимизация нижней оценки	129
5.2.3. Проблема повторной параметризации и метод ее решения	129
5.2.4. Анализ функционирования и ограничения VAE	131
5.3. Генеративно-сопоставительные сети	131
5.3.1. Принцип минимаксной игры генератора и дискриминатора	132
5.3.2. Проблемы обучения: нестабильность и коллапс мод	133
5.3.3. Усовершенствованные архитектуры GAN	134
5.4. Авторегрессионные модели и большие языковые модели	139
5.4.1. Принцип авторегрессии	139
5.4.2. Эволюция генеративных языковых моделей ...	140
5.4.3. Методы управления генерацией	142
5.5. Диффузионные модели	143
5.5.1. Принцип диффузионного процесса: прямой и обратный ход	144
5.5.2. Обучение модели предсказания шума	144
5.5.3. Практические реализации: Stable Diffusion и управляемая генерация	145
5.6. Гибридные архитектуры	146
5.6.1. Нормализующие потоки	146
5.6.2. Energy-Based Models	147
5.6.3. Мультимодальные модели	147
Контрольные вопросы	149
Практические задания	150
Глава 6. Алгоритмы речевого интеллекта	155
6.1. Алгоритм серых волков	155

6.2. Алгоритм светлячков	156
6.3. Метод опыления цветов	157
6.4. Искусственная иммунная система	159
6.5. Гибридная искусственная иммунная система	163
Контрольные вопросы	164
Практические задания	165
Глава 7. Экспертная система	169
7.1. Методы представления знаний	170
7.2. Механизмы вывода	171
Контрольные вопросы	173
Практические задания	173
Раздел II. БИОМЕТРИЧЕСКАЯ АУТЕНТИФИКАЦИЯ	176
Глава 8. Аутентификация по статическим физиометрическим признакам	180
8.1. Распознавание по отпечатку пальца	180
8.1.1. Классификация макроскопических характеристик	180
8.1.2. Микроскопические особенности отпечатка пальца	182
8.1.3. Математический аппарат классического анализа папиллярного узора	182
8.1.4. Нейросетевые подходы	193
8.1.5. Обучающая выборка	197
8.1.6. Аппаратная реализация: типы сканеров	198
8.2. Аутентификация по рисунку вен ладони	201
8.2.1. Физиологические основы метода	201
8.2.2. Математическая постановка задачи	202
8.2.3. Алгоритмическая обработка данных	202
8.2.4. Классические методики аутентификации по рисунку вен ладони	204
8.2.5. Нейросетевые методы	205
8.2.6. Обучающая выборка	206
8.2.7. Аппаратная реализация	207
8.3. Аутентификация по геометрии лица в статике	209
8.3.1. Ключевые признаки и основные методы	209

8.3.2. Математическое обоснование классического алгоритм 2D-распознавания	210
8.3.3. Математическое обоснование классического алгоритма 3D-распознавания лица	214
8.3.4. Нейросетевые методы распознавания лиц	218
8.3.5. Нетривиальные задачи	220
8.3.6. Обучающие выборки	222
8.3.7. Аппаратная реализация	224
8.4. Распознавание по радужной оболочке глаза	226
8.4.1. Физиологические основы	226
8.4.2. Математическая постановка задачи и алгоритмическая обработка данных	227
8.4.3. Классические методы аутентификации по радужной оболочке глаза	231
8.4.4. Нейросетевые методы аутентификации по радужной оболочке глаза	232
8.4.5. Обучающие выборки	234
8.4.6. Аппаратная реализация	235
Контрольные вопросы	237
Практические задания	237
Глава 9. Аутентификация по динамическим и поведенческим признакам	241
9.1. Анализ лиц в видеопотоке	242
9.1.1. Алгоритм распознавания лиц в видеопотоке ...	242
9.1.2. Нейросетевые подходы	244
9.1.3. Аппаратная реализация	249
9.2. Распознавание эмоций как фактор непрерывной аутентификации	250
9.2.1. Алгоритмизация процесса распознавания эмоций	251
9.2.2. Нейросетевые подходы	254
9.2.3. Аппаратная реализация	256
9.3. Динамика клавиатурного почерка	257
9.3.1. Статистические методы анализа динамики нажатия клавиш	259
9.3.2. Нейросетевые методы анализа динамики нажатия клавиш	261
9.3.3. Аппаратная реализация	261

9.4. Распознавание личности по голосу	263
9.4.1. Математическая постановка задачи	264
9.4.2. Нейросетевые подходы к голосовой аутентификации	268
9.4.3. Аппаратная реализация	270
Контрольные вопросы	271
Практические задания	272
Глава 10. Перспективные и гибридные методы биометрической аутентификации личности	274
10.1. Аутентификация по словесному описанию	274
10.1.1. Математическое обоснование метода	275
10.1.2. Нейросетевая архитектура метода	277
10.2. Фоторобот	279
10.2.1. Математическое обоснование метода	280
10.2.2. Нейросетевая архитектура на основе GAN	281
Контрольные вопросы	282
Практические задания	283
Глава 11. Безопасность биометрических систем	285
11.1. Угрозы и атаки на биометрические системы	285
11.1.1. Классификация угроз и атак на биометрические системы	285
11.1.2. Физические спуфинг-атаки	287
11.1.3. Атаки на каналы передачи и обработки данных	288
11.1.4. Атаки на базы биометрических шаблонов	290
11.2. Методы защиты	291
Контрольные вопросы	292
Практические задания	292
Раздел III. СОЦИАЛЬНАЯ ИНЖЕНЕРИЯ	299
Глава 12. Теоретические основы и векторы атак социальной инженерии	301
12.1. Психологические принципы манипуляции	301
12.2. Классические векторы атак	303
12.3. Эволюция методов в цифровую эпоху	308
12.4. Цели и последствия успешных атак	310
Контрольные вопросы	312
Практические задания	313

Глава 13. Генеративный ИИ для создания фишинговых атак	316
13.1. LLM для генерации персонализированных фишинговых писем	316
13.2. Динамическая адаптация контента на основе OSINT	317
13.2.1. Сбор и анализ данных (OSINT-разведка)	318
13.2.2. Интеграция OSINT с генеративным ИИ	319
13.2.3. Динамическая адаптация и жизненный цикл атаки	320
13.3. Детектирование ИИ-генерируемого фишинга	322
13.3.1. Анализ контента	322
13.3.2. Поведенческий и технический анализ	323
13.3.3. Проактивные стратегии и организационные меры	324
Контрольные вопросы	325
Практические задания	326
Глава 14. Технологии Deepfake	329
14.1. Определение технологии Deepfake	329
14.2. Архитектурные основы Deepfake: от GAN к Few-Shot-обучению	330
14.3. Классификация deepfake-контента и сферы его применения	334
Контрольные вопросы	335
Практические задания	335
Глава 15. Голосовой спуфинг и аудиодипфейк	339
15.1. Технологии синтеза и клонирования голоса	339
15.2. Методы детекции аудиодипфейков	341
15.3. Атаки на системы биометрической аутентификации	344
Контрольные вопросы	345
Практические задания	346
Глава 16. Видеоспуфинг и видеодипфейк	349
16.1. Нейросетевые архитектуры для генерации видео	349
16.2. Детекция видеодипфейков	350
16.3. Защита систем видеоидентификации от дипфейков	352
Контрольные вопросы	354
Практические задания	355

Раздел IV. ОБНАРУЖЕНИЯ УГРОЗ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ИНЦИДЕНТОВ	358
Глава 17. Фундаментальные задачи анализа угроз информационной безопасности	360
17.1. Классификация вредоносного программного обеспечения	360
17.2. Компьютерные вирусы	361
17.2.1. Особенности и виды компьютерных вирусов ..	361
17.2.2. Классические методы распознавания вирусов ..	362
17.2.3. Распознавание вирусов с применением искусственной иммунной системы	365
17.3. Распознавание вредоносных Android-приложений ...	367
17.3.1. Задача обнаружения вредоносных Android-приложений	367
17.3.2. Нейросетевые подходы к детекции вредоносных Android-приложений	370
17.4. Распознавание спама	371
17.4.1. Задача обнаружения спама	371
17.4.2. Искусственная иммунная система распознавания спама	375
17.5. Обнаружение угроз облачных вычислительных систем	377
17.5.1. Задача защиты облачных вычислительных систем	377
17.5.2. Искусственная нейронная сеть обнаружения угроз облачных вычислительных систем	378
Контрольные вопросы	381
Практические задания	382
Глава 18. Сетевые системы обнаружения аномалий ..	388
18.1. Анализ сетевого трафика: особенности и классические методы	388
18.1.1. Особенности сетевого трафика как объекта анализа	388
18.1.2. Классические методы обнаружения аномалий в сетевом трафике	389
18.2. Нейросетевые подходы к анализу сетевого трафика ..	390
18.2.1. Обработка потоковых данных рекуррентными сетями	391
18.2.2. Применение сверточных сетей для анализа трафика	393

18.2.3. Автокодировщики для выявления сетевых аномалий	395
18.3. Обнаружение DDoS-атак с помощью рекуррентных сетей	396
18.4. Выявление скрытых угроз методами обучения без учителя	398
18.5. Детектирование аномалий в потоковых данных в реальном времени	399
Контрольные вопросы	402
Практические задания	402
Глава 19. Поведенческий анализ конечных устройств	405
19.1. ML-методы обнаружения вредоносного ПО	405
19.2. Анализ поведения процессов с помощью алгоритма изолирующих лесов	409
19.3. Детектирование атак, использующих легитимные инструменты	411
19.4. Прогнозный анализ угроз на основе временных рядов	415
Контрольные вопросы	417
Практические задания	418
Глава 20. Интеллектуальные SIEM-системы	422
20.1. Корреляция событий безопасности с помощью графовых нейросетей	422
20.2. NLP для анализа логов и инцидентов	425
20.3. Автоматическая классификация инцидентов ИБ	428
20.4. Прогнозирование векторов атак	430
Контрольные вопросы	432
Практические задания	432
Глава 21. SOAR-платформы и автоматизация реагирования	435
21.1. Оптимизация сценариев реагирования с использованием ИИ	435
21.2. Генеративные модели ИИ для создания сценариев автоматизации	437
21.3. Динамическое планирование контрмер	438
21.4. Оценка эффективности мер защиты	442
Контрольные вопросы	444
Практические задания	445

Раздел V. АТАКИ НА СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	450
Глава 22. Модель угроз	452
22.1. Активы систем искусственного интеллекта	452
22.2. Профиль нарушителя	454
22.3. Цели атаки: интерпретация триады CIA для ИИ	456
22.4. Векторы и сценарии угроз (Как осуществляется атака?)	458
Контрольные вопросы	460
Практические задания	461
Глава 23. Таксономия атак на системы искусственного интеллекта	464
23.1. Описание таксономии атак	464
23.2. Классификация атак	465
23.3. Практическое применение системы координат: анализ и противодействие угрозам	467
Контрольные вопросы	468
Практические задания	469
Глава 24. Конкретные примеры атак и их анализ в координатах таксономии	471
24.1. Data Poisoning	471
24.2. Adversarial Examples	473
24.3. Membership Inference	475
24.4. Model Stealing	477
Контрольные вопросы	478
Практические задания	479
Литература	481