МОДЕЛЬ КОМПЕТЕНЦИЙ В СФЕРЕ БОЛЬШИХ ДАННЫХ И ИСКУСТВЕННОГО ИНТЕЛЛЕКТА

АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ

Компетенция	Data Engineer	ML Engineer	Data Scientist	Data Analyst	Технологический стек	Методологический стек
БЛОК КОМПЕТЕНЦИЙ 1. БАЗОВЫЕ КОМПЕТЕНЦИИ РАБОТЫ С ДАННЫМИ						
Программирование на Python для задач работы с данными	2	2	3	2	Jupyter Notebook, Pandas, NumPy	Построение предиктивных моделей, демонстрация полученных результатов
Использование языков программирования и алгоритмической базы	2	2	2	2	C/C++	Алгоритмы и структуры данных
Использование SQL для работы с данными	2	2	2	2	SQL	Нормальные формы БД, вложенные запросы
Понимание ключевых метрик качества моделей, данных, дрифта, скорости ответа	2	3	3	1	Pandas, PyPlot	PSI, F1-мера, ROC-AUC, дрифт данных, RPS
Работа с готовыми генеративными моделями, написание базовых промптов	1	2	2	2	DeepSeek, GigaChat, YandexGPT	Базовые генеративные модели, понимание разницы и ограничения на модели
БЛОК КОМПЕТЕНЦИЙ 2. СБОР, ХРАНЕНИЕ И ИНТЕГРАЦИЯ ДАННЫХ						
Работа с экосистемой больших данных и распределенной обработкой данных	3	1	2	1	Apache Hudi, HDFS, YARN, Hadoop, Hive, Impala, Spark, PySpark	-
Проектирование и реализация data-пайплайнов	3	1	1	1	Apache Airflow, NiFi, Sqoop, Oozie	Знание ELT и ETL процессов
Работа с реляционными и аналитическими СУБД	3	1	1	2	ClickHouse, Greenplum, MS SQL, Oracle, PostgreSQL	Навык чтения карт отображений, понимание нормальных форм и ключей таблиц
Работа с нереляционными БД	2	1	1	2	S3, графовые БД, колоночные БД	Хранение и индексирование информации
Проектирование хранилищ, витрин и архитектура данных	2	0	0	1	Visio	Понимание первичного ключа и нормальных форм. Подход BPMN
Оптимизация запросов и производительности работы с данными	2	1	1	2	-	Понимание индекса, партиций
Программирование для распределенной обработки данных	3	1	1	0	Scala, Java, Kafka	Понимание работы высоконагруженных систем
Работа с потоковыми данными	3	2	0	0	Kafka	Понимание брокера сообщений инструментов обработки очереди FIFO, LIFO
Интеграция данных из внешних и внутренних источников	2	1	1	2	API, OpenAPI, GraphQL	-
Работа с хранилищами данных	3	2	2	1	S3, Hadoop	Понимание разницы в типах данных, типах обработки и хранения данных. Знание нормальных форм, индексирования, партиций
БЛОК КОМПЕТЕНЦИЙ 3. ОБЕСПЕЧЕНИЕ КАЧЕСТВА, БЕЗОПАСНОСТИ И ЭТИЧНОГО ИСПОЛЬЗОВАНИЯ ДАННЫХ						
Ознакомление с нормативными документами Регулятора по разработке моделей	1	1	1	1	3-МР, 243-ФЗ	-
Обеспечение качества, очистки и подготовки данных	3	2	2	3	-	Подготовка проверок при создании таблиц и источников для моделей
Цензор-модули LLM	0	1	1	1	-	Понимание архитектуры взаимодействия LLM-моделей
БЛОК КОМПЕТЕНЦИЙ 4. АНАЛИТИКА ДАННЫХ И BI						
Разработка BI-отчетности, дашбордов и визуализаций	2	1	2	3	Excel, Tableau, Grafana	Понимание построения OLAP-кубов
Анализ данных и интерпретация результатов	1	1	2	3	-	-
Проектирование метрик и отчетности для LLM-моделей	1	1	2	1	-	Оценка необходимости ручной верификации, понимание работы цензор-модуля
Проектирование метрик и отчетности	2	1	3	3	-	-
Формирование бизнес-выводов на основе данных	1	1	2	3	-	-
Проведение продуктовой аналитики	0	1	3	3	-	-
БЛОК КОМПЕТЕНЦИЙ 5. СТАТИСТИКА, ЭКСПЕРИМЕНТЫ И МАТЕМАТИЧЕСКИЕ МЕТОДЫ						
Проведение и анализ экспериментов	1	1	2	2	NumPy	Применение статистических распределений при оценке результативности
Прогнозирование и анализ временных рядов	0	1	2	2	-	-
Проведение A/B-тестов	1	1	3	2	-	Применение статистических распределений при оценке результативности
БЛОК КОМПЕТЕНЦИЙ 6. МАШИННОЕ ОБУЧЕНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ						
Подготовка эксперимента	1	1	2	2	Python	Умение подготовить пилот для понимания необходимости масштабировать решение
Разработка и применение моделей машинного обучения без применения GenAI	1	2	3	1	-	-
LLM-решения	1	2	2	2	Использование внутренних решений ОС	Умение формулировать системные промпты. Понимание возможностей передачи данных во внешние сервисы
Разработка NLP	0	1	2	1	-	Оценка финансовой целесообразности применения LLM-моделей. Учёт галлюцинаций в NLP или GenAI
Разработка рекомендательных и ранжирующих систем	0	3	2	0	-	-
Разработка моделей глубокого обучения	0	2	2	0	-	-
Формирование признакового пространства	2	2	2	1	-	-
Использование ML/DL-фреймворков	0	3	2	0	-	-
БЛОК КОМПЕТЕНЦИЙ 7. РАЗРАБОТКА И ЭКСПЛУАТАЦИЯ DATA/AI-РЕШЕНИЙ						
Работа с инструментами разработки и версионирования	2	2	2	2	Bash, Linux, GIT, IDE, Unix, CLI	-
Внедрение и сопровождение ML/AI-решений	2	2	2	0	FastAPI, Yandex Cloud	-
Автоматизация разработки, тестирования и доставки решений	3	2	2	2	-	-
Развертывание решений в контейнерных и облачных средах	3	2	3	0	Docker, Kubernetes, OpenShift	-
Мониторинг, логирование и наблюдаемость решений	3	2	0	2	Tableau, Grafana	-
БЛОК КОМПЕТЕНЦИЙ 8. УПРАВЛЕНИЕ ПРОЕКТАМИ И ПРОДУКТАМИ						
Документирование и презентация результатов	2	2	2	3	Confluence	Структурирование версий отчетов
Построение плана работ	1	1	1	2	Project, Jira	Построение плана работ, умение вести спринты и отслеживать прогресс по задачам, декомпозировать сложные задачи, работа с матрицей RACI
Коммуникация со стейкхолдерами	1	1	2	2	Confluence, Jira, Project	Структурирование ведения проекта, определение ключевых точек, согласование работ по матрице RACI
Понимание предметной области и бизнес-контекста	2	1	3	3	Confluence, Jira	Умение пользоваться технической документацией, Confluence, версионностью
Определение уровня сложности и самостоятельности задач	2	0	3	0	-	
Постановка и формализация задач	1	1	3	2	-	
Проактивность и развитие решений	0	0	0	3	-	

Необходимый уровень владения компетенцией

0	не является целевой	1	базовый	2	продвинутый	3	экспертный