

Памятка: что проверить перед внедрением Azure OpenAI через Foundry

Памятка предназначена для команд, которые переводят эксперимент с OpenAI в корпоративный контур и должны согласовать модель, deployment, безопасность, стоимость и эксплуатацию.

Архитектура и deployment

- Определить, через какой Azure-контур доступна модель.
- Проверить deployment name, endpoint, окружения dev, test и prod.
- Назначить владельца deployment и ответственных за эксплуатацию.
- Описать lifecycle: создание, обновление, мониторинг и вывод из эксплуатации.

Безопасность и governance

- Проверить, какие данные отправляются в модель.
- Описать роли доступа и хранение ключей.
- Зафиксировать требования к журналированию и incident response.
- Проверить human oversight и правила ответственного использования.

Стоимость и закупочный контур

- Сравнить OpenAI direct и Azure Foundry по договорной рамке, поддержке и внутренним процедурам.
- Понять, как потребление попадает в Azure billing.
- Настроить лимиты, бюджет и наблюдение за расходами.
- Проверить SLA и доступную поддержку Microsoft.

Качество и production

- Подготовить evaluation-наборы и метрики качества.
- Проверить сценарий на ошибочные ответы и ограничения модели.
- Не считать наличие Foundry Models доказательством готового внедрения.
- Запланировать наблюдение в production и процедуру отката.