Бюджет нейропомощника: расчёт Б791 и рабочие листы

Б791: тарифная карточка, три нагрузки и полная смета с обслуживанием.



Практический гайд с заполненным учебным примером и чистыми рабочими листами. Подготовлено редакцией Нейросистемы риелтора.

Учебные цифры не являются данными о текущем рынке. Обложка представляет концептуальную иллюстрацию.

Б791: паспорт бюджета

Самостоятельный комплект «Нейросистемы риелтора». Нагрузка, расходы платформы и обслуживания вымышлены. Только тарифная строка модели проверена в первичном источнике. Подписки не покупались, модель не запускалась.

Граница	Условие
Назначение	Расчёт эксплуатации текстового помощника
Учёт	Платформа и модель оплачиваются раздельно
Нагрузка	400 диалогов, 5 основных вызовов в среднем
Дополнительно	10% оплачиваемых вызовов
Не включено	Отдельный поиск, голос, изображения и другие инструменты

Моя функция / период	Владелец бюджета

Нагрузка: от диалогов к вызовам

Шаг	Расчёт
Основные вызовы	400 × 5 = 2 000
Дополнительные оплачиваемые	2 000 × 10% = 200
Всего	2 200
Вход на вызов	1 500 токенов, условно
Выход на вызов	300 токенов, условно
Итоговый вход	3 300 000 токенов
Итоговый выход	660 000 токенов

Что считается
Диалог не равен посетителю, заявке или одному вызову. 10% здесь не обозначает долю серверных ошибок; это дополнительные оплачиваемые обращения.

Мой объём / источник оценки	Что пока является предположением

Тарифная карточка для расчёта

Поле	Проверено 19.10.2026
Провайдер	Yandex AI Studio
Модель	YandexGPT Lite
Режим	Синхронный
Вход	0,2 ₽ за 1000 токенов
Выход	0,2 ₽ за 1000 токенов
Налоговый признак	С НДС
Источник	Официальная страница тарификации

Граница использования
Это пример подстановки тарифа, не рекомендация модели для всех задач. Качество ответов проверяется отдельно. Перед подключением сверить действующую строку и дополнительные услуги.

Мой тариф / дата / единица	Что включено / что оплачивается отдельно

Стоимость генерации Б791

Компонент	Формула	Итого
Вход	3 300 000 / 1000 × 0,2	660 ₽
Выход	660 000 / 1000 × 0,2	132 ₽
Сумма	660 + 132	792 ₽

Без необоснованной скидки
Кеширование и отдельные инструменты не заложены в эту простую модель нагрузки. В реальной смете нужно учитывать применимые виды потребления.

Вход: объём / ставка / сумма	Выход: объём / ставка / сумма

Другие начисления	Основание / сумма или не оценено

Полная месячная смета

Статья	Учебная сумма	Статус
Генерация	792 ₽	Расчёт на проверенном тарифе
Платформа	1 200 ₽	Условно
Инфраструктура	300 ₽	Условно
Обслуживание	2 ч × 600 ₽ = 1 200 ₽	Условно
Месяц	3 492 ₽	При заданной нагрузке
Разовая подготовка	6 000 ₽	Условно, отдельно
Первый месяц	9 492 ₽	Месяц плюс подготовка

Не считать дважды
Если ваш пакет включает модельное потребление, не добавляйте его повторно. В Б791 оно оплачивается отдельно.

3 492 / 400 = 8,73 ₽ на диалог в этом сценарии. Это не стоимость привлечённого клиента.

Три сценария объёма

Диалоги	Вызовы с +10%	Генерация	Месяц
200	1 100	396 ₽	3 096 ₽
400	2 200	792 ₽	3 492 ₽
800	4 400	1 584 ₽	4 284 ₽

Одинаковые допущения
5 основных вызовов на диалог; 1500 входящих и 300 исходящих токенов на вызов; фиксированная часть 2700 ₽. Пороги платформы и увеличение поддержки не учтены.

Мой порог нагрузки	Как меняется фиксированная часть

Вопросы исполнителю до запуска

Вопрос	Какое подтверждение нужно
Что входит в платёж?	Разработка, модель, платформа, сопровождение по строкам
Есть ли объёмные пороги?	Условия изменения пакета
Что считается вызовом?	Описание шагов и повторов
Как получим факт расходов?	Доступная отчётность по периоду
Как действует ограничение?	Проверенное поведение при достижении
Кто поддерживает базу?	Объём и стоимость обслуживания

Неоценённая строка	Кто уточнит / основание / срок

Чистый лист бюджета

Функция / месяц / число диалогов	Вызовы на диалог / дополнительные

Вход / выход / тарифные единицы	Модельная сумма / дата тарифа

Платформа / инфраструктура / поддержка	Месячная сумма / неучтённое

Разовый запуск	Первый месяц / условия пересмотра

Контроль фактических расходов

Наблюдение	Что разбирать
Вызовов больше плана	Повторы и лишние шаги
Вход растёт	История и передаваемые материалы
Иная ставка	Модель, режим и дата тарифа
Больше обслуживания	Причины ручных исправлений

Факт / план / отклонение	Причина / действие / ответственный

Качество не сокращается по смете
После уменьшения контекста проверяйте сохранение фактов и ограничений. Уведомление о бюджете не называйте автоматической остановкой без подтверждённого поведения.

Проверено 19.10.2026. Учебная смета не является коммерческим предложением. Перед использованием обновите ставки и измерьте потребление своего сценария.

Источники и проверка

Yandex AI Studio: правила тарификации