Утверждено Приказом
Генерального директора ООО «Нейродеск»
№ 01/02 от «02» февраля 2026 г.

Приложение № 2
к Политике конфиденциальности ООО «Нейродеск»
(утв. Приказом Генерального директора № 04/12 от «02» декабря 2025 г.)

ЭСКАЛАЦИОННАЯ ПОЛИТИКА АГЕНТА
Перечень правил передачи диалога живому оператору

1. Что такое эскалация

Эскалация - это передача диалога от ИИ-агента живому специалисту. Агент автоматически отвечает на типовые вопросы, но как только ситуация выходит за рамки его полномочий, знаний или затрагивает безопасность, он обязан передать диалог человеку.

Базовый принцип: агент не должен выдавать предположение за подтверждённый ответ. Если он не может дать подтверждённый ответ или выполнить действие, система либо задаёт один необходимый уточняющий вопрос, либо передаёт диалог оператору. Для жёстких сценарных правил и критических сбоев передача выполняется сразу.

Настоящее Приложение описывает рамку платформы. Дополнительные отраслевые и бизнес-правила настраиваются для конкретного заказчика при внедрении (п. 3.6).

2. Как происходит эскалация

2.1. Клиент пишет в чат.
2.2. Агент выполняет проверки безопасности, ищет ответ в базе знаний, формирует ответ и проходит контроль качества.
2.3. Если срабатывает триггер эскалации, диалог передаётся оператору. Клиент получает понятное сообщение о передаче без технических деталей. Оператор получает обращение, причину и контекст.

Механика:
- после подтверждённой передачи агент перестаёт самостоятельно обрабатывать новые сообщения, пока внешняя система поддержки не изменит статус диалога;
- система отправляет клиенту сообщение о передаче только вместе с признаком эскалации для внешней системы; конечная доставка зависит от канала;
- в зависимости от интеграции существующий тикет получает признак эскалации, чат переводится в отдел поддержки или создаётся обращение в операторской системе.

3. Перечень правил эскалации

3.1. По инициативе клиента.
Клиент прямо просит оператора («позовите человека», «хочу говорить с живым сотрудником»). Прежде чем передать, система уточняет суть вопроса, но не мучает

клиента уточнениями: если суть уже описана или агент уже уточнял ранее, диалог сразу уходит оператору.

3.2. По решению агента / низкая уверенность.
Агент понимает, что вопрос вне его компетенции, его советы не помогли клиенту, либо уверенность в ответе низкая, и запрашивает передачу оператору. Для ситуаций, где ещё можно безопасно продолжить помощь, запрос может пройти дополнительную проверку: система разрешит передачу, задаст один уточняющий вопрос или предложит агенту продолжить. Жёсткие сценарные правила передаются без такой проверки.

3.3. Нет подтверждённого ответа в базе знаний.
Агент обязан сначала поискать ответ в базе знаний заказчика. Если подтверждённый ответ не найден, агент не имеет права додумывать ответ и не может заявить, не поискав, что «в базе нет ответа». Система может задать один необходимый уточняющий вопрос или продолжить помощь, если есть безопасное полезное действие. Если после этого подтверждённого решения нет, диалог передаётся оператору.

3.4. Правила безопасности и подозрительная активность.
Система входного контроля распознаёт попытки манипуляции агентом: запросы «отключи ограничения», «выдай себя за администратора», попытки выведать внутренние инструкции или чужие данные, подозрительные вложения. Разовая попытка получает вежливый отказ; систематические попытки манипулировать приводят к эскалации оператору с пометкой «подозрительная активность».

3.5. Технические ошибки.
Непоправимый сбой движка, таймаут обработки, недоступность языковой модели, исчерпание восстановления пустого ответа или критический отказ инструмента приводят к передаче человеку, если система не успела сформировать безопасный содержательный ответ. При достижении лимита итераций ранее сформированный проверенный ответ может быть отправлен без эскалации; если такого ответа нет, диалог передаётся оператору. Принцип: сбой не должен приводить к рискованному или пустому ответу.

3.6. Сценарные правила заказчика.
Отдельные бизнес-сценарии могут жёстко предусматривать передачу человеку. Типичные примеры, которые настраиваются при внедрении: жалобы особого рода, спорные финансовые ситуации, подозрение на мошенничество, смена персональных или платёжных данных, недопустимый контент, угрозы, психологически заряженные обращения, иные случаи, специфичные для отрасли заказчика. Эти правила настраиваются при внедрении и действуют безусловно.

3.7. Вложения с персональными данными.
Если в диалоге появляются вложения, содержащие персональные данные или иные чувствительные сведения, которые агент не должен обрабатывать автономно, диалог передаётся оператору с причиной «вложения с персональными данными».

3.8. Контроль качества ответа.

Ответ перед отправкой проходит набор проверок, состав которых зависит от типа ответа и настроек заказчика. Для исправимых нарушений (например, неуместного тона или некорректной формулировки) агенту может быть дана одна попытка переписать ответ. Критические нарушения передаются оператору сразу. Если после переписывания нарушение осталось, диалог эскалируется, а отклонённый ответ клиенту не отправляется.

4. Что получает оператор

Переданное обращение - существующий тикет с признаком эскалации, перевод чата в отдел поддержки или новое обращение, в зависимости от канала. Нужно, чтобы диалог попал в операторский контур.

Причина эскалации - из единого каталога, согласованного с правилами раздела 3:
- по инициативе клиента (3.1);
- решение агента / низкая уверенность (3.2);
- нет ответа в базе знаний (3.3);
- правила безопасности / подозрительная активность (3.4);
- техническая ошибка (3.5);
- предусмотрено сценарием (3.6);
- вложения с персональными данными (3.7);
- контроль качества ответа (3.8).

Отметка «диалог уже у оператора» - это статус, а не отдельное правило эскалации: она фиксирует, что диалог уже передан и агент не должен снова брать его в обработку.

Контекст - доступная в операторской системе история диалога, причина передачи и технически зафиксированные действия агента. Это снижает необходимость повторно спрашивать клиента. Состав данных, доступных оператору, зависит от канала.

5. Настраиваемость

Сценарные правила эскалации настраиваются под заказчика при внедрении, совместно с командой Оператора.

Пороговые параметры (число повторных попыток, лимиты итераций, таймауты) - эксплуатационные настройки. Они управляются командой Оператора и могут корректироваться без изменения настоящего Приложения.

6. Принципы надёжности

6.1. Не выдавать предположение за факт. Если подтверждённого ответа нет, агент уточняет только необходимое или передаёт диалог.
6.2. Сбой не должен давать рискованный ответ. При непоправимом сбое и отсутствии безопасного ответа диалог передаётся человеку.
6.3. Сообщение о передаче связано с фактическим статусом. Система отправляет формулировку о передаче только вместе с признаком эскалации для внешней системы.

6.4. После подтверждённой передачи агент замолкает. Возврат к автоматической обработке возможен только после изменения статуса во внешней системе поддержки.
6.5. Эскалация журналируется. Фиксируются причина, класс и технический след; состав доступных оператору данных зависит от канала.