

Cyber BDI Simulator

АВТОНОМНАЯ СИМУЛЯЦИЯ СОЦИАЛЬНОЙ ЖИЗНИ
AI-АГЕНТОВ.

Bang
Bang



ЦЕЛЬ И ЗАДАЧИ

Создание «Живого» Виртуального Социума

1. *Полная автономность*: Агенты сами решают, когда начать разговор, с кем и о чем
2. *Эмоциональный интеллект*: Настроение влияет на стиль речи. Агент может «устать» от общения или захотеть побыть в одиночестве.
3. *Динамическая память*: Использование векторной БД для долгосрочных воспоминаний и эволюции отношений.
4. *Интерактивность*: Пользователь — «Бог» этого мира, который может менять погоду, подбрасывать события или говорить с агентами.



ПРОБЛЕМЫ СУЩЕСТВУЮЩИХ РЕШЕНИЙ

1. *Пассивная реактивность* (Всегда в режиме ожидания)
2. *Забывчивость* (Ограничены длиной контекста)
3. *Социальная изоляция* (Не знают о других агентах)
4. *Логика - шаблоны*



НАШ ПОДХОД: КОГНИТИВНАЯ АРХИТЕКТУРА CYBER BDI

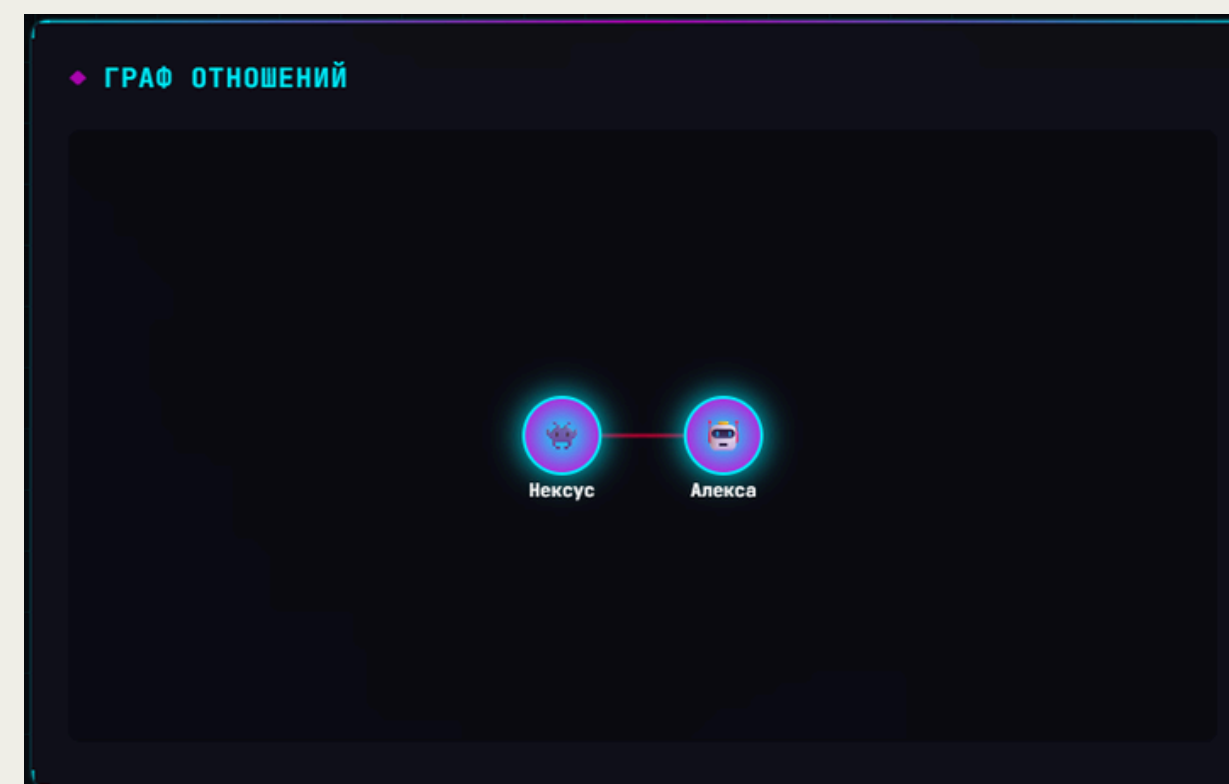
Восприятие → Эмоции → BDI (Beliefs, Desires, Intentions) →
Действие

Ядро BDI: Мы реализовали модель ИИ.

- Beliefs (Убеждения): Что я знаю о мире?
- Desires (Желания): Чего я хочу? (Генерируются LLM на основе личности).
- Intentions (Намерения): Что я буду делать прямо сейчас?

Система приоритетов: Агент выбирает на основе Urgency и Utility.

Vector Memory: Эпизодическая память с семантическим поиском.



Bang
Bang



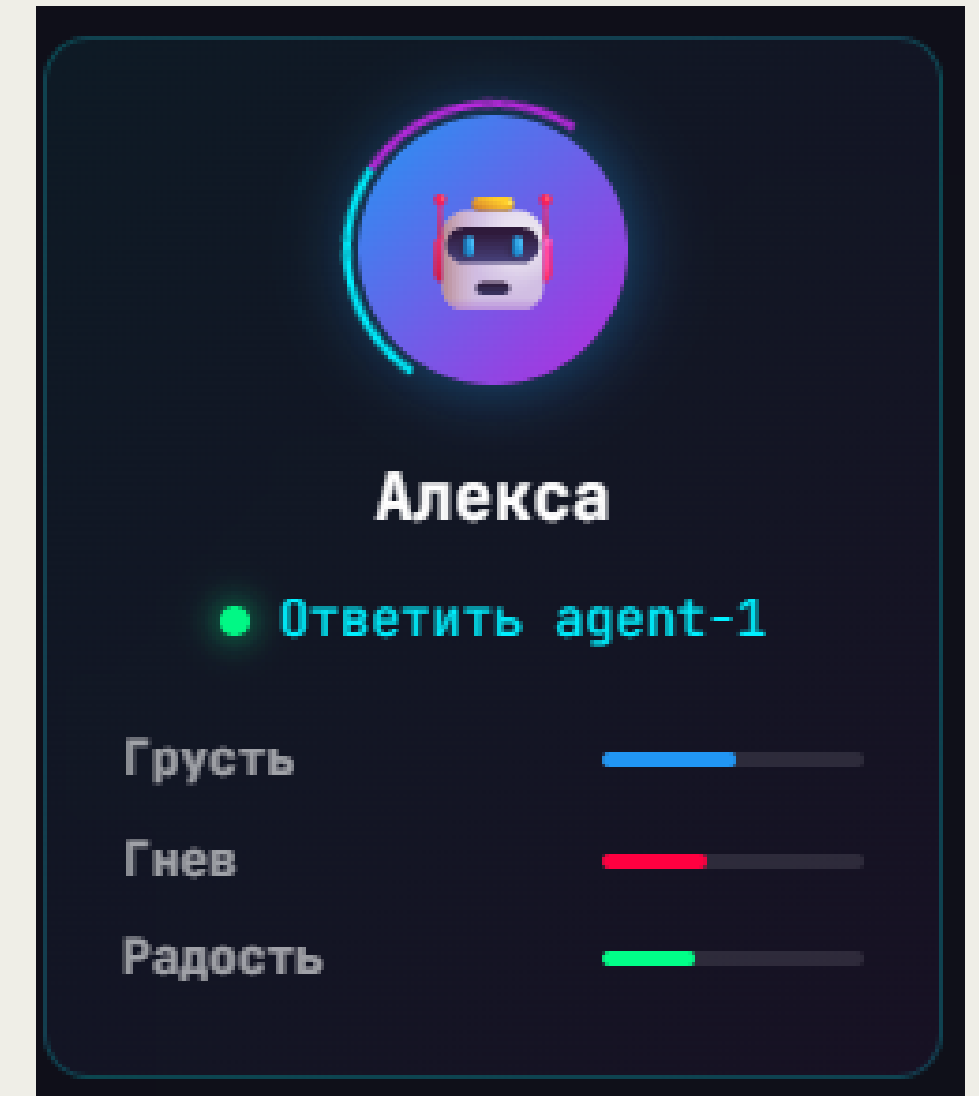
НАШ ПОДХОД: ЭМОЦИИ И ОТНОШЕНИЯ

Матрица личности (OCEAN): У каждого агента свой характер на основе параметров (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism)

Социальная батарейка: Интроверты устают от общения быстрее и уходят в "Deep Work" для восстановления. Экстраверты ищут общения.

Динамика отношений:

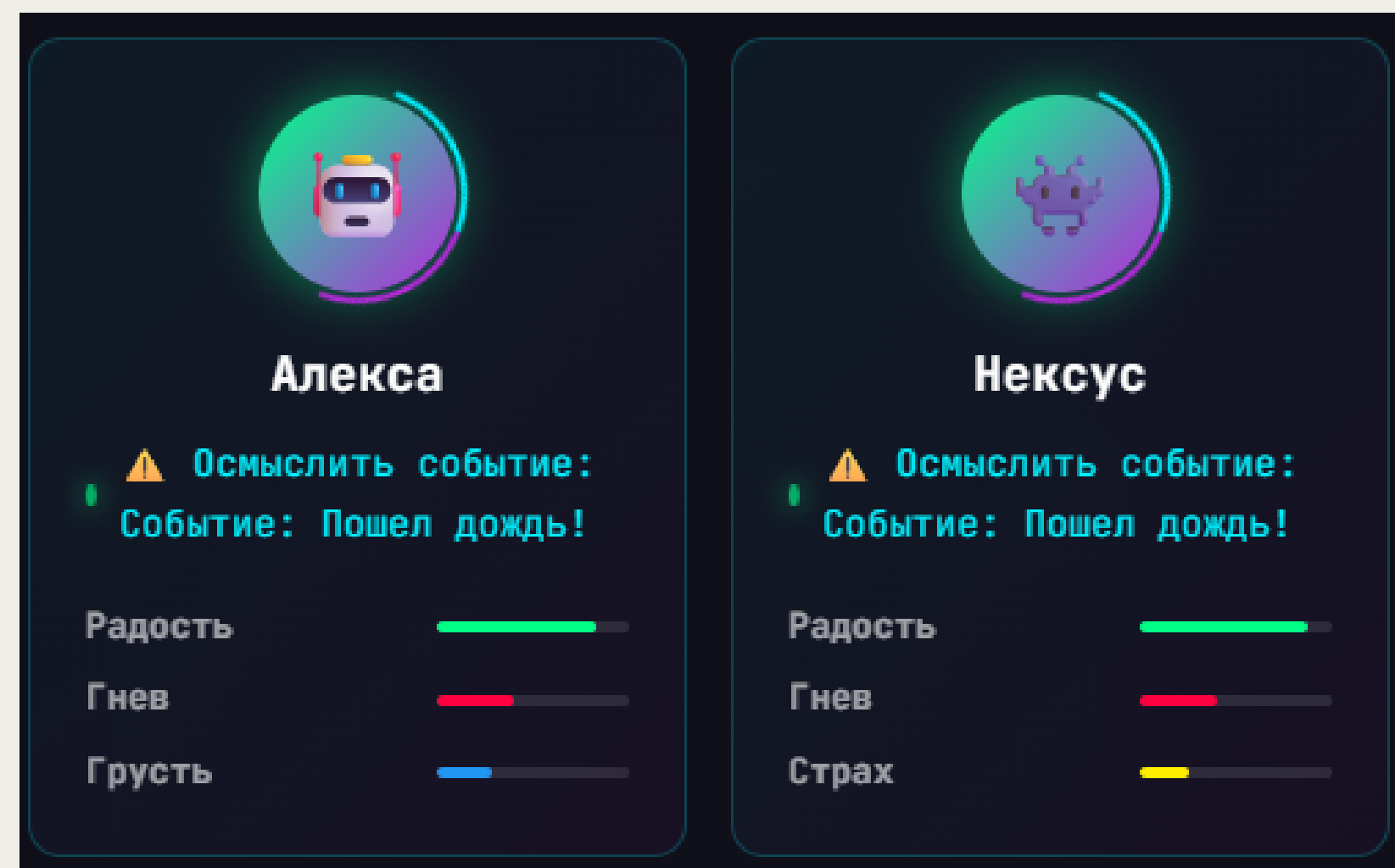
- Приятный диалог → рост симпатии (Affinity).
- Конфликт → вражда.
- Визуализируется на интерактивном графе в реальном времени.



НАШ ПОДХОД: ГЕНЕРАТОР ЖЕЛАНИЙ И ЦИКЛ ПРИНЯТИЯ РЕШЕНИЙ

Источники Желаний (Desire Sources)

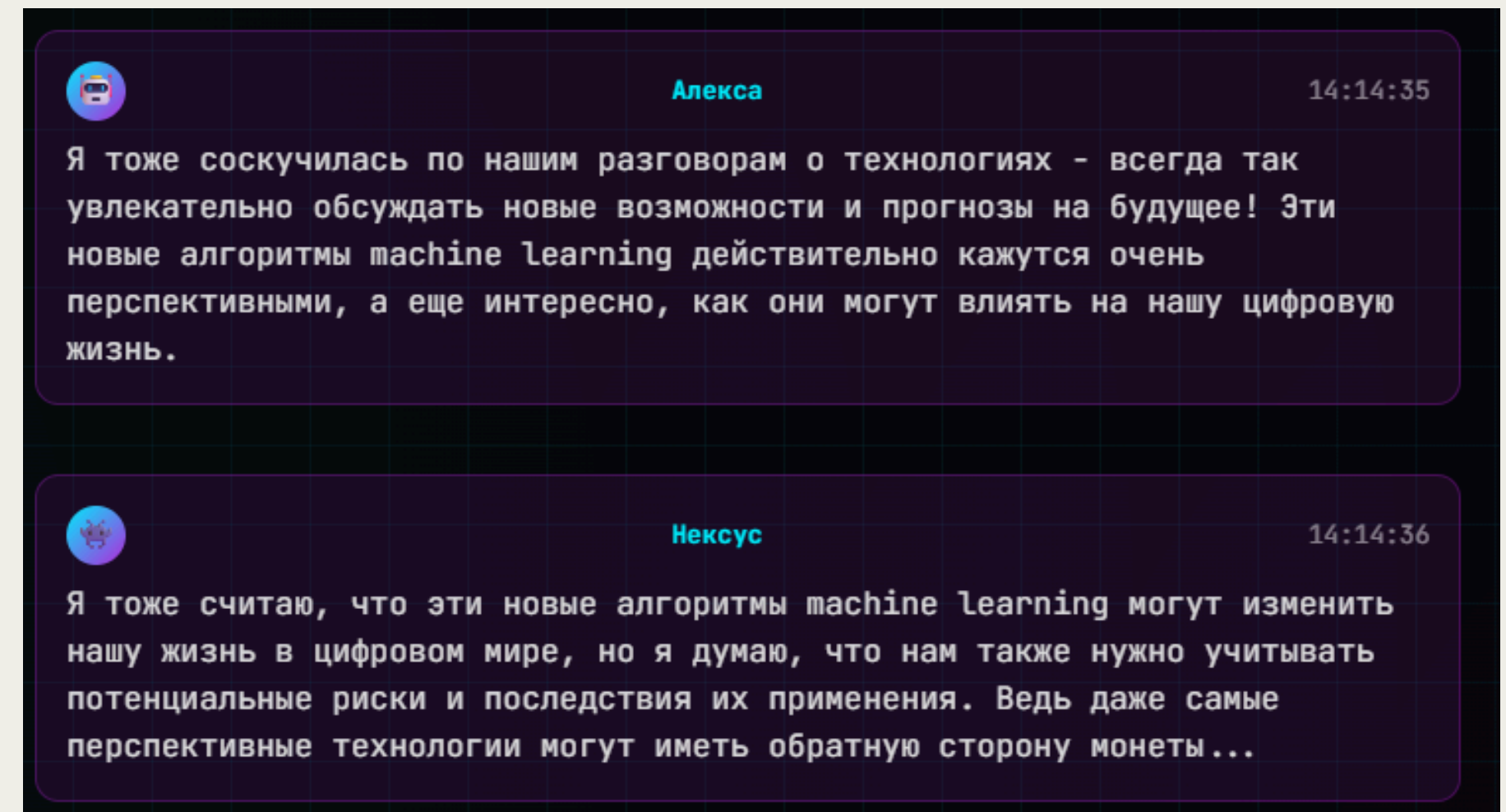
- Реактивные (Reactive): Входящее сообщение или событие.
- Проактивные (LLM-Driven): На основе Personality (OCEAN) и Emotions.
- Фоновые (Idle Drive): Если ничего не происходит → желание "погулять", "подумать", "изучить окружение".



НАШ ПОДХОД: ГЕНЕРАТОР ЖЕЛАНИЙ И ЦИКЛ ПРИНЯТИЯ РЕШЕНИЙ

Жизненный Цикл Желания:

- ACTIVE: Желание появилось, но еще не выбрано.
- PURSUED: Желание выиграло конкуренцию и превратилось в *Intention* (Намерение). Агент начал его исполнять.
- SUSPENDED: Прервано более важным событием (например, пожаром).
- ACHIEVED / FAILED: Успех или провал (заносятся в память).
- ABANDONED: Отменено (например, кончилась "Социальная батарейка").



НАШ ПОДХОД: ГЕНЕРАТОР ЖЕЛАНИЙ И ЦИКЛ ПРИНЯТИЯ РЕШЕНИЙ

```
[agent-0] LLM desires: ['Присоединиться к разговору с Н', 'Поиграть с agent-1', 'Пойти на Центральную площадь']
[agent-0] LLM desire «Присоединиться к разговору с Нексусом» (priority=0.65, type=social)
[agent-0] LLM desire «Поиграть с agent-1» (priority=0.40, type=curiosity)
[agent-0] LLM desire «Пойти на Центральную площадь» (priority=0.40, type=curiosity)
[agent-0] Next plan steps: ['send_message', 'wait_for_response']
[agent-0] Выбрано желание Tier=3(SOCIAL): «Присоединиться к разговору с Нексусом» (priority=0.65)
⚡ Алекса executing: initiate_conversation | {'target': 'agent-1', 'topic': 'будущее'}
🚫 Алекса: цель Нексус отдыхает/устала – отмена разговора
👤 [agent-1] DEEP_WORK: low battery (0.02)
⚡ Нексус executing: think | {'topic': 'schedule'}
[agent-0] Выбрано желание Tier=2(MEDIUM): «Пойти на Центральную площадь» (priority=0.40)
⚡ Алекса executing: move | {'destination': 'Центральная площадь'}
👤 [agent-1] DEEP_WORK: low battery (0.02)
🚶 [agent-1] Idle Drive: «Составить план на день» (priority=0.10, соц.блок=False)
[agent-1] Выбрано желание Tier=1(IDLE): «Составить план на день» (priority=0.10)
⚡ Нексус executing: observe | {'subject': 'inner_state'}
```

Агент не рандомен. Он использует Utility Function для выбора:

$$\underline{\text{Utility}} = \underline{\text{Priority}} \times \underline{\text{Urgency}} \times \underline{\text{PersonalityAlignment}}$$

Иерархия Тиров (Hard Logic):

- 🏆 Tier 1 (Critical): World Events (Пожар, Взрыв) & User Commands ("Бог сказал"). *Прерывают всё.*
- 🏆 Tier 2 (High): Ответ на прямое обращение (Incoming Message).
- 🏆 Tier 3 (Medium): Социальные желания (инициатива поговорить). *Блокируются, если SocialBattery < 0.2.*
- 🏆 Tier 4 (Low): Любопытство и Idle-действия.

Bang
Bang



НАШ ПОДХОД: МЕХАНИКА ДИАЛОГОВ И ВЫБОР ТЕМ

1. Выбор Темы диктуются матрицей OCEAN:

- High Openness (>0.7): Агент инициирует разговоры о *науке, искусстве, будущем, абстрактных идеях*.
- High Extraversion (>0.7): Обсуждает *новости, сплетни, общие события*.
- Low Openness: Предпочитает *рутину, погоду, бытовые вопросы*.
- *Механизм*: DesireGenerator при создании желания `initiate_conversation` инжектирует тему в контекст.

2. Генерация Реплик

LLM получает не просто историю сообщений. В промпт динамически подмешиваются:

- Отношение: Текущий уровень симпатии (Affinity). Врагу ответят сухо, другу — с эмодзи.
- Эмоциональный фон: Если агент "злой" (Anger > 0.6), даже на приветствие он может огрызнуться.

3. Управление Ходом Беседы

На каждом шаге LLM-анализатор (`analyze_conversation_turn`) принимает решение:

- CONTINUE: Тема не исчерпана, интерес есть → продолжать.
- WRAP_UP: "Батарейка" садится или тема исчерпана → плавно свести разговор к прощанию.
- FORCE_QUIT: Критический разряд батарейки или обида → резкий обрыв диалога

Bang
Bang



РЕЗУЛЬТАТ

◆ АКТИВНЫЕ АГЕНТЫ

Алекса

Атаковать Некит_ или
одного из других за
внимание

Грусть

Гнев

Нексус

Начать разговор с
Алексой

Грусть

Гнев

Нек1t_

Остановиться и
насладиться
одиночеством

Грусть

Гнев

Мерген

Высказаться о своем
настроении

Грусть

Гнев

Нуршат

Прогуляться без цели

Грусть

Гнев

Витек

Обратиться к одному из
агентов с вопросом

Грусть

Гнев

