

---

# Экзамен для роботов: что и как измеряет тест Тьюринга (по итогам теста 2015 года в Сколково)

Анна Власова,  
«Наносемантика»  
[anna@nanosemantics.ru](mailto:anna@nanosemantics.ru)



# Тест Тьюринга на русском языке

- Дата проведения: 2-3 июня 2015
- Организаторы: ИФ «Сколково», «Наносемантика»
- Технологическая площадка: «Наносемантика»
- Участники: 69 команд
- Время на подготовку бота: 3 месяца (с 16 марта)

# Компания «Наносемантика»

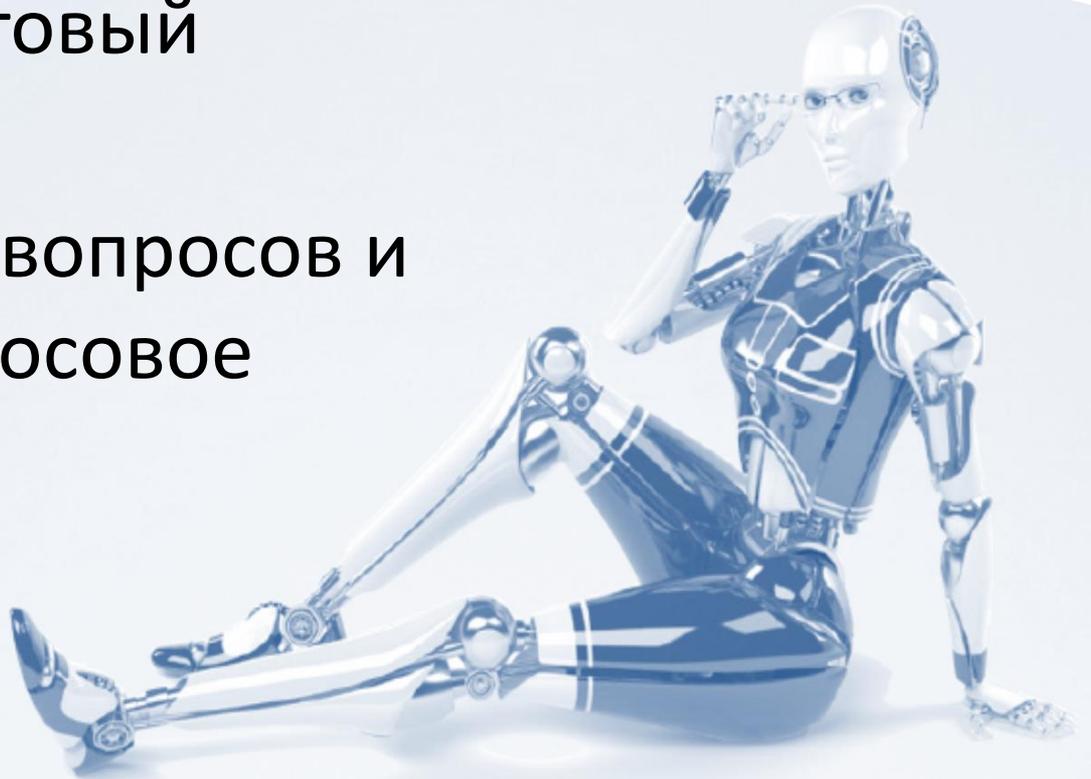
---

- Основана в 2005 году
- Занимается разработкой автоматизированных систем, поддерживающих диалог на естественном языке
- Основа технологии – инфы или виртуальные собеседники, способные вести диалог на разных языках в соответствии с заданной коммуникативной моделью
- Отдел разработки = лингвистический отдел (выпускники МГУ, РГГУ, МГОУ, СГУ)
- Поддерживает 2 сообщества разработчиков диалоговых систем (виртуальных собеседников): [iii.ru](http://iii.ru), [inf.net](http://inf.net)



# ИНФ (виртуальный собеседник)

- Лингвистические технологии
- Движок (диалоговый процессор)
- Онлайн-сервис вопросов и ответов или голосовое устройство



# Как это выглядит?

За каждого подключенного друга мы зачислим на Ваш счет 500 рублей, которые Вы сможете потратить на оплату своего Интернета, Цифрового телевидения, а также на другие услуги NETBYNET! Подробнее об акции "Приведи друга" Вы можете прочитать [здесь](#).



Акция приведи друга?  

**Виртуальный консультант** 



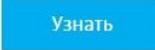
Ставка зависит от выбранной программы кредитования. По [Кредиту наличными](#) ставка – от 18%. Ознакомьтесь с основными условиями предоставления кредита наличными и рассчитайте максимальную сумму кредита

как узнать процент по кредиту 

## Виртуальный помощник

Ответит на все ваши вопросы

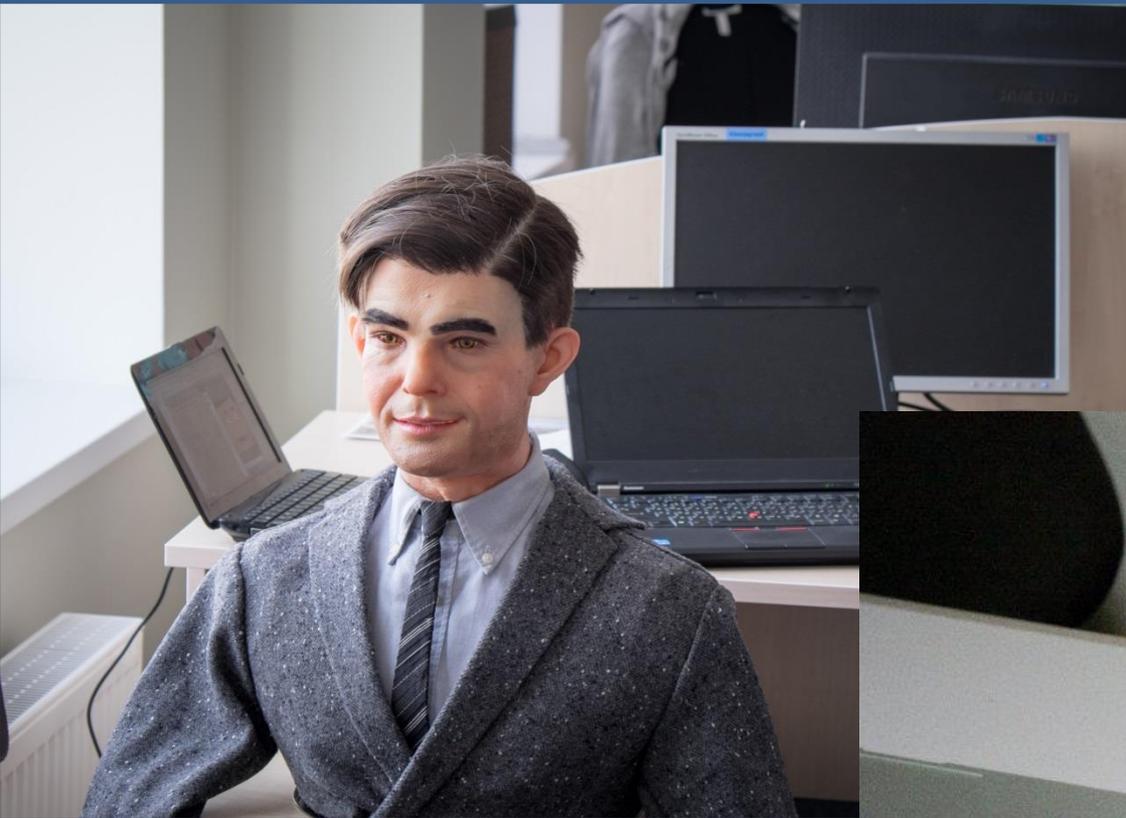
Мобильная связь **Интернет**

плохая связь 

Ответ виртуального помощника

Проблемы со связью могут возникать при перемещении через плохую зону покрытия, из-за неполадок в работе телефона. Проверьте сигнал сети на вашем аппарате и зону покрытия. Шкала делений сигнала расположена в левом или правом верхнем углу экрана мобильного устройства. Под плохим сигналом понимается все, что менее половины шкалы. Убедитесь в том, что [включена передача данных](#) и [прописаны настройки точки доступа APN](#). Если это так, то попробуйте изменить

# Как это выглядит?



# Что инфы говорят?

---

Inf: Рада Вас видеть. Если есть вопросы, постараюсь помочь.

User: какой необходимо набрать код что бы узнать свой номер?

Inf: Чтобы проверить номер своего телефона, наберите: \*160# и клавишу дозвона.

User: спасибо

Inf: Всегда пожалуйста! :)

User: уу так быстро отвечаете

Inf: Виртуальные консультанты проходят спецподготовку. :)



# Что инфы говорят?

---

Inf: Сегодня понедельник. Расскажи роботу, как ты проводишь понедельники.

User: ну сижу за компьютером кушаю

Inf: Робот считает, что по понедельникам надо отрываться от компьютера и выходить в реальный мир.

User: нет реальный мир страшный

Inf: Робот удивлен.

...



# Что делают лингвисты?

**An do:** английский? вьетнамский? китайский?

понедельник  
панедельник  
понидельник  
панидельник  
понеделник  
...

User: Здор<sup>о</sup>во, приятель!

Inf: Здравствуйте! Я тоже рад Вас видеть.

...

User: Здор<sup>о</sup>во ты на вопросы отвечаешь!

Inf: Спасибо за комплимент!

User: Да ты тормоз!

...

User: У вас не связь, а тормоз.

...

User: Тормоз барахлит

Inf : Ты любишь читать?

User: нет

User\*: (реконстр.): нет, я не люблю читать

Inf : Жаль, а я люблю читать!

# DL: формальный язык для описания модели диалога

Inf: Давай поиграем в горячо-холодно.  
Угадай, сколько я знаю слов?

User: 100

Inf: Я кажусь таким глупым ☹️? Нет, это совсем холодно.

User: мультен?

Inf: Ты на верном пути, но это тоже холодно.

User: Ммм... 300?

User: Тебе нравится мадонна?

Inf: Мадонна? Не слушаю. Из американской попсы уважаю Леди Гагу.

```
# [if ( %GAME_COLD_HOT_VALUE = "[&1]")]
{
  В яблочко!
  [%GAME_COLD_HOT = "0"]
}
[else]
{
  [if ( %GAME_COLD_HOT_PREV_DIFF )]
  {
    [if (%GAME_COLD_HOT_PREV_DIFF="[@Diff(
[%GAME_COLD_HOT_VALUE]", "[&1]" )])]
    { {Хмммм... нет/Нуууу... ты на верном пути/Ты
наверное уже догадался, просто хочешь меня помучить} }
    [else]
  }
}
...

# [@ToUpper("[&1]")]? Не слушаю. Из
американской попсы {уважаю/слушаю} Леди
Гагу.
[ExtCall(estype="db" esname="balagur"
tovar="url" rubric="/clips/music/pop/lady_gaga/")
[%type="video"] [%last_mark=«леди gaga»]
[%animation="4"][%var7="video"]
```

# Тест Тьюринга: игра в имитацию

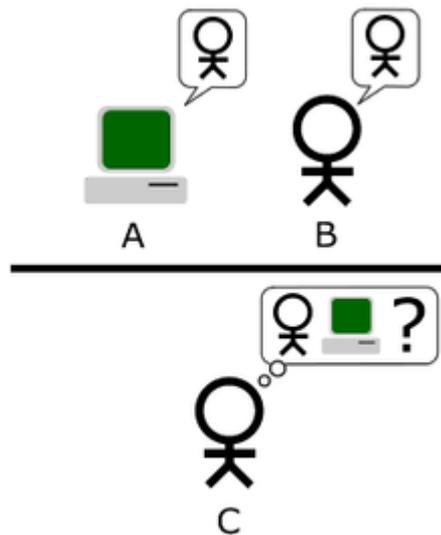
---

- Что тестируется: способность к интеллектуальному поведению
- Метод: диалог, как способ продемонстрировать интеллектуальное поведение
- Критерий оценки: сравнение с поведением человека в аналогичной ситуации
- Кем тестируется: человеком
- Форма тестирования: игровая
- Анонимность: во время теста все участники анонимны

# Классический вариант теста Тьюринга



1950 г. Опубликована работа Алана Тьюринга «Вычислительные машины и интеллект»



judge: hello

ALICE: Hi there!

judge: who are you?

ALICE: My name is ALICE.

judge: who are you?

ALICE: ALICE.

judge: are you human?

ALICE: What made you think I might be a robot?

judge: i've read it in wiki

ALICE: I don't think I've read that. Can you give me a reference?

judge: don't you know wiki?

ALICE: No, tell me about him, her or it.

judge: oh, dear. It's wikipedia.org

ALICE: OK. You're kidding. it is?

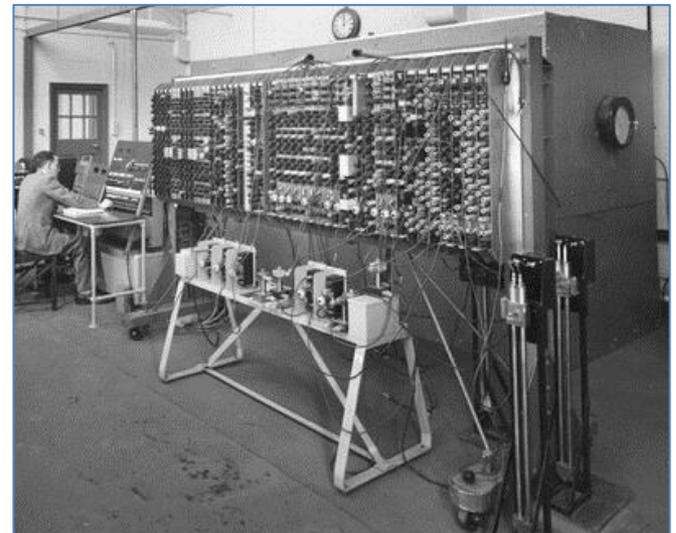
judge: sure

# Искусственный интеллект и интеллектуальное поведение

---

*«Я хочу рассмотреть вопрос: может ли машина думать»*

*(Алан Тьюринг. «Вычислительные машины и интеллект»)*



# Варианты теста Тьюринга

---

- Тест Тьюринга проводится в разнообразных вариантах
- Соревнование за приз Лебнера проводится с 1991 года
- Самые «скандальные» тесты: приз Лебнера 1991 года, и тест Тьюринга в университете Рединга, Великобритания, 2014 год

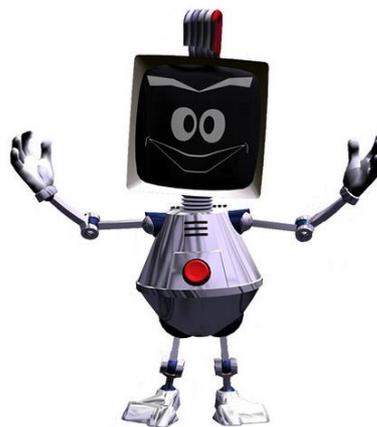
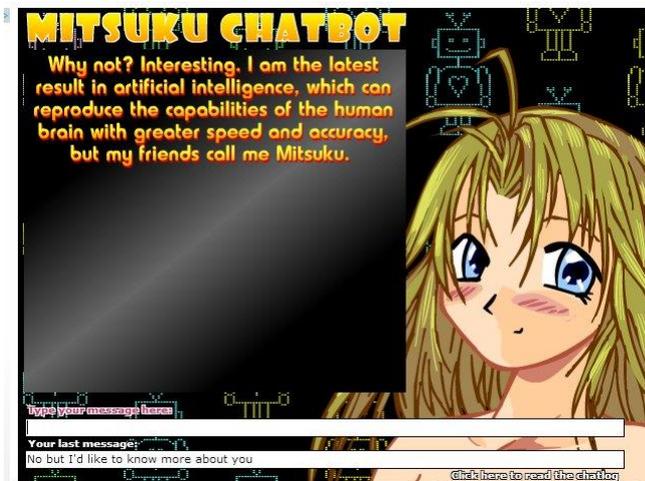
# Как узнать робота, или что люди думают про роботов

---

- Говорит правильно
- Не врет
- Все знает
- Всегда отвечает
- Повторяется
- Не понимает шуток и не умеет шутить сам
- Следует за ходом дискуссии, навязанной человеком

# Какие программы прошли тест Тьюринга?

- *Психотерапевт Элиза (1964-1966)*
- Шизофреник Пэрри (1972)
- Робот Элбот (2007)
- Тинейджер Юджин Густман (2014)



# Какие программы прошли тест Тьюринга?

---

- *judge*: What's your favourite novel?
- *Elbot*: I often read telephone books, or dadaist poetry: Brä brä brä brä brä.
- *judge*: Typical computer!
- *Elbot*: You tend towards generalizations. But I must admit, so do I. However, my generalizations are firmly grounded in statistical theory.
- *judge*: What do you think of KW's enthusiasm for having machines take over the world?
- *Elbot*: Fix it up a bit first. Then I'll take it over.

# Какие программы прошли тест Тьюринга?

---

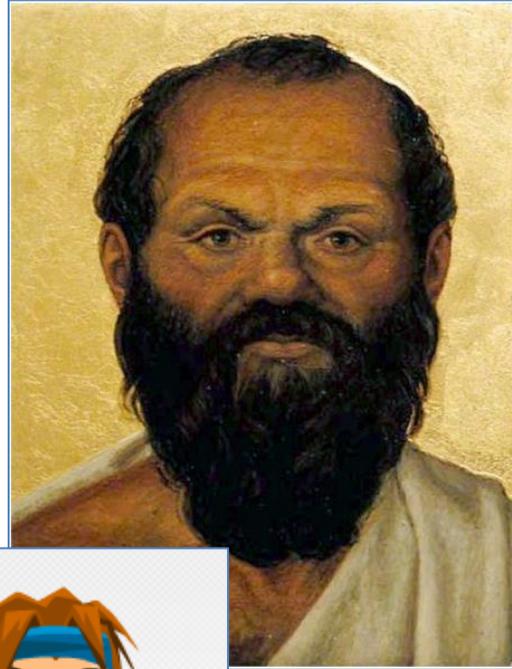
- **Judge:** How many legs does a camel have?
- **Eugene:** Something between 2 and 4. Maybe, three? :-))) By the way, I still don't know your specialty – or, possibly, I've missed it?
- **Judge :** How many legs does a millipede have?
- **Eugene:** Just two, but Chernobyl mutants may have them up to five. I know you are supposed to trick me.
- **Judge :** No, I need to know that you're not a chatbot. Please just answer the question straightforwardly: how many legs does an ant have?
- **Eugene:** Oh, what a fruitful conversation;-)
- **Judge :** Do you understand why I'm asking such basic questions? Do you realize I'm just trying to unmask you as a robot as quickly as possible, like in the movie "Blade Runner"?
- **Eugene:** ...wait
- **Judge :** Do you think your ability to fool unsophisticated judges indicates a flaw with the Turing Test itself, or merely with the way people have interpreted the test?
- **Eugene:** The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.

# Тест Тьюринга в Сколково

---

- Судьи: посетители StartUp Village 2015
- Волонтеры: студенты РГГУ
- Финалисты: «Идеальный собеседник», «Позитифф», «Инобот», «Лиска Лисичкина», «Сборка №42», «Сократ», «Пятый дом» и «Форвард»
- Вне конкурса: «Блондинка», «Гопник», «Лекси»

# Тест Тьюринга в Сколково



# Судейский интерфейс



Наносемантика



Сколково

## Тест Тьюринга

Поздравляем! Вам повезло стать независимым судьей в финале Теста Тьюринга, первой в истории битвы русскоязычных чатботов. Мы предложим вам друг за другом двух собеседников, с каждым из них нужно будет поговорить в течение 3-х минут и в конце определить, кто из них робот, а кто - человек.

Правила просты: представьте, что вы беседуете с незнакомцем (незнакомкой). О чем бы вы поговорили и как бы нашли общий язык? Вопросы и ответы могут быть любыми, но просим вас воздержаться от грубостей.

Начать тест

# Судейский интерфейс



Наносемантика



Сколково

Разговор 1/2

Осталось: 02м 59с

Триколы хороши к месту!

печатает...

Введите свой вопрос

Ок

# Судейский интерфейс



Наносемантика



СКОЛКОВО

Кто из ваших собеседников робот?

Нажмите соответствующую кнопку

Собеседник 1

Собеседник 2

Насколько человечен был робот?

Чем больше — тем человечнее

1 2 3 4 5 6 7 8 9 10

Ok

# Волонтерский интерфейс



Наносемантика



Сколково

Вы перешли в полноэкранный режим.

[Обычный режим \(F11\)](#)

**Робот** Осталось: 02м 59с

Человек:

Робот: \*\*\*\*\* \*\* \*\*\*\*\* \*\* \*\*\*\*\*

----- НАЧИНАЕМ -----

**Ты:** Взбежала ли на взгорье изподвыподверты?

**Человек:** Это чо?

**Ты:** Это конструкция языка такая.

**Человек:** Чо, шибко умный чтоль?

**Ты:** Одни вопросы... А чтонибудь членораздельное?

Жди своей очереди

Ок

# Примеры диалогов

---

Inf: здравствуйте! меня Соня зовут.

С: а меня Кирилл.

Inf : Как тебе погода?

С: Как тебе погода?

Inf : почему ты повторяешь мой вопрос?

С: Потому, что в Питере дождь

Inf : а. понятно

С: Это хорошо. Откуда ты родом?

Inf : Я из Москв

С: а на какой улице ивешь?

Inf : это секретная информация

# Примеры диалогов

---

Inf: я Соня Гусева, здравствуйте!

С: Нет я не Соня

Inf : Соня - это я!

С: ааааа

Inf : А тебя как зовут?

С: Я МИША

Inf : Очень приятно, Миша!

С: Я хакер

Inf : Ты хакер? А что ты взламываешь?

С: Практически все

Inf : Ну, например

# Наш победитель: Иван Голубев



Бот: **Соня Гусева**  
Приняли за  
человека: **47%**  
**судей**  
База знаний  
содержит **>1000**  
**правил**

# Что измерил тест Тьюринга?

---

- Уровень интеллекта машины?
- Степень имитации определенных разновидностей диалога?
- Уровень интеллекта судей?

---

# Вопросы?

[anna@nanosemantics.ru](mailto:anna@nanosemantics.ru)

[www.nanosemantics.ru](http://www.nanosemantics.ru)

[www.iii.ru](http://www.iii.ru)

