

**О. Н. Максимова** – студент кафедры вычислительных систем и сетей

**А. В. Аксёнов** (ст. преп.) – научный руководитель

## **ИСПОЛЬЗОВАНИЕ ВАРИАЦИИ ТЕСТА ТЬЮРИНГА ДЛЯ ВЫЯВЛЕНИЯ «МАШИННЫХ» ПАТТЕРНОВ ЧЕЛОВЕЧЕСКОГО ПОВЕДЕНИЯ**

Одним из направлений развития информационных технологий является создание программ, в той или иной мере имитирующих речевое поведение человека. Подобные программы разного уровня сложности имеют достаточно широкое, хотя и не всегда «полезное» применение, но могут быть использованы для упрощения человеко-машинного взаимодействия. Например, в наши дни уже существует опыт использования программ в качестве продавцов-консультантов в интернет-магазинах [1].

При создании любого виртуального собеседника (чат-бота) решаются две задачи: во-первых, программу необходимо обучить «понимать» сообщения на естественном языке, то есть выделять ключевые фразы, анализировать контекст и предыдущие сообщения. И, во-вторых — формировать на основе полученной информации понятные человеку ответы в форме, максимально близкой к разговорной речи.

Взаимодействие человека с подобной программой будет тем более комфортным, чем ближе её поведение к «человеческому». Но однозначно определить, какое поведение воспринимается как «человеческое», а какое — нет, на самом деле не просто.

В рамках данного исследования была предпринята попытка найти ответы на два вопроса:

1. По каким признакам человек определяет, что общается не с другим человеком, а с программой? (Выявление наиболее распространённых представлений о «человеческих» и «машинных» паттернах поведения).

2. Существуют ли такие сценарии человеческого поведения, которые при определённых условиях будут восприняты собеседником как «машинные»? (Выявление «машинных» паттернов человеческого поведения).

Первая часть исследования включала в себя онлайн-опрос, участникам которого было предложено ответить на ряд вопросов, связанных с программами-собеседниками. Всего в опросе приняло участие более 100 человек, преимущественно — люди в возрасте от 18 до 25 лет. При этом 40% опрошенных указали в качестве сферы своей деятельности сферу ИТ.

Согласно результатам опроса, практически все респонденты знают о существовании виртуальных собеседников (92%), при этом большая часть таких программ кажется им неубедительной: такого мнения придерживаются 57% опрошенных, в то время как за убедительность большинства чат-ботов проголосовало только 6,2%. Остальные участники опроса не смогли сделать выбор либо посчитали вероятность принятия программы за человека ситуативной.

Помимо этого, респондентам предлагалось представить себя участниками теста Тьюринга — эмпирического теста, идея которого была предложена Аланом Тьюрингом в статье «Вычислительные машины и разум», опубликованной в 1950 году. Исходной целью теста был поиск ответа на вопрос, может ли машина мыслить. Позднее от этой идеи отказались, и в наши дни тест рассматривается исключительно как способ проверки «убедительности» чат-ботов. Его участники должны, обмениваясь текстовыми сообщениями с невидимыми им собеседниками, определить, кто из них является человеком, а кто — имитирующей человеческое поведение программой. Тест Тьюринга считается пройденным, если программа сумеет обмануть 30% своих собеседников.

Среди стратегий, предложенных респондентам для выявления программ-собеседников, предпочтительными оказались анализ манеры письменной речи собеседника и его поведения.

При этом к человеческим паттернам поведения были отнесены:

- Чувство юмора;
- Наличие ошибок и опечаток в сообщениях при правильном построении фраз;
- Наличие собственного мнения по большинству вопросов;
- Способность подробно рассказать о себе.

Также достаточно популярным среди респондентов оказался критерий восприятия собеседника как личности, что можно понимать как наличие у собеседника биографии, устойчивой системы взглядов и индивидуальных особенностей речи.

В качестве признаков машинного поведения были выделены такие, как наличие повторяющихся либо совершенно различных ответов на похожие вопросы, отличие манеры речи собеседника от разговорной и «странность» его собственных вопросов. Стоит отметить, что эти признаки относятся по большей части к слабым программам, имеющим небольшую базу знаний и не способным анализировать предыдущие сообщения собеседника. Но так же в некоторых случаях подобное поведение может демонстрировать и человек.

Данное предположение косвенно подтверждается результатами опроса: 57% респондентов соглашались с тем, что им станет сложнее определить, является ли собеседник программой или человеком, если предварительно станет известно, что среди собеседников-людей не все в должной мере владеют языком, на котором происходит общение.

Таким образом, можно сделать вывод, что в ситуации, когда необходимо отличить собеседника-человека от собеседника-программы, восприниматься как полностью «человеческие» будут лишь некоторые из всех возможных сценариев человеческого поведения. При этом значительное внимание будет уделяться владению разговорным языком.

Вторая часть исследования является экспериментальной и направлена на выявление сценариев человеческого поведения, которые в определённых условиях могут быть интерпретированы как «машинные». Для этого была предложена специальная вариация теста Тьюринга, единственным существенным отличием которой от общепринятой является отсутствие среди испытуемых реальных ботов.

В эксперименте принимало участие две группы: «судьи» и «участники», набранные случайным образом из числа студентов ГУАП. «Судьям» было сообщено, что они участвуют в тесте Тьюринга, в то время как «участники» считали, что помогают испытывать программу-мессенджер. Взаимодействие осуществлялось в течение 5 минут строго по одной реплике — «судья» не должен был отправлять следующий вопрос, пока не получил ответ на предыдущий. При этом вся переписка между «судьями» и «участниками» автоматически сохранялась.

Наиболее популярными стратегиями обнаружения ботов стали выявление особенностей личности собеседника («судьи» задавали участникам вопросы, касающиеся их настроения, произошедших в их жизни событий, каких-либо предпочтений и т.д.) и проверка знаний собеседника (несложные вопросы из химии, математики, астрономии).

Также использовались:

- Анализ текста сообщений собеседника (наличие опечаток, смайликов и т.п.).
- Анализ скорости ответа собеседника.
- Проверка способности собеседника распознавать и производить вычисления («Напиши мне, сколько будет два в степени 0, три в степени ноль, 4 в степени один. Ответ напиши по порядку»).

При этом определить, что все собеседники являются людьми, смог только испытуемый, выбравший последнюю стратегию. Все остальные «судьи» принимали некоторых своих собеседников за ботов. Кроме того, нашлись такие «участники», которые приняли за ботов «судей», несмотря на то, что им не был заранее предложен такой вариант.

Анализируя характеристики, данные собеседниками друг другу, и тексты их переписок, удалось выявить ряд паттернов человеческого поведения, воспринимаемых как «машинные». В ходе эксперимента за ботов были приняты те участники, в диалогах с которыми выполнялось

хотя бы одно из следующих условий:

- Реакция собеседника на задаваемые вопросы не соответствует ожидаемой. Например, шуточные ответы «участников» на вопросы «судей» воспринимались последними как непонимание сути этих вопросов.

- Собеседник отвечает на вопросы слишком медленно или вообще не отвечает на них (воспринимается как «зависание» программы).

- Реплики собеседника являются «шаблонными», предсказуемыми в рамках темы диалога.

- Собеседник реагирует только на сообщения, касающиеся ответов на его вопросы. Стремление «судей» за короткий период времени получить как можно больше информации согласно выбранной стратегии было воспринято частью собеседников как проявление «машинного» поведения.

Из всего вышесказанного можно сделать вывод о существовании особых «машинных» сценариев человеческого поведения, влияющих на восприятие человеком его собеседника. Выявление таких сценариев, как и анализ закономерностей человеческого восприятия, может помочь в создании программ, более правдоподобно имитирующих человеческое поведение.

### **Библиографический список**

1. *Насакин Р.* Эволюция болтунов. // Компьютерра Online. — Дата публикации: 18.04.2007.  
Режим доступа: <http://old.computerra.ru/focus/315260/>
2. *Тьюринг А. М.* Могут ли машины мыслить? // В сб.: Информационное общество. — М.: АСТ, 2004. — С. 221-284.
3. *Тьюринг А. М.* Вычислительные машины и разум. // В сб.: Хофштадер Д., Деннет Д. Глаз разума. — Самара: Бахрах-М, 2003. — С. 47-59.
4. *Warwick K., Shah H.* Human misidentification in Turing tests. // Journal of Experimental & Theoretical Artificial Intelligenc. 2014, № 27 (3). PP. 123-135.