НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2024 Том 24 № 2 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2024 Vol. 24 No 2 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

COMPUTER SCIENCE

doi: 10.17586/2226-1494-2024-24-2-214-221
УДК 004.89

# RuPersonaChat: корпус диалогов для персонификации разговорных агентов

**Кирилл Сергеевич Апанасович[1], Олеся Владимировна Махныткина[2]✉, Владимир Иосифович Кабаров[3], Ольга Петровна Далевская[4]**

[1,2,3,4] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] apanasovich.k@yandex.ru, https://orcid.org/0000-0001-7966-3488
[2] makhnytkina@itmo.ru✉, https://orcid.org/0000-0002-8992-9654
[3] vikabarov@itmo.ru, https://orcid.org/0000-0001-6300-9473
[4] opdalevskaia@itmo.ru, https://orcid.org/0000-0001-5246-9212

**Аннотация**

**Введение.** Одним из способов повышения качества разговорных агентов является персонификация. Персонификация улучшает качество взаимодействия пользователя с разговорным агентом и повышает удовлетворенность пользователей за счет повышения консистентности и специфичности ответов. Диалог с агентом становится более последовательным, минимизируется противоречивость ответов, которые оказываются более конкретными и интересными. Для обучения и тестирования персонифицированных разговорных агентов требуются специфичные наборы данных, содержащие факты о персоне и тексты диалогов персон, в репликах которых используются факты о персонах. Существует несколько наборов на английском и китайском языках, содержащие в описании персоны в среднем пять фактов. Диалоги в наборах данных составлены пользователями краудсорсинга, которые многократно имитировали различные персоны. **Метод.** В данной работе предложена методика сбора оригинального корпуса данных, содержащего расширенный набор фактов о персоне и естественные диалоги между персонами. Новый корпус данных RuPersonaChat основан на трех различных сценариях записи: интервью, короткая беседа, длинная беседа. Впервые собран корпус данных для персонификации разговорных агентов, включающий естественные диалоги и расширенное описание персоны. Предложена дополнительная разметка набора данных, которая ставит в соответствие реплики персоны и факты о персоне, на основе которых она была сформулирована. **Основные результаты.** Разработана методика сбора оригинального корпуса тестовых данных, позволяющего осуществлять тестирование языковых моделей для решения большего количества задач в рамках разработки персонифицированного разговорного агента. Собранный набор данных включает 139 диалогов и 2608 реплик. Корпус использован для тестирования моделей генерации ответов и вопросов. Наилучшие результаты получены с использованием модели Gpt3-large (перплексия равна 15,7). **Обсуждение.** Собранный корпус данных RuPersonaChat может быть использован для тестирования персонифицированных разговорных агентов на возможность рассказать о себе собеседнику, ведения диалога с собеседником и использования фактической речи, учета длинного контекста при ведении диалога с пользователем.

**Ключевые слова**

методика сбора данных, диалоговые данные, разговорные агенты, персонификация, генерация вопросов и ответов

**Благодарности**

Исследование выполнено за счет гранта Российского научного фонда (№ 22-11-00128, https://www.rscf.ru/project/22-11-00128/).

**Ссылка для цитирования:** Апанасович К.С., Махныткина О.В., Кабаров В.И., Далевская О.П. RuPersonaChat: корпус диалогов для персонификации разговорных агентов // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 2. С. 214–221. doi: 10.17586/2226-1494-2024-24-2-214-221

# RuPersonaChat: a dialog corpus for personalizing conversational agents

**Kirill S. Apanasovich[1], Olesia V. Makhnytkina[2]✉, Vladimir I. Kabarov[3], Olga P. Dalevskaya[4]**

[1,2,3,4] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] apanasovich.k@yandex.ru, https://orcid.org/0000-0001-7966-3488
[2] makhnytkina@itmo.ru✉, https://orcid.org/0000-0002-8992-9654
[3] vikabarov@itmo.ru, https://orcid.org/0000-0001-6300-9473
[4] opdalevskaia@itmo.ru, https://orcid.org/0000-0001-5246-9212

**Abstract**

Personalization is one of the keyways to improve the performance of conversational agents. It improves the quality of user interaction with a conversational agent and increases user satisfaction by increasing the consistency and specificity of responses. The dialogue with the agent becomes more consistent, the inconsistency of responses is reduced, and the responses become more specific and interesting. Training and testing personalized conversational agents requires specific datasets containing facts about a persona and texts of persona's dialogues where replicas use those facts. There are several datasets in English and Chinese containing an average of five facts about a persona where the dialogues are composed by crowdsourcing users who repeatedly imitate different personas. This paper proposes a methodology for collecting an original dataset containing an extended set of facts about a persona and natural dialogues between personas. The new RuPersonaChat dataset is based on three different recording scenarios: an interview, a short conversation, and a long conversation. This is the first dataset for dialogue agent personalization collected which includes both natural dialogues and extended persona's descriptions. Additionally, in the dataset, the persona's replicas are annotated with the facts about the persona from which they are generated. The methodology for collecting an original corpus of test data proposed in this paper allows for testing language models for various tasks within the framework of personalized dialogue agent development. The collected dataset includes 139 dialogues and 2608 replicas. This dataset was used to test answer and question generation models and the best results were obtained using the Gpt3-large model (perplexity is equal to 15.7). The dataset can be used to test the personalized dialogue agents' ability to talk about themselves to the interlocutor, to communicate with the interlocutor utilizing phatic speech and taking into account the extended context when communicating with the user.

**Keywords**

data collection methodology, dialog data, conversational agents, personalization, question and answer generation

**Acknowledgements**

This study was funded by a grant from the Russian Science Foundation (22-11-00128, https://www.rscf.ru/project/22-11-00128/).

**For citation:** Apanasovich K.S., Makhnytkina O.V., Kabarov V.I., Dalevskaya O.P. RuPersonaChat: a dialog corpus for personalizing conversational agents. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 2, pp. 214–221 (in Russian). doi: 10.17586/2226-1494-2024-24-2-214-221

## Введение

В последние несколько лет получили значительное развитие системы разговорных агентов. Одним из подходов, позволяющим сделать человеко-компьютерное общение более реалистичным и естественным, является персонификация разговорных агентов [1, 2]. Представление персоны разговорного агента включает описание личностных и профессиональных характеристик, предпочтений и т. д. Персонификация позволяет моделям решить проблему обычных разговорных агентов, которые регулярно обладают непоследовательной индивидуальностью и не обладают явной долговременной памятью, что приводит к различным ответам на один и тот же вопрос и склонности к неконкретным ответам типа «я не знаю». Персонификация в интерактивных системах является важным фактором, улучшающим качество взаимодействия с системой и повышающим удовлетворенность пользователей за счет повышения согласованности и конкретности ответов. Персонификация делает диалог с системой более связным, минимизирует противоречивость ответов, делает их более конкретными и интересными.

Для обучения и тестирования персонифицированных разговорных агентов необходимы наборы данных, содержащие информацию о персоне и примеры диалогов. В настоящее время существует несколько наборов данных с диалогами, в которых в том или ином виде содержится информация о персоне, или другая информация, которая позволяет моделям, обученным на таких данных, вести с пользователем диалог, который наиболее приближен к естественному диалогу между двумя людьми.

Одним из самых используемых наборов данных является PersonaChat [3] и его расширенная версия ConvAI2 [4], в которых к каждому собеседнику в диалогах ставится в соответствие информация о его персоне. В данном случае персона представлена в виде пяти коротких предложений, например «I like to ski» или «I have a computer science degree». Описания персон, как и диалоги, были собраны методом краудсорсинга. Предполагается, что с помощью присвоения персоны, языковая модель сможет генерировать реплики, которые будут в меньшей степени противоречить друг другу.

Другим возможным способом персонификации разговорного агента, может служить его обучение эмоциональному диалогу. Так, например, в наборе данных Empathetic Dialogues [5] каждому диалогу ставится в соответствие эмоция, которая отражает настроение одного из собеседников. Таким образом, у двух собеседников появляется возможность вести эмоциональный диалог.

Между тем, если обучать модель нескольким навыкам, то она может показывать хорошие результаты в диалогах, используя эти навыки отдельно друг от друга, но у нее все еще могут быть проблемы при их смешивании в одном диалоге, где в зависимости от ситуации необходимо проявлять эмпатию или рассказать что-то о своей персоне. В наборе данных Blended Skill Talk [6] собраны диалоги, где собеседники могут демонстрировать различные навыки: показать свои знания, чуткость, личный опыт или персону.

Языковые модели ограничены тем, что могут работать с текстовыми последовательностями относительно небольшой длины, из-за чего модель при ведении длинного диалога забывает, о чем велась речь. В наборах данных Multi-Session Chat [7] и DuLeMon [8] данная проблема решена путем обучения модели запоминать в ходе ведения диалога новые факты о персонах своего собеседника и себе самой.

Существенной проблемой в развитии разговорных агентов является то, что большинство собранных наборов данных содержат диалоги только на английском и китайском языках, в то время как на русском языке существует только один набор данных — Toloka Persona Chat Rus[1], который по своей структуре схож с наборами PersonaChat и ConvAI2. При этом он не подразумевает возможности вести эмоциональный диалог, и не позволяет обучить модель получать из реплик новые факты о персонах.

Все вышеперечисленные наборы данных имеют достаточно краткое описание персоны (пять фактов), при сборе данных один и тот же диктор имитирует речь нескольких персон, диалоги представляют собой короткие беседы. В настоящей работе предложена методика сбора корпуса данных RuPersonaChat, отличающегося от существующих большей естественностью за счет использования реальных персон дикторов и разнообразием тем диалогов. Это позволит использовать собранный набор данных для оценки качества языковых моделей на нескольких задачах: вести короткие и длинные диалоги, уметь использовать факты о своей персоне и понимать факты о персоне собеседника с учетом, того, что персоны будут содержать значительно большее количество фактов, чем в наборах данных PersonaChat и ConvAI2.

### Методика сбора корпуса данных

Общая схема сбора корпуса данных RuPersonaChat показана на рис. 1. Одной из задач сбора данного корпуса является тестирование того, насколько хорошо разговорный агент способен понимать заданную ему персону и персону собеседника и учитывать их при ведении диалога. Проведение подобных экспериментов было невозможно на наборах данных, содержавших пять фактов о персоне и короткие диалоги, в которых для генерации ответа можно было использовать только эти факты. При генерации ответа пользователю разговорный агент осуществляет выбор факта о персоне, в зависимости от контекста диалога, приближенного к естественному. Для оценки способности разговорного агента необходимо иметь существенно большее количество фактов о персоне.

В рамках разрабатываемой методики сбора корпуса данных были выделены основные тематики блоков вопросов: демографическая информация, образование, профессия и опыт работы, интересы, навыки, жизненная позиция, жизненные ситуации. Для более детального описания персоны, чем в наборах данных PersonaChat и ConvAI2, была составлена анкета из 94 вопросов по сформулированным тематикам.

Методика сбора корпуса включает три сценария ведения диалога, два из них (короткие диалоги-интервью и длинные диалоги-беседы) являются оригинальными и отсутствуют в существующих наборах данных.

1. Короткие диалоги-интервью: один из собеседников задает вопросы на одну тему другому собеседнику. Цель такого сценария — протестировать, насколько хорошо разрабатываемая модель способна рассказать о себе собеседнику.
2. Короткие диалоги-беседы, которые более похожи на обычное общение двух человек. Здесь оба собеседника обсуждают одну конкретную тему, задавая друг другу вопросы по ней. В данном сценарии тестируются навыки языковой модели в ведении диалога с собеседником и использовании фатической речи.
3. Длинные диалоги-беседы, схожие с короткими беседами, но отличающиеся длительностью записи, которая в данном случае должна превышать 20 мин. Такие беседы необходимы, чтобы проверить, как хорошо модель работает с длинным контекстом при ведении диалога с собеседником.

Запись диалогов осуществлена двумя способами: с использованием программы для организации видеоконференций Zoom и мессенджера Telegram.

При первом способе выполнена запись речи собеседников, при этом речь каждого собеседника записана как отдельным файлом, так и в один общий, что в дальнейшем использовано для задачи диаризации речи собеседников. Запись осуществлена в тихой комнате, без посторонних шумов и речи других людей. Далее полученные записи автоматически переведены в текстовый формат с помощью модели автоматического распознавания речи. Для выделения произнесений и их транскрибирования использована автоматическая система распознавания речи и фреймворк NeMo [9]. Система распознавания речи основана на предобученной нейросетевой модели архитектуры Conformer [10].

При использовании второго способа сбор данных совершен путем ведения текстового диалога между двумя пользователями. Перед записью диалога собеседникам предлагалось выбрать тему, на какую они будут вести диалог. Список таких тем для обсуждения включал темы об интересах, ситуациях из повседневной жизни (погода, технологии и др.), образовании, работе. Пример диалога представлен на рис. 2.

Для включения в диалоги эмоциональной речи были сформулированы дополнительные темы, через

[1] [Электронный ресурс]. Режим доступа: https://toloka.ai/datasets/ (дата обращения: 26.02.2024).

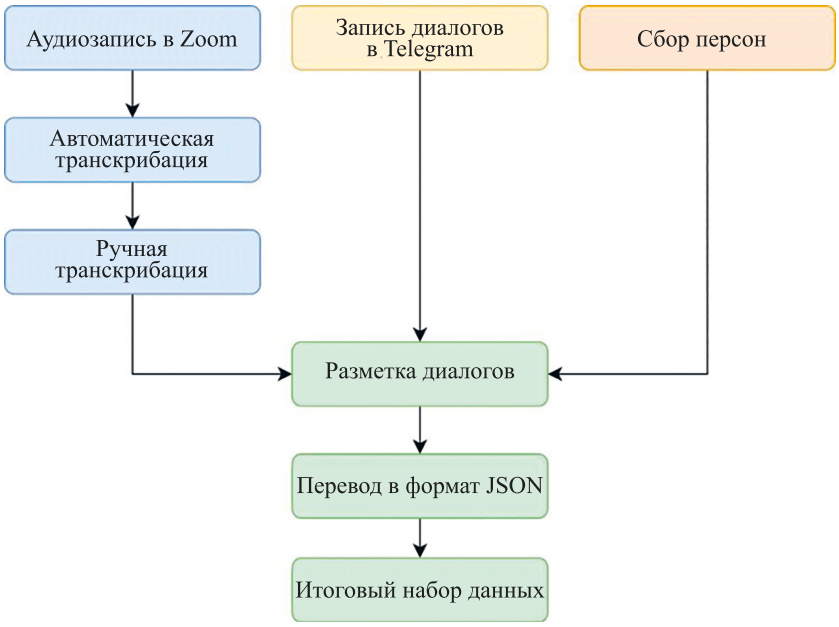


*Рис. 1*. Схема сбора корпуса данных RuPersonaChat
*Fig. 1*. A scheme for collecting the corpus

обсуждение которых можно вызвать естественные эмоции у собеседников. Сами эмоции определялись в соответствии с моделью Экмана (нейтральная, радость, злость, отвращение, страх, грусть, удивление). Пример диалога представлен на рис. 3. Репликам ставились в соответствие вызванные эмоции, для каждой реплики определялось, какой факт или факты в ней упомянуты. В процессе сбора диалогов выяснилось, что в репликах часто использовались такие факты, которые не были предусмотрены анкетой ввиду того, что многие темы, на которые велись диалоги, были очень специфичны.

### Описание корпуса данных

В настоящий момент корпус данных RuPersonaChat содержит 63 диалога, записанных с использованием

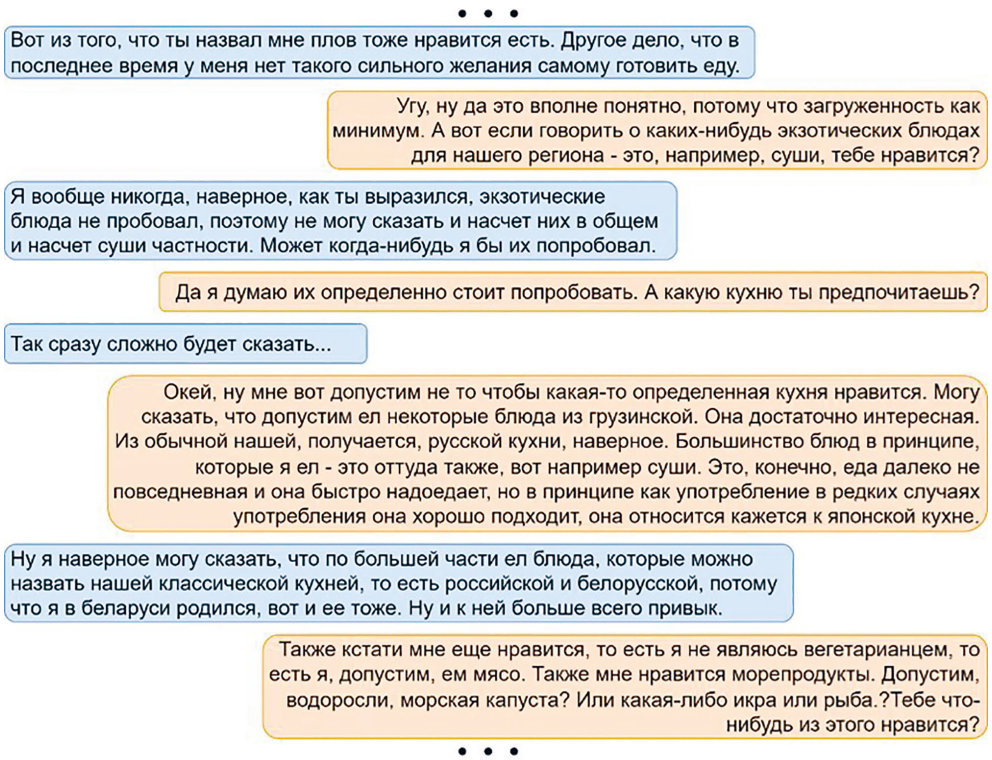


*Рис. 2*. Фрагмент из диалога на тему предпочтений в еде
*Fig. 2*. Fragment of a dialog on food preferences

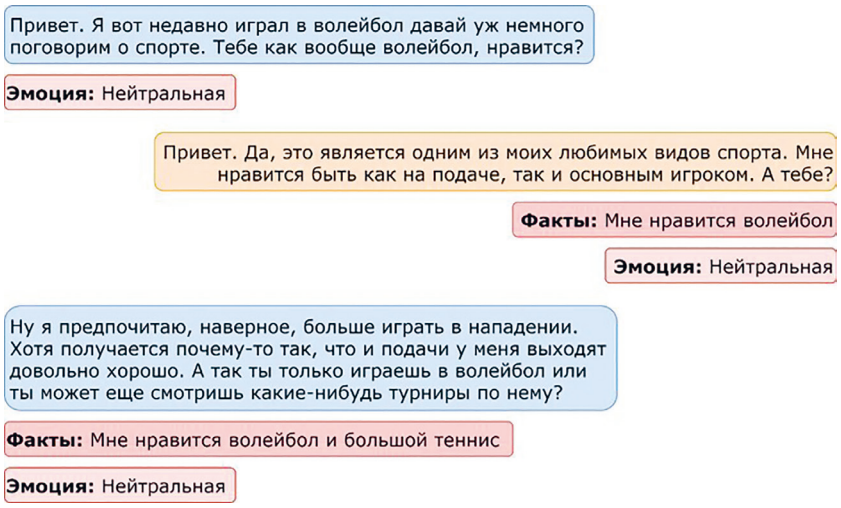


*Рис. 3*. Фрагмент из диалога с разметкой по эмоциям и упоминаемым фактам
*Fig. 3*. A fragment from a dialog with markup by emotions and facts mentioned

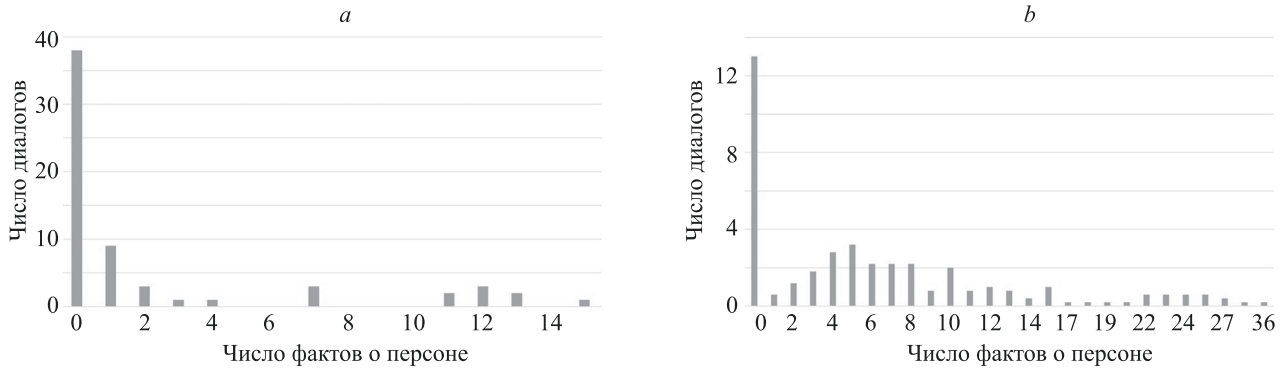


*Рис. 4*. Распределение встречаемости числа фактов о персоне в диалогах, собранных в Zoom (*a*) и в Telegram (*b*)
*Fig. 4*. Distribution of facts occurrence in dialogs collected in Zoom (*a*) and in Telegram (*b*)

программ для проведения видеоконференций Zoom, общая длительность диалогов в аудиоформате составляет 10 ч. Полученные диалоги включают в себя 1396 реплик. Каждая реплика в среднем состоит из 36 слов. При этом один диалог мог включать до 15 фактов из анкеты для одного собеседника (рис. 4, *a*).

В сборе диалогов с использованием мессенджера Telegram (рис. 4, *b*) участвовало пять собеседников, всего было собрано 76 диалогов, которые суммарно содержат 1212 реплик. В среднем длина одной реплики составила 15,4 слова.

Наиболее часто встречались факты, которые затрагивали темы интересов собеседников, их работы и образования. На рис. 5, *a* показана гистограмма с наиболее часто задаваемыми вопросами при сборе в Zoom, а на рис. 5, *b* — через Telegram.

Эмоциональная речь имитировалась только при сборе диалогов в Zoom. В диалогах чаще всего использовалась нейтральная эмоция — в 1358 репликах. При этом в 17 диалогах хотя бы в одной реплике встречалась другая эмоция.

В табл. 1 представлено сравнение собранного корпуса данных RuPersonaChat с другими наборами данных с диалогами, в которых каждому собеседнику поставлена в соответствие его персона.

Размер полученного корпуса данных RuPersonaChat не позволяет использовать его для полноценного обучения, но полученных данных достаточно для проведения тестирования уже обученных моделей.

### Тестирование моделей генерации ответов и вопросов

Собранный корпус данных использован для решения задачи генерации вопросов и ответов с учетом характеристики персоны, речь которой будет имитировать разговорный агент на основе генеративных моделей. Для проведения экспериментов рассмотрены модели двух архитектур: Encoder-Decoder [11] модели T5-base и T5-large с 222 и 737 млн параметров [12] и Decoder-only модели Gpt3-small, Gpt3-medium и Gpt3-large с 125, 355 и 760 млн параметров [13]. Все модели были дообучены на наборе данных Toloka Persona Chat Rus. Для сравнения результатов работы моделей использованы метрики: перплексия (perplexity, PPL) [14] и Bilingual Evaluation Understudy (BLEU)-1 [15]:

— PPL — безразмерная величина, которая является обратной к величине средней вероятности, приписываемой каждому слову тестовой строки. Чем

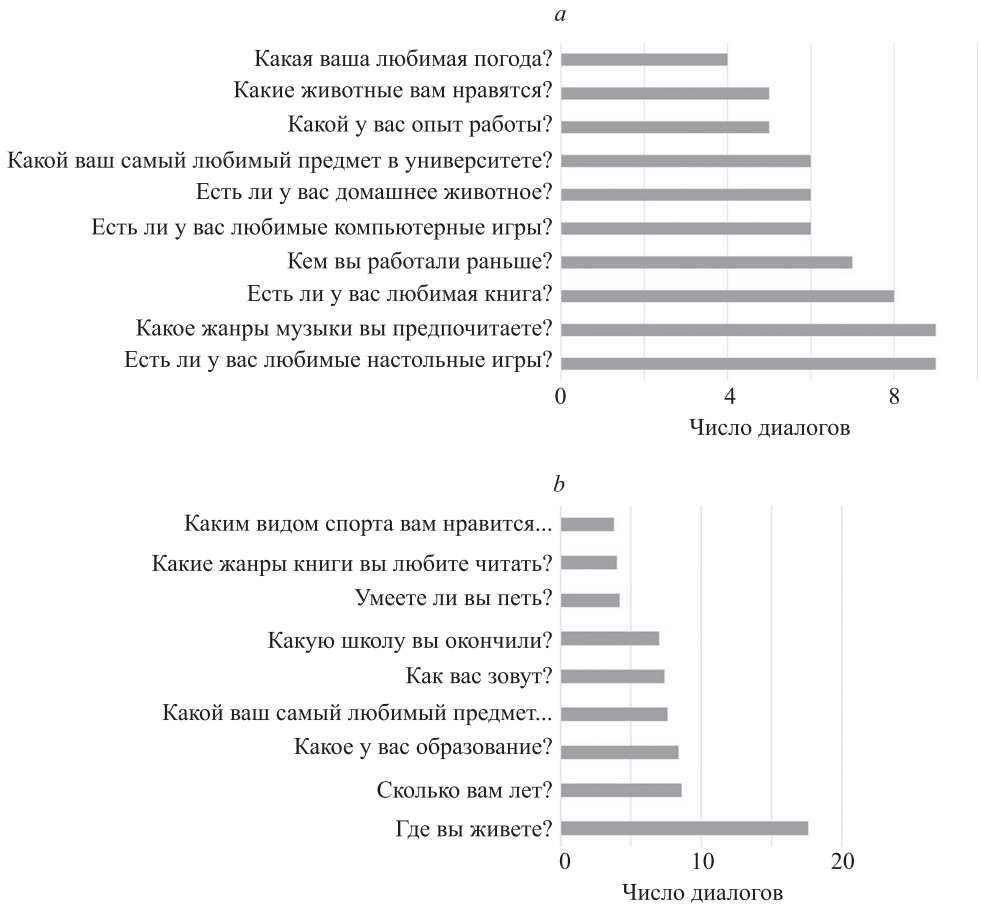


*Рис. 5*. Гистограмма с наиболее часто встречающимися вопросами при сборе диалогов в Zoom (*a*) и Telegram (*b*)
*Fig. 5*. Histogram with the most frequent questions in dialog collection in Zoom (*a*) and Telegram (*b*)

*Таблица 1*. Статистика по наборам данных
*Table 1.* Statistics on datasets

| Название набора данных | Язык набора данных | Количество фактов о персоне | Число | | Среднее количество слов | |
|---|---|---|---|---|---|---|
| | | | диалогов | реплик | в реплике | в диалоге |
| PersonaChat | Английский | 5 | 10 907 | 162 064 | 11,9 | 177 |
| ConvAI2 | Английский | 5 | 19 893 | 145 873 | 13,7 | 100 |
| DuLeMon | Китайский | 5 | 27 501 | 400 472 | 19,7 | 287 |
| Empathetic Dialogues | Английский | 1 | 24 850 | 64 609 | 15,3 | 40 |
| Toloka Persona Chat rus | Русский | 5 | 10 013 | 168 783 | 6,1 | 103 |
| RuPersonaChat | Русский | 94 | 139 | 2608 | 26,4 | 495 |

*Таблица 2*. Результаты экспериментов по генерации ответных реплик
*Table 2.* Results of experiments on answer generation

| Модель | Toloka Persona Chat Rus | RuPersonaChat | |
|---|---|---|---|
| | PPL ↓ | PPL ↓ | BLEU-1 ↑ |
| Gpt3-small | 13,6 | 17,9 | 0,191 |
| Gpt3-medium | 10,6 | 16,4 | 0,207 |
| Gpt3-large | 9,5 | **15,7** | 0,218 |
| T5-base | 16,8 | 20,7 | 0,201 |
| T5-large | 13,6 | 16,4 | **0,229** |

меньше значение данной метрики, тем качество работы модели считается выше (↓);
— BLEU принимает значения от 0 до 1, рассчитывается как доля униграмм, совпавших с эталоном. Чем больше значение данной метрики, тем качество работы модели считается выше (↑).

Полученные результаты представлены в табл. 2. Выполнено сравнение результатов собранного корпуса данных RuPersonaChat и набора данных Toloka Persona Chat Rus.

Из табл. 2 видно, что для собранного корпуса данных RuPersonaChat получено более высокое значение PPL, что может быть объяснено тем, что диалоги в нем значительно длиннее, чем в корпусе Toloka Persona Chat Rus (среднее количество слов в диалоге в наборе данных Toloka Persona Chat Rus — 103, а в RuPersonaChat — 495) (табл. 1), и не всегда вся история диалога может быть помещена в модель.

## Заключение

В работе представлена методика сбора корпуса текстовых данных, отличающего от существующих большей естественностью, и позволяющего тестировать модели для большого класса задач. Впервые предложено использовать в качестве описаний персон не искусственно созданные профили, а факты о реальных персонах. Анкеты, содержащие факты о персоне, заполнялись участниками сбора корпуса и далее записывались диалоги между этими людьми, в отличие от существующих наборов данных, при сборе которых одному и тому же человеку предлагалось многократно записывать диалоги имитируя разных персон. При использовании искусственных описаний персон достаточно часто факты повторялись в репликах диалогов дословно, а не были встроены в речь как при естественных диалогах. Такой подход позволил осуществить запись диалогов по различным сценариям и существенно расширить описание персон с пяти фактов до 94.

Предложена оригинальная разметка корпуса данных, включающая указание факта, который использован при ответе, и эмоциональную окраску ответа. Это существенно расширило потенциальный набор задач, для которых он может быть использован: извлечение информации о персоне из текстовых данных, генерация эмоциональных реплик, распознавание эмоционального состояния человека в диалоге.

Проведение экспериментов с собранным корпусом данных RuPersonaChat для задачи генерации ответов и вопросов с использованием генеративных моделей показало, что набор данных имеет более сложную структуру и для решения задачи более качественной человеко-машинной коммуникации необходима модификация существующих и разработка новых подходов к персонификации разговорных агентов.

## Литература


1. Posokhov P., Apanasovich K., Matveeva A., Makhnytkina O., Matveev A. Personalizing dialogue agents for Russian: retrieve and refine // Proc. of the 31st Conference of Open Innovations Association (FRUCT). 2022. P. 245–252. https://doi.org/10.23919/fruct54823.2022.9770895
2. Matveev Y., Makhnytkina O., Posokhov P., Matveev A., Skrylnikov S. Personalizing hybrid-based dialogue agents // Mathematics. 2022. V. 10. N 24. P. 4657. https://doi.org/10.3390/math10244657
3. Zhang S., Dinan E., Urbanek J., Szlam A., Kiela D., Weston J. Personalizing Dialogue Agents: I have a dog, do you have pets too? // Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. V. 1. 2018. P. 2204–2213. https://doi.org/10.18653/v1/p18-1205
4. Dinan E., Logacheva V., Malykh V., Miller A., Shuster K., Urbanek J., Kiela D., Szlam A., Serban I., Lowe R., Prabhumoye S., Black A.W., Rudnicky A., Williams J., Pineau J., Burtsev M., Weston J. The second conversational intelligence challenge (ConvAI2) // The NeurIPS'18 Competition. Springer, Cham, 2020. P. 187–208. https://doi.org/10.1007/978-3-030-29135-8_7
5. Rashkin H., Smith E.M., Li M., Boureau Y-L. Towards empathetic open-domain conversation models: A new benchmark and dataset // Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. P. 5370–5381. https://doi.org/10.18653/v1/p19-1534
6. Smith E.M., Williamson M., Shuster K., Weston J., Boureau Y-L. Can you put it all together: evaluating conversational agents' ability to blend skills // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 2021–2030. https://doi.org/10.18653/v1/2020.acl-main.183
7. Xu J., Szlam A., Weston J. Beyond goldfish memory: Long-term open-domain conversation // Proc. of the 60th Annual Meeting of the Association for Computational Linguistics. V. 1. 2022. P. 5180–5197. https://doi.org/10.18653/v1/2022.acl-long.356
8. Xu X., Gou Z., Wu W., Niu Z-Y., Wu H., Wang H., Wang S. Long time no see! open-domain conversation with long-term persona memory // Findings of the Association for Computational Linguistics:


## References


1. Posokhov P., Apanasovich K., Matveeva A., Makhnytkina O., Matveev A. Personalizing dialogue agents for Russian: retrieve and refine. *Proc. of the 31st Conference of Open Innovations Association (FRUCT)*, 2022, pp. 245–252. https://doi.org/10.23919/fruct54823.2022.9770895
2. Matveev Y., Makhnytkina O., Posokhov P., Matveev A., Skrylnikov S. Personalizing hybrid-based dialogue agents. *Mathematics*, 2022, vol. 10, no. 24, pp. 4657. https://doi.org/10.3390/math10244657
3. Zhang S., Dinan E., Urbanek J., Szlam A., Kiela D., Weston J. Personalizing Dialogue Agents: I have a dog, do you have pets too? *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*. *V. 1*, 2018, pp. 2204–2213. https://doi.org/10.18653/v1/p18-1205
4. Dinan E., Logacheva V., Malykh V., Miller A., Shuster K., Urbanek J., Kiela D., Szlam A., Serban I., Lowe R., Prabhumoye S., Black A.W., Rudnicky A., Williams J., Pineau J., Burtsev M., Weston J. The second conversational intelligence challenge (ConvAI2). *The NeurIPS'18 Competition*. Springer, Cham, 2020, pp. 187–208. https://doi.org/10.1007/978-3-030-29135-8_7
5. Rashkin H., Smith E.M., Li M., Boureau Y-L. Towards empathetic open-domain conversation models: A new benchmark and dataset. *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5370–5381. https://doi.org/10.18653/v1/p19-1534
6. Smith E.M., Williamson M., Shuster K., Weston J., Boureau Y-L. Can you put it all together: evaluating conversational agents' ability to blend skills. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 2021–2030. https://doi.org/10.18653/v1/2020.acl-main.183
7. Xu J., Szlam A., Weston J. Beyond goldfish memory: Long-term open-domain conversation. *Proc. of the 60th Annual Meeting of the Association for Computational Linguistics*. *V. 1*, 2022, pp. 5180–5197. https://doi.org/10.18653/v1/2022.acl-long.356
8. Xu X., Gou Z., Wu W., Niu Z-Y., Wu H., Wang H., Wang S. Long time no see! open-domain conversation with long-term persona memory. *Findings of the Association for Computational Linguistics:*

ACL 2022. 2022. P. 2639–2650. https://doi.org/10.18653/v1/2022.findings-acl.207
9. Kuchaiev O., Li J., Nguyen H., Hrinchuk O., Leary R., Ginsburg B., Kriman S., Beliaev S., Lavrukhin V., Cook J., Castonguay P., Popova M., Huang J., Cohen J.M. NeMo: A toolkit for building ai applications using neural modules // arXiv. 2019. arXiv:1909.09577. https://doi.org/10.48550/arXiv.1909.09577
10. Gulati A., Qin J., Chiu C.-C., Parmar N., Zhang Y., Yu J., Han W., Wang S., Zhang Z., Wu Y., Pang R. Conformer: Convolution-augmented transformer for speech recognition // Proc. of the Interspeech 2020. P. 5036–5040. https://doi.org/10.21437/interspeech.2020-3015
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems. 2017. V. 30.
12. Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P.J. Exploring the limits of transfer learning with a unified text-to-text transformer // Journal of Machine Learning Research. 2020. V. 21. P. 140.
13. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D. Language models are few-shot learners // Advances in Neural Information Processing Systems. 2020. V. 33. P. 1877–1901.
14. Jelinek F., Mercer R.L., Bahl L.R., Baker J.K. Perplexity – a measure of the difficulty of speech recognition tasks // The Journal of the Acoustical Society of America. 1977. V. 62. N S1. P. S63–S63. https://doi.org/10.1121/1.2016299
15. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation // Proc. of the 40th Annual Meeting on Association for Computational Linguistics. 2002. P. 311–318. https://doi.org/10.3115/1073083.1073135

*ACL 2022*, 2022, pp. 2639–2650. https://doi.org/10.18653/v1/2022.findings-acl.207
9. Kuchaiev O., Li J., Nguyen H., Hrinchuk O., Leary R., Ginsburg B., Kriman S., Beliaev S., Lavrukhin V., Cook J., Castonguay P., Popova M., Huang J., Cohen J.M. NeMo: A toolkit for building ai applications using neural modules. *arXiv*, 2019. arXiv:1909.09577. https://doi.org/10.48550/arXiv.1909.09577
10. Gulati A., Qin J., Chiu C.-C., Parmar N., Zhang Y., Yu J., Han W., Wang S., Zhang Z., Wu Y., Pang R. Conformer: Convolution-augmented transformer for speech recognition. *Proc. of the Interspeech 2020*, pp. 5036–5040. https://doi.org/10.21437/interspeech.2020-3015
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017, vol. 30.
12. Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020, vol. 21, pp. 140.
13. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 1877–1901.
14. Jelinek F., Mercer R.L., Bahl L.R., Baker J.K. Perplexity — a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, 1977, vol. 62, no. S1, pp. S63–S63. https://doi.org/10.1121/1.2016299
15. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation. *Proc. of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 311–318. https://doi.org/10.3115/1073083.1073135

## Авторы

**Апанасович Кирилл Сергеевич** — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57698703700, https://orcid.org/0000-0001-7966-3488, apanasovich.k@yandex.ru
**Махныткина Олеся Владимировна** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57208002090, https://orcid.org/0000-0002-8992-9654, makhnytkina@itmo.ru
**Кабаров Владимир Иосифович** — старший преподаватель, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57210787844, https://orcid.org/0000-0001-6300-9473, vikabarov@itmo.ru
**Далевская Ольга Петровна** — старший преподаватель, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0001-5246-9212, opdalevskaia@itmo.ru

## Authors

**Kirill S. Apanasovich** — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57698703700, https://orcid.org/0000-0001-7966-3488, apanasovich.k@yandex.ru
**Olesia V. Makhnytkina** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57208002090, https://orcid.org/0000-0002-8992-9654, makhnytkina@itmo.ru

**Vladimir I. Kabarov** — Senior Lecturer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57210787844, https://orcid.org/0000-0001-6300-9473, vikabarov@itmo.ru

**Olga P. Dalevskaya** — Senior Lecturer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0001-5246-9212, opdalevskaia@itmo.ru

*Статья поступила в редакцию 28.10.2023*
*Одобрена после рецензирования 04.02.2024*
*Принята к печати 17.03.2024*

*Received 28.10.2023*
*Approved after reviewing 04.02.2024*
*Accepted 17.03.2024*