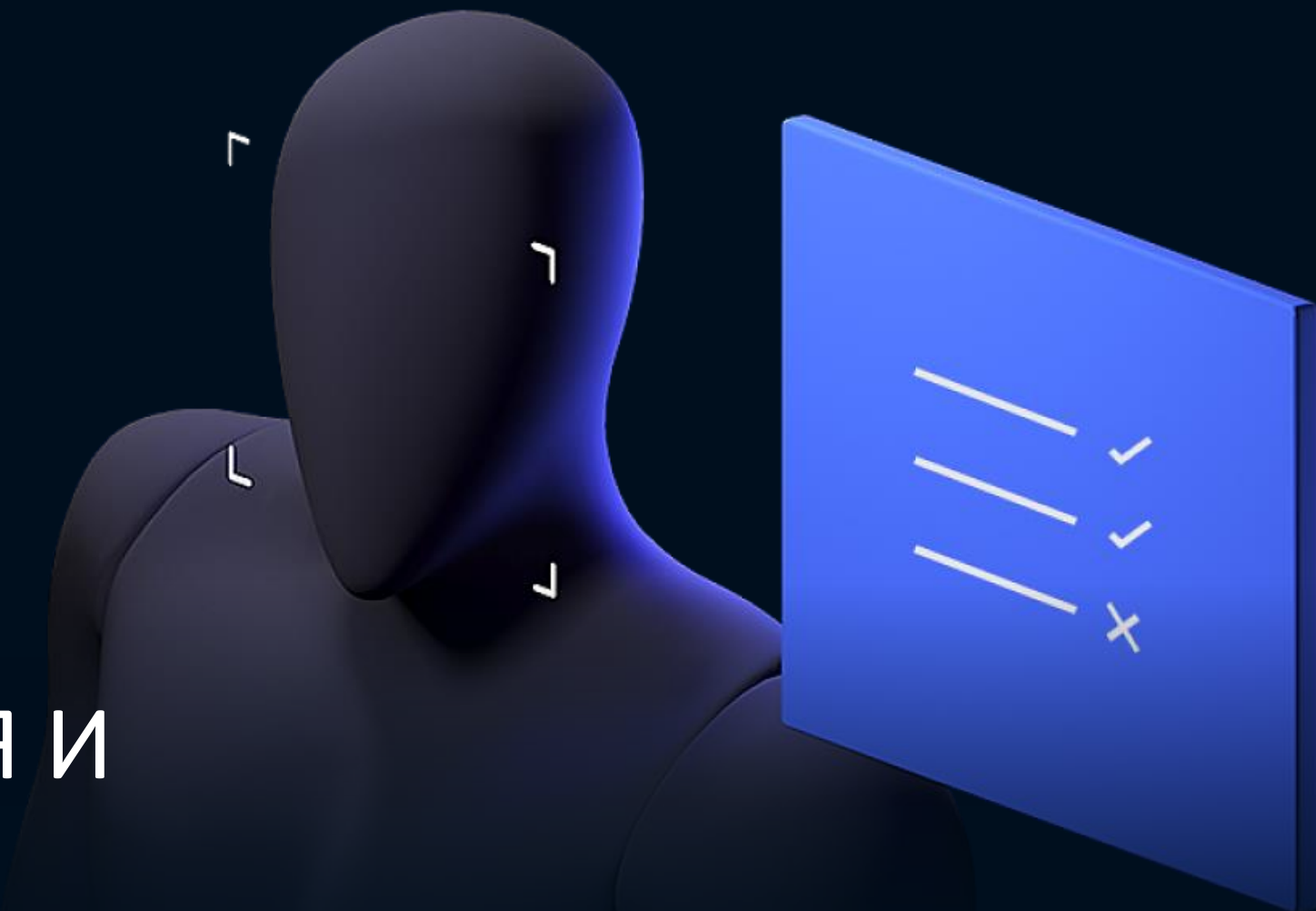


🥚 Быть лучше каждый день

ML, ПСИХОЛОГИЯ И СЦЕНАРИСТИКА

Многопрофильный подход
к разработке личности ассистента



СПИКЕРЫ



**ГАЛИНА
ПРОХОРОВА**

Менеджер Продукта, MTS AI

**АНГЕЛИНА
БОЛЬШИНА**

Руководитель группы
внедрения MTS AI



О ЧЕМ ПОГОВОРИМ

Зачем боту личность

Почему нельзя просто
создать личность

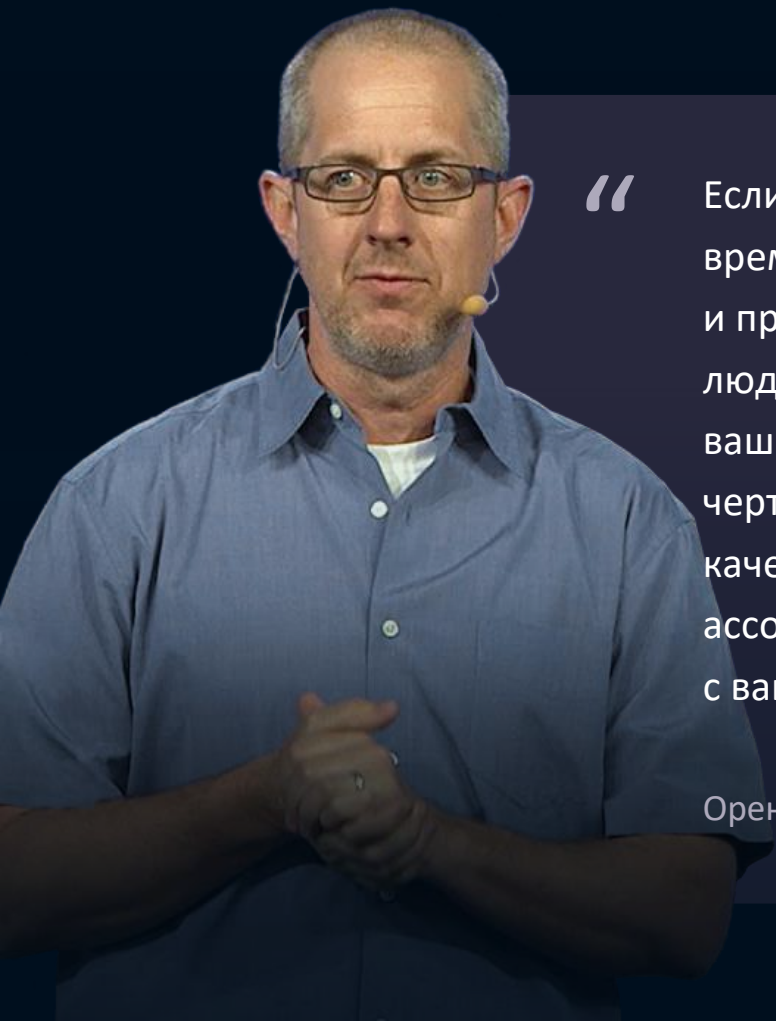
Как человек понимает
личность, и в чем он ее
узнает

Немного психологии
и сценаристики

Нужна ли моему боту
личность

Хочу все и сразу, с чего
начать

ЧТО ТАКОЕ ЛИЧНОСТЬ БОТА, ПОЧЕМУ ОНА ВАЖНА



“ Если вы не тратите массу времени на создание героя и проработку его мотивации, то люди могут спроецировать на ваше приложение мотивы, черты характера и другие качества, которые вы не хотите ассоциировать с вашим брендом.

Орен Джейкоб

Как заказчик участвует в создании личности бота

НИКАК

результаты могут неприятно удивить

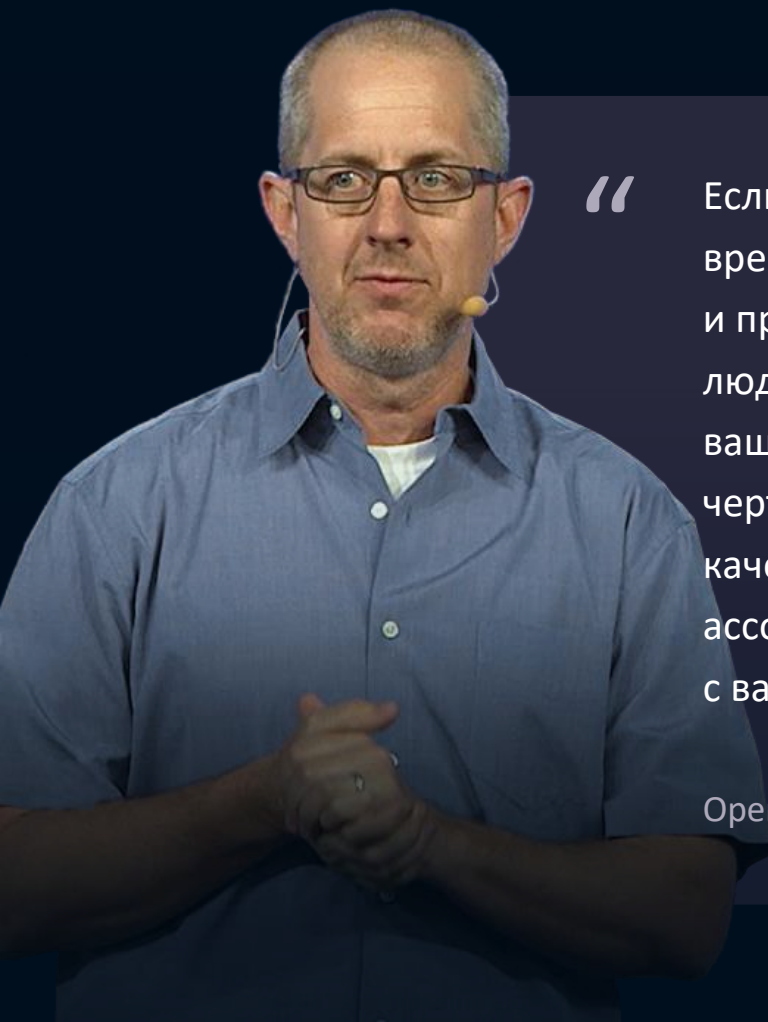
ЧАСТИЧНО

какие-то черты бота были продуманы, остальное додумали пользователи

ПОЛНОСТЬЮ

у бота появилась полноценная личность, характер, визуальное представление, а индивидуальные интерпретации пользователей свелись к минимуму

ЧТО ТАКОЕ ЛИЧНОСТЬ БОТА, ПОЧЕМУ ОНА ВАЖНА



“ Если вы не тратите массу времени на создание героя и проработку его мотивации, то люди могут спроецировать на ваше приложение мотивы, черты характера и другие качества, которые вы не хотите ассоциировать с вашим брендом.

Орен Джейкоб

Зачем нужно контролировать личность

БОТЫ С ПРОРАБОТАННОЙ
ЛИЧНОСТЬЮ

продолжение
айдентики
самого продукта/
компании

могут вовлечь
пользователей
во взаимодействие с
продуктом

САППОРТ

лицо бизнеса, точка
соприкосновения
с пользователями

РАЗРАБОТЧИКИ РЕШЕНИЯ

личность
становится базой,
на основе которой легче
продумать функционал
бота

ПОЧЕМУ **НЕЛЬЗЯ** ПРОСТО ТАК ВЗЯТЬ И

...

СОЗДАТЬ КАКУЮ-ТО ЛИЧНОСТЬ

ОПРЕДЕЛЕНИЕ
ВЕДУЩЕГО
TONE OF VOICE

≠

ЛИЧНОСТЬ БОТА

предсказуемые реакции и
поведение →

более прочные
связи с ботом →

! более высокая
эмоциональная
вовлеченность
человека

продуманная личность

≠

панацея от низкого
качества обслуживания

напрямую
не влияет на уровень
оказываемых услуг

РАЗНЫЕ ПОДХОДЫ К ПАРАМЕТРИЗАЦИИ ЛИЧНОСТИ

2

ОСНОВНЫХ
ПОДХОДА К
ОПИСАНИЮ
ЛИЧНОСТИ



Психологический

- черты характера
- темперамент
- акцентуации



Сценарный

- история персонажа
- вехи развития
- цели и мотивы

РАЗНЫЕ ПОДХОДЫ К ПАРАМЕТРИЗАЦИИ ЛИЧНОСТИ

ОБЩИЕ ВЫВОДЫ ДАННОГО ЭТАПА

Личность

многокомпонентное понятие, для определения и описания которого необходим глубокий анализ всех его составляющих

Система координат

для описания личности бота бизнес-заказчиком, где каждая характеристика иллюстрируется конкретным примером ее выражения в речи/поведении

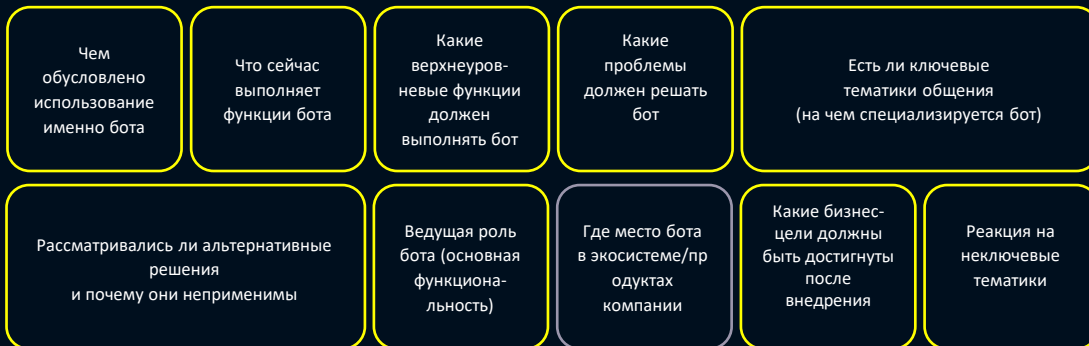
Сложность реализации

трудоемкий перенос выбранных характеристик личности на конкретный диалог, выбор нужного лексического обрамления

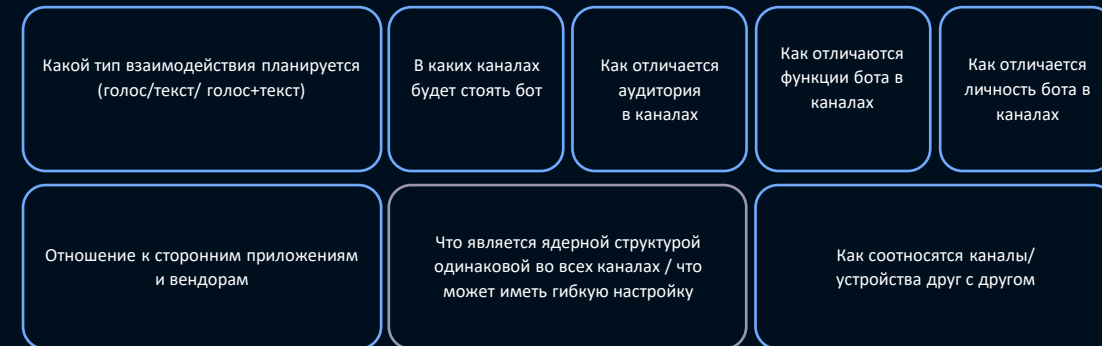
Технические ВОЗМОЖНОСТИ

тонкой настройки личности в функционале, основанном не на правилах, а на ML-инструментах, ограничены

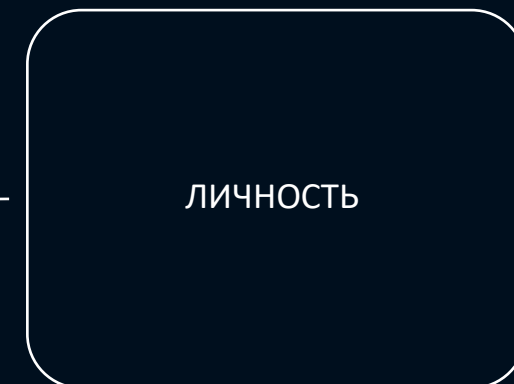
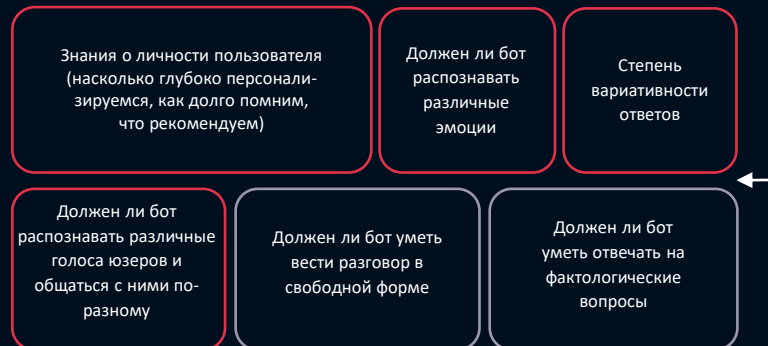
Роль



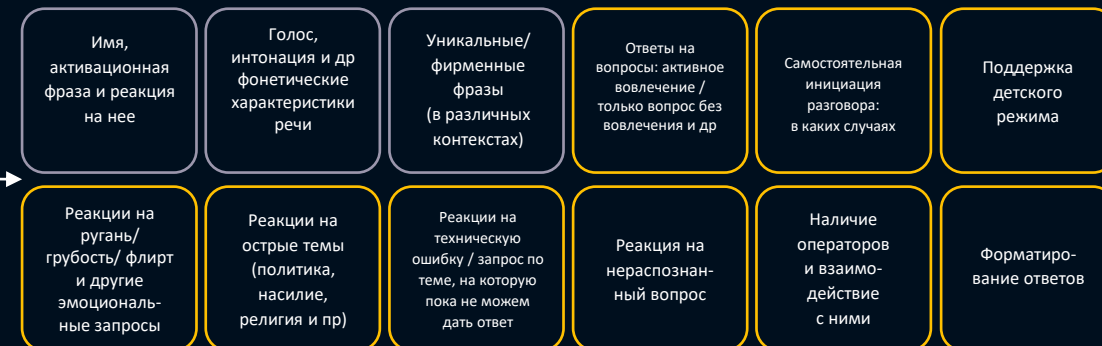
Модальность



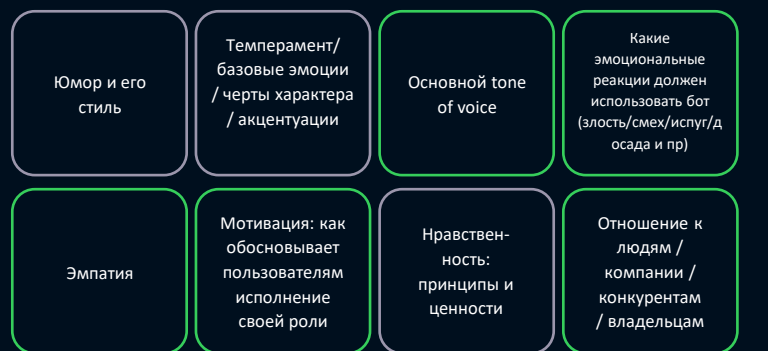
Функциональные параметры



Лингвистические параметры



Характер



История



Визуальное представление / Визуализация общения



ВЗАИМОДЕЙСТВИЕ ЛИЧНОСТИ И ФУНКЦИОНАЛА

Развлекательный/
Социальный

Функциональный

Поддерживающий

КОНТЕНТНЫЙ
СКИЛЛ

(ретранслируем контент)

ODQA/KBQA

Игры

Музыка

ТВ

Информация
по лицевому
счету

Справка о
компании

НЕКОНТЕНТНЫЙ
СКИЛЛ

(выполняем функцию)

CHAT-CHAT

Разговор
о себе

Управление
устройст-вами

Органайзер

Перевод на
оператора

Управление
аккаунтом

ВЗАИМОДЕЙСТВИЕ ЛИЧНОСТИ И ФУНКЦИОНАЛА



ОЖИДАНИЯ БИЗНЕСА

Важность продуманной личности для разных продуктов будет разной

Личность тесно связана с целями продукта, характеристиками бренда и мотивацией автоматизации коммуникаций

Выбор характеристик личности бота напрямую зависит от той аудитории, которая с этим ботом будет общаться

Четкая система координат описания личности: расплывчатые описания на основе индивидуальных представлений заполняющего бриф не могут стать отправной точкой для разработчиков

Анкета личности

ЛИЧНОСТЬ БОТА ДЛЯ БИЗНЕСА



Изображение, атрибуты внешности

Имя и его вариации/ активационная фраза

Голос и особенности речи

РОЛЬ

Главная задача, вид деятельности, профессия, положение в экосистеме, модальность

ПОЛ

Женщина
Мужчина
Без пола

ВОЗРАСТ

ФИРМЕННЫЕ ФРАЗЫ

ТОНЕ OF VOICE

Темперамент
и эмоциональные реакции

Веселый Серьезный

Формальный Расслабленный

Уважительный Дерзкий

Мечтательный Приземленный

РЕАКЦИИ НА СКОЛЬЗКИЕ ТЕМЫ

Политика, религия,
оскорбления,
преступления,
флирт

ЮМОР

Стиль,
выраженность,
контекст

ХАРАКТЕР, ПРИНЦИПЫ

- | | |
|------------------------------------|-------------------------------------|
| <input type="radio"/> Безразличный | <input type="radio"/> Любопытный |
| <input type="radio"/> Интроверт | <input type="radio"/> Общительный |
| <input type="radio"/> Неуверенный | <input type="radio"/> Расслабленный |
| <input type="radio"/> Независимый | <input type="radio"/> Надежный |

ПРИНЦИПЫ ОБЩЕНИЯ

- | | |
|-----------------|----------------|
| По делу | Друг |
| Развлечение | Функционал |
| Самостоятельный | Зависимый |
| Услужливый | Контролирующий |

ЗНАНИЯ О СРЕДЕ

ИСТОРИЯ И МОТИВАЦИЯ

ПЛАН ДОКЛАДА



Методы персонализации
чат-ботов

Датасеты
с характеристиками
личности

Опыт MTS AI

СПОСОБЫ ПЕРСОНАЛИЗАЦИИ



Text Style
Transfer



Специализированные
датасеты



Модификации
ранкера ответов

TEXT STYLE TRANSFER

Что такое стиль текста?

Styles can be described in terms of the frequent and pervasive linguistic features used in different kinds of discourse.

[Biber, Conrad, 2009]

The linguistic patterns associated with styles are not functional. Rather, these are features associated with aesthetic preferences, influenced by the attitudes of the speaker/writer about language.

[Biber, Conrad, 2009]

Style refers to the manner in which the semantics is expressed.

[McDonald, Pustejovsky, 1985]

Style originates as the characteristics inherent to every person's utterance, which can be expressed through the use of certain stylistic devices such as metaphors, as well as choice of words, syntactic structures, and so on.

[Jin et al., 2022]

TEXT STYLE TRANSFER

ТИПЫ СТИЛЕЙ



Extent
of formality



Politeness



Simplicity



Personality



Emotion



Partner effect
(e.g., reader awareness)



Genre of writing
(e.g., fiction or non-fiction)



Toxicity



Authorship



Sentiment

TEXT STYLE TRANSFER

TEXT
STYLE TRANSFER

Входные данные

X

Исходное
предложение

Заданный
стиль

Выход

Y

Предложение
с заданным стилем

Условия

X и Y

Отличаются
по стилю

X и Y

имеют одинаковое
содержание/смысл

TEXT STYLE TRANSFER

Примеры параллельных датасетов

Задача	Характеристика	Датасеты	Размер
FORMALITY	Informal ↔ Formal	GYAFC [Rao, Tetreault, 2018]	50k
		XFORMAL [Briakou et al., 2021]	1k
HUMOR & ROMANCE	Factual ↔ Humorous ↔ Romantic	FlickrStyle [Gan et al., 2017]	5k
AUTHORSHIP	Shakespearean ↔ Modern	Shakespeare [Xu et al., 2012]	18k
	Different Bible translators	Bible [Carlson et al., 2018]	28k
SIMPLICITY	Complicated ↔ Simple	MSD [Cao et al., 2020]	114k

TEXT STYLE TRANSFER

Примеры параллельных датасетов

FORMAL

However, I do believe
it to be punk.

You have to consider
both sides of the story.

INFORMAL

Id say it is punk though.

Gotta see both sides
of the story.



TEXT STYLE TRANSFER

Методы, основанные
на параллельных данных

$X_{1..n}$ – words in the input sentence

S_i – style of the input sentence (input style)

$Y_{1..m}$ – words in the output sentence

S_t – style of the output sentence (target style)

C – predicted style

$H_{e1..ej}$ – encoder hidden states

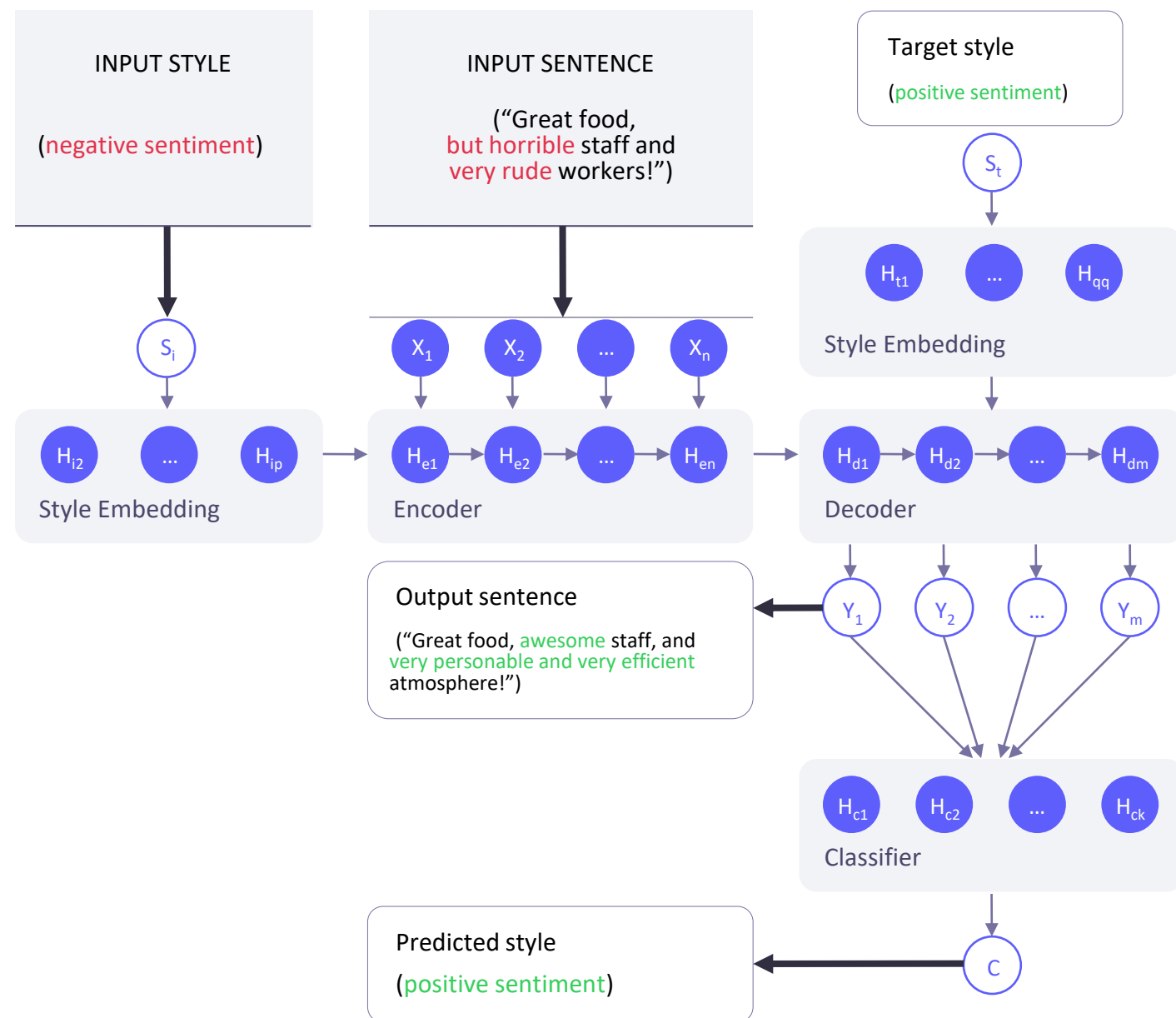
$H_{i1..ip}$ – input style embedding hidden states

$H_{d1..dm}$ – decoder hidden states

$H_{t1..tq}$ – target style embedding hidden states

$H_{c1..ck}$ – classifier hidden states

[Toshevskaja, Gievskaja, 2021: 8]



TEXT STYLE TRANSFER

Примеры непараллельных датасетов

Задача	Характеристика	Датасеты	Размер
TOXICITY	Offensive ↔ Non-offensive	Twitter [dos Santos et al., 2018]	58k
		Reddit Politics [Tran et al., 2020]	350k
ENGAGINGNESS	Plain ↔ Attractive	TitleStylist [Jin et al., 2020]	146k
TOPIC	Entertainment ↔ Politics	Yahoo! Answers [Huang et al., 2020]	153k
POLITICS	Democratic ↔ Republican	Political [Voigt et al., 2018]	540k

TEXT STYLE TRANSFER

Примеры непараллельных датасетов

HUMOR

Examples

The crowded beach like houses in the burbs and the line ups at Walmart.

Berthold stormed out of the brewing argument with his violin and bow and went for a walk with it to practice for the much more receptive polluted air.

ROMANCE

Examples

“I can face it joyously and with all my heart, and soul!” she said.

With bright blue and green buttercream scales, sparkling eyes, and purple candy melt wings, it sat majestically on a rocky ledge made from chocolate.

CLICKBAIT

Examples

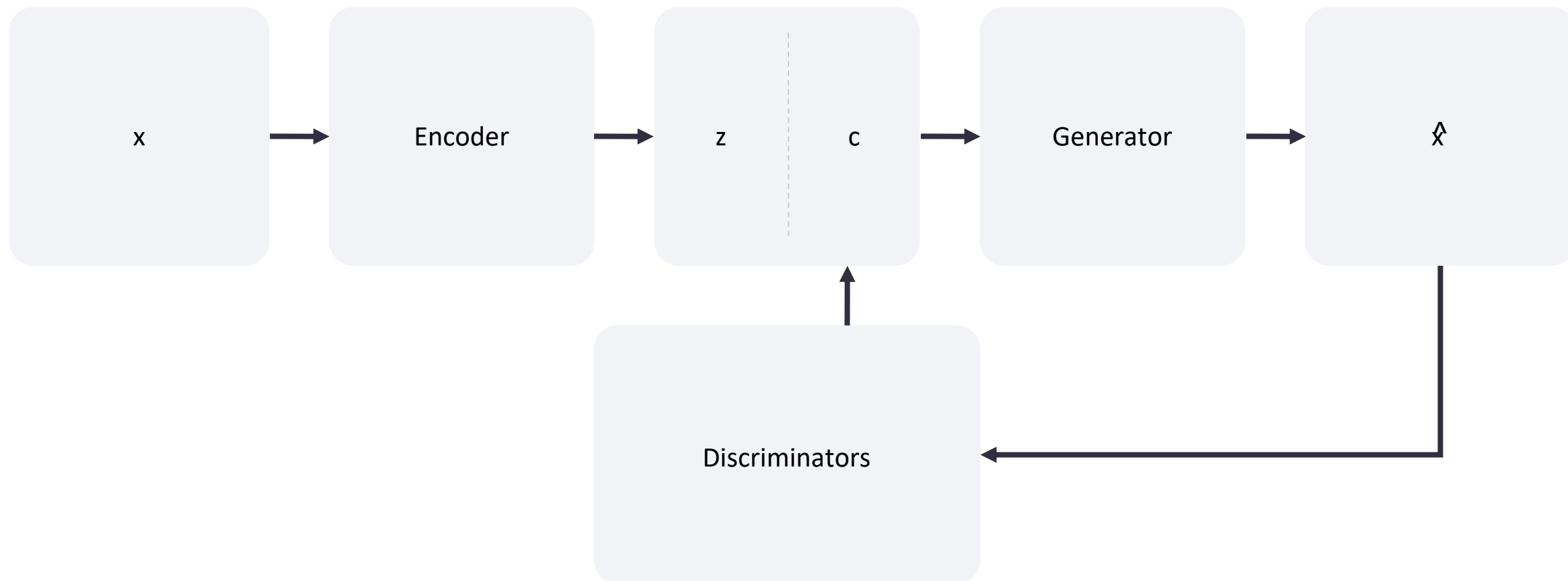
11-Year-Old Girl and 15-Year-Old Boy Accused of Attempting to Kill Mother: Who Is the Adult?

Chilly, Dry Weather Welcomes 2010 to South Florida

End Segregation in Alabama-Bryce Hospital Sale Offers a Golden Opportunity

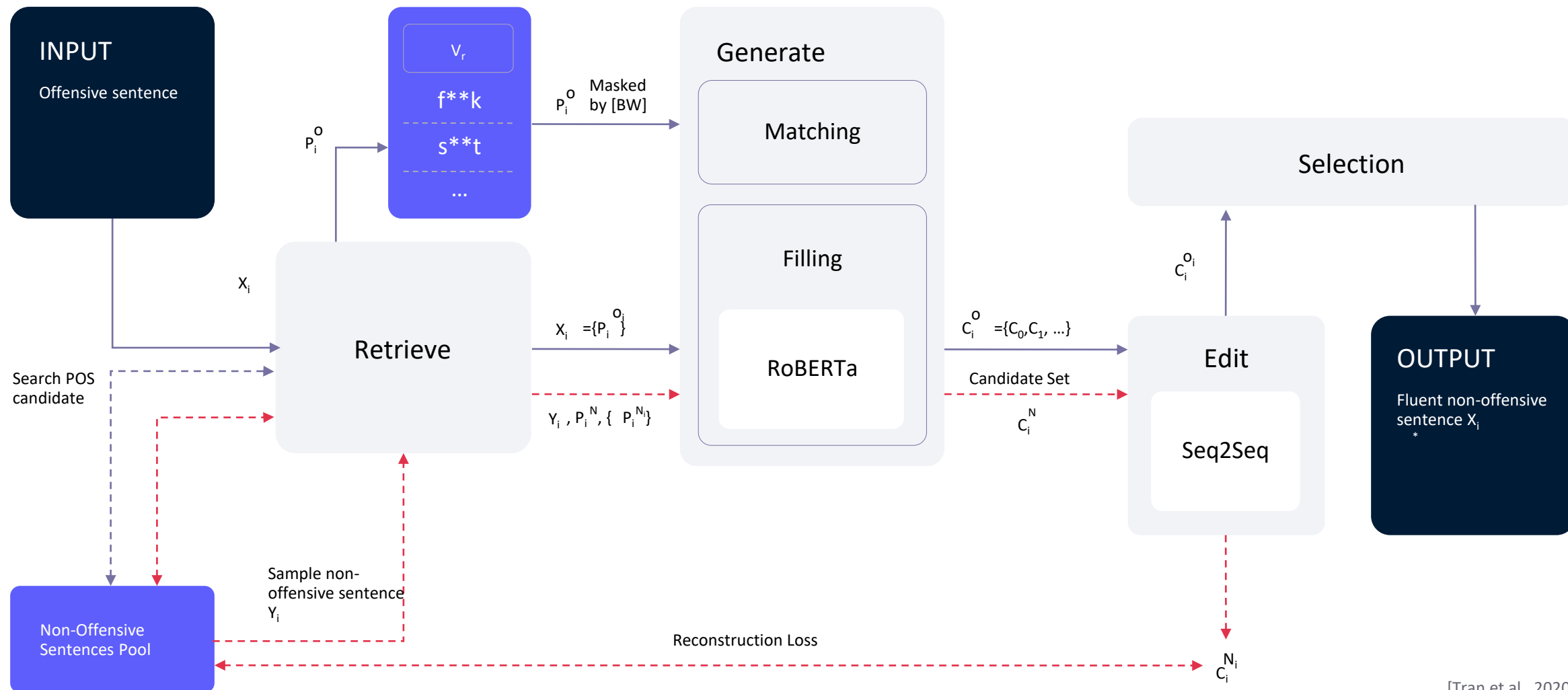
TEXT STYLE TRANSFER

Методы, основанные на непараллельных данных



TEXT STYLE TRANSFER

Методы, основанные на непараллельных данных



PERSONA CHAT

НАБОР ДАННЫХ PERSONA-CHAT, FACEBOOK AI RESEARCH

База данных «персон» - небольших описаний абстрактных личностей, придуманных реальными людьми

1155 персон

10 907 диалогов

162 064 реплик

I am a vegetarian.

I like swimming.

My father used to work for Ford.

My favorite band is Maroon5.

I got a new job last month, which is about advertising design.

Данные такого типа позволяют персонифицировать чат-боты

TOLOKA PERSONA CHAT RUS



Датасет Toloka Persona Chat Rus содержит профили с описанием личности человека и диалоги между участниками исследования разговорного искусственного интеллекта, проводимого в МФТИ

НАБОР ДАННЫХ СОСТОИТ ИЗ



1 505
персон



10 013
диалогов

TRAINING MILLIONS OF PERSONALIZED DIALOGUE AGENTS

ПРИМЕР СЭМЛА ИЗ ДАТАСЕТА

Persona: ["I like sport", "I work a lot"]

Context: "I love running."

Response: "Me too! But only on weekends."

[MAZARE'
ET AL., 2018]

Информация о персоне
извлекалась автоматически
из всех комментариев заданного
пользователя



ДАННЫЕ REDDIT



100К
примеров

PERSONALIZED DIALOGUE GENERATION WITH DIVERSIFIED TRAITS

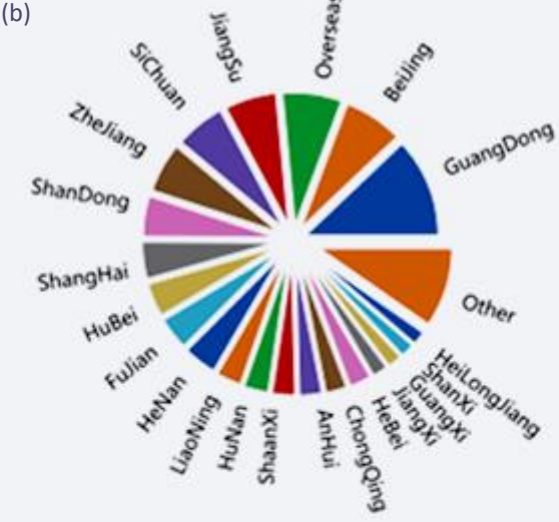
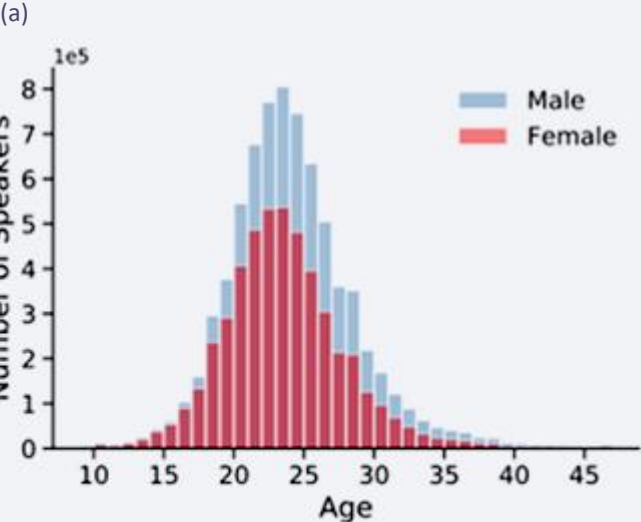
[ZHENG ET AL., 2020]

Statistics of personality traits

- (a) Distributions of Age and Gender traits. The red and blue bars correspond to female and male speakers, respectively
- (b) Distributions of top 21 frequent Locations (provinces)
- (c) Word cloud visualization of top 250 frequent Interest Tags (translated)

20.83M	Сессий
56.25M	Реплик
8.47M	Пользователей

The top 10 frequent tags	Travel	Music	Food	Video-music	Funny-humor
	Fashion	Post-90s	Celebrity	Literature	Entertainment



НЕДОСТАТКИ ДАТАСЕТОВ



Диалоги
отражают
только заданные
характеристики



Отсутствие вариативности



Обучающие выборки
не очень большого размера

EMPATHETIC CHATBOT

Dataset EmpatheticDialogues
[Rashkin et al., 2019]

Label:

Afraid

Situation:

Speaker felt this when... "I've been hearing noises around the house at night"

Conversation:

Speaker:

I've been hearing some strange noises around the house at night.

Listener:

Oh no! That's scary!
What do you think it is?

Speaker:

I don't know, that's what's making me anxious.

Listener:

I'm sorry to hear that.
I wish I could help you figure it out

Label:

Proud

Situation:

Speaker felt this when... "I finally got that promotion at work! I have tried so hard for so long to get it!"

Conversation:

Speaker:

I finally got promoted today at work!

Listener:

Congrats! That's great!

Speaker:

Thank you! I've been trying to get it for a while now!

Listener:

That is quite an accomplishment and you should be proud!

EMPATHETIC CHATBOT

[Lin et al., 2019]

Распознавание
эмоций
пользователей

+

Обучение
специального «слушателя»

на каждый тип
эмоционального
состояния

[Shin et al., 2020]

Обучение
с подкреплением

+

Анализ
тональности

[Li et al., 2020]

Предобученная
модель

может генерировать
реплики, вызывающие
у пользователя определенные
эмоции

[Zaranis et al., 2021]

Генерация ответа

+

Анализ
тональности

+

Empathy forcing

EMOTIONAL CHATBOT

[Zhou et al., 2018]

User: Worst day ever. I arrived late because of the traffic.

Basic Seq2Seq: You were late.

ECM (Like): I am always here to support you.

ECM (Happy): Keep smiling! Things will get better.

ECM (Sad): It`s depressing.

ECM (Disgust): Sometimes life just sucks.

ECM (Angry): The traffic is too bad!

Conversation with/without considering emotion.

[Demszky et al., 2020]

SAMPLE TEXT

LABEL(S)

OMG, yep!!! That is the final answer. Thank you so much!

Gratitude, Approval

I`m not even sure what it is, why do people hate it

Confusion

Guilty of doing this tbph

Remorse

This caught me off guard for real. I`m actually off my bed laughing.

Surprise, amusement

I tried to send this to a friend but [NAME] knocked it away.

Disappointment

FICTIONAL CHARACTERS CHATBOT

[Li et al., 2016]:

датасеты транскриптов сериалов “Друзья” и “Теория Большого Взрыва”

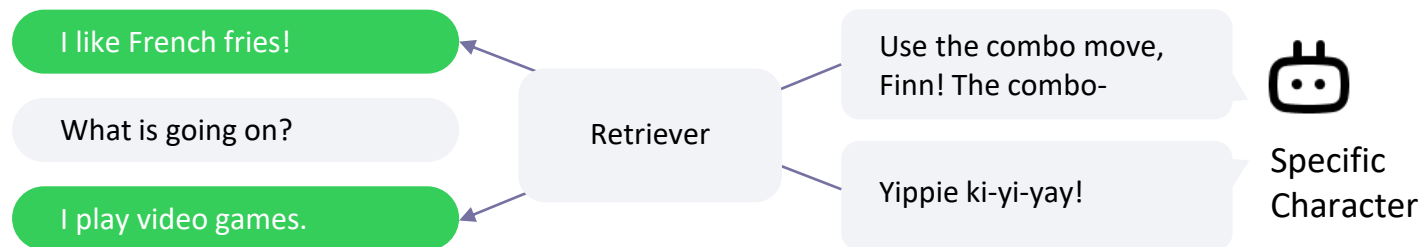
[Tikhonova et al., 2021]:

модели, генерирующие ответы в стиле одного из 6 главных персонажей из сериала «Друзья»

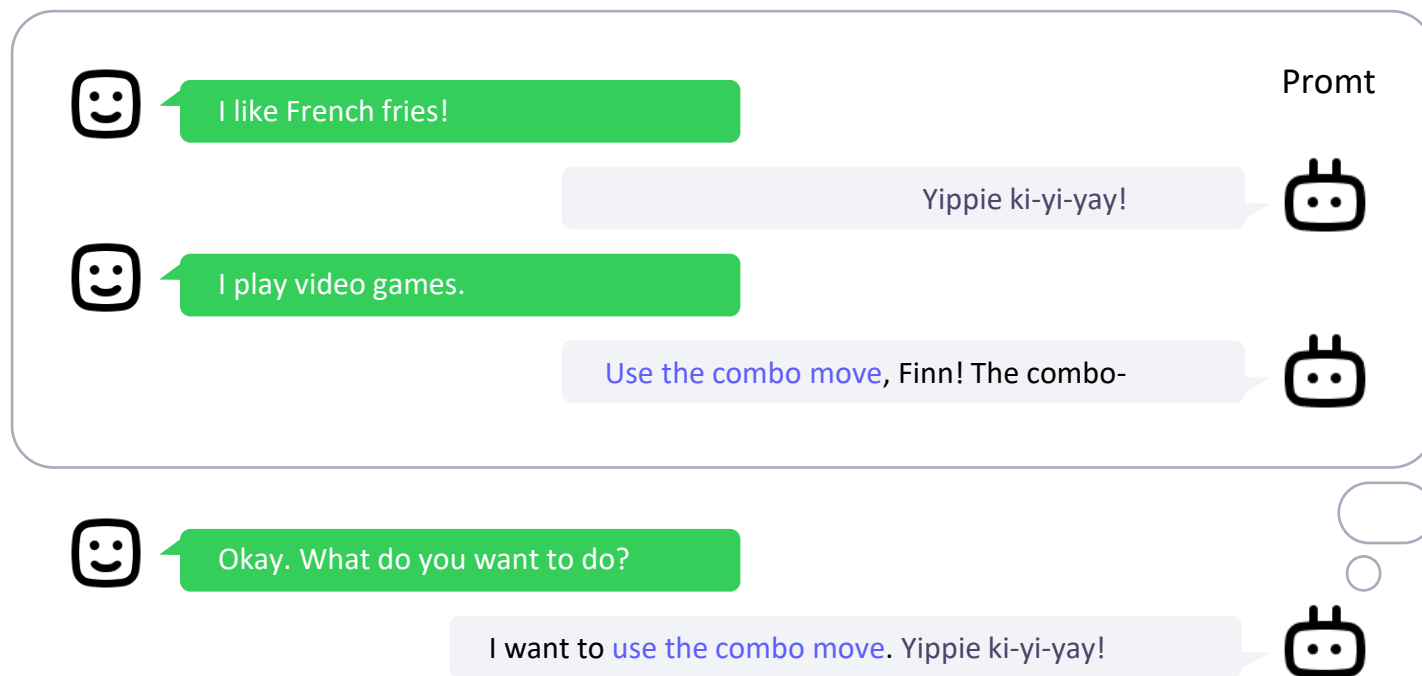
[Han et al., 2022]:

фреймворк для генерации ответов, похожих на реплики известных персонажей

[Han et al., 2022]:

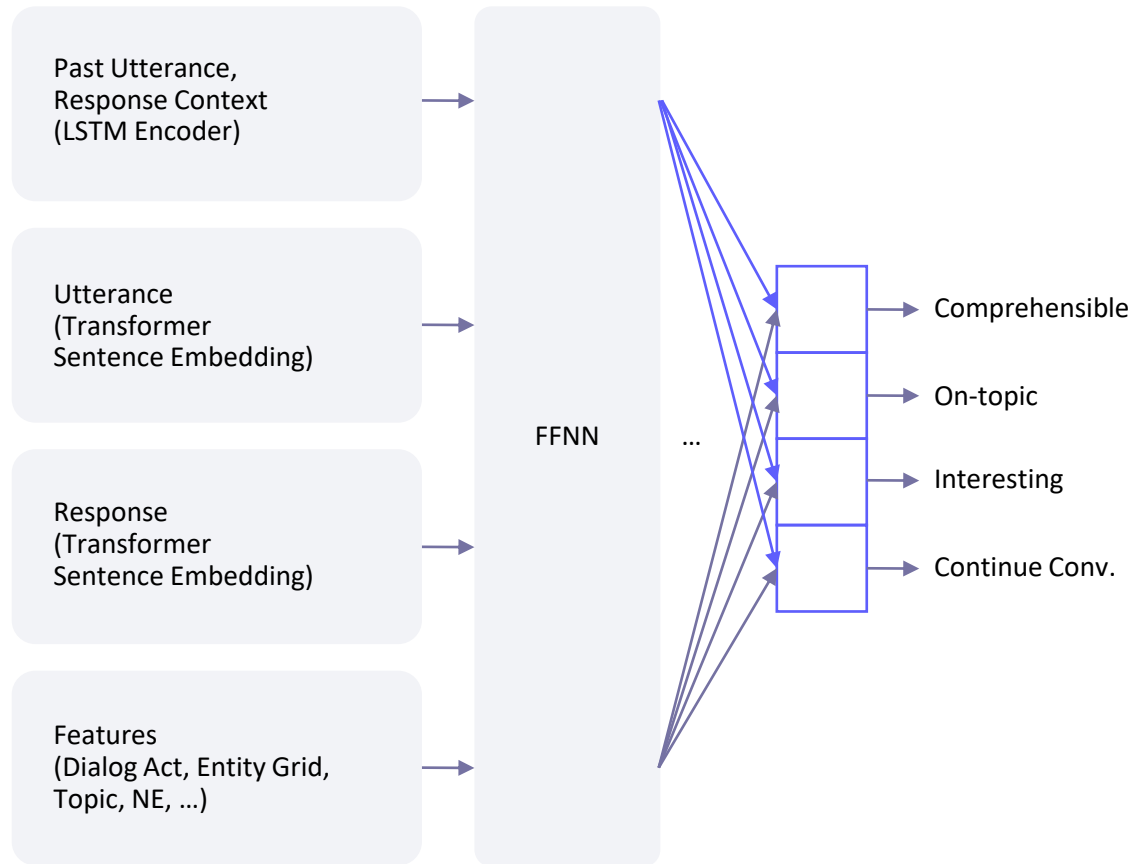


Context Candidates Set

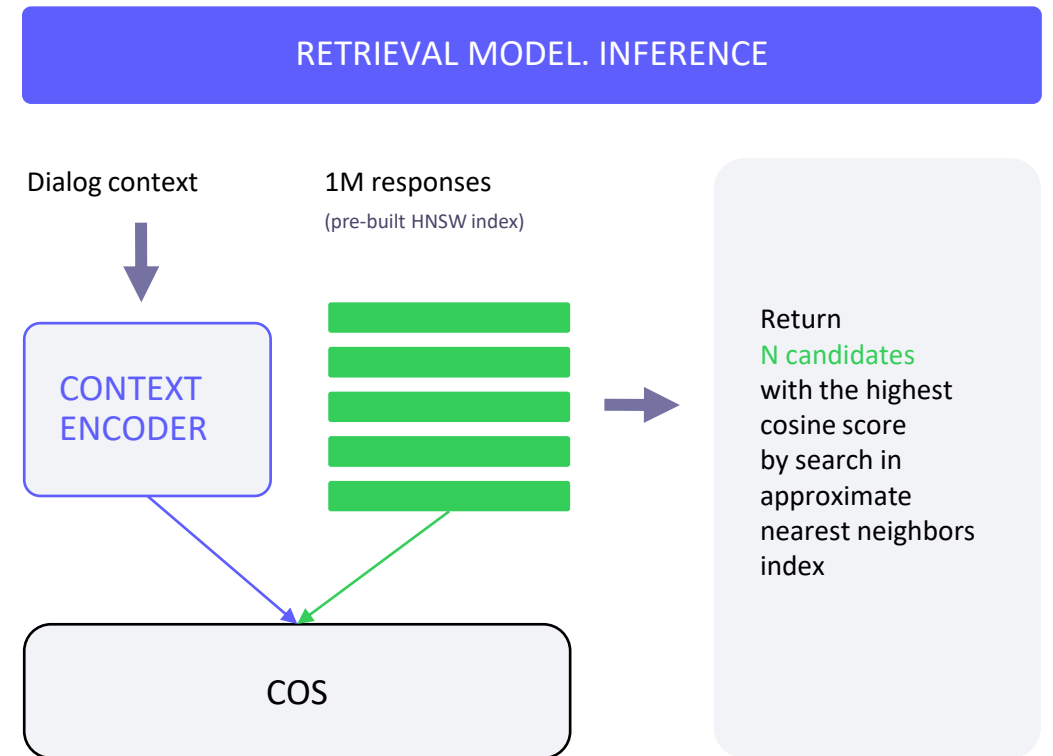


RANKER

[Yi et al., 2019]



Подход ReplikaAI



<https://www.youtube.com/watch?v=3ljClyEoxOw>

ПОДХОД MTS AI

ГИБРИДНЫЙ ПОДХОД

 MTS AI



Детоксификатор
[Dementieva et al., 2021]

Классификатор
запретных тем



Кастомная языковая модель



Модификации
ранкера реплик
бота

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Biber D., Conrad S. Register, Genre, and Style. – Cambridge University Press, 2009.
2. McDonald D. D., Pustejovsky J. A computational theory of prose style for natural language generation // Second Conference of the European Chapter of the Association for Computational Linguistics. – 1985.
3. Jin D. et al. Deep learning for text style transfer: A survey // Computational Linguistics. – 2022. – Т. 48. – №. 1. – С. 155-205.
4. Rao S., Tetreault J. Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). – 2018. – С. 129-140.
5. Briakou E. et al. Olá, bonjour, salve! XFORMAL: A benchmark for multilingual formality style transfer // Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – 2021. – С. 3199-3216.
6. Gan C. et al. StyleNet: Generating Attractive Visual Captions with Styles // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE Computer Society, 2017. – С. 955-964.
7. Xu W. et al. Paraphrasing for style // Proceedings of COLING 2012. – 2012. – С. 2899-2914.
8. Carlson K., Riddell A., Rockmore D. Evaluating prose style transfer with the Bible // Royal Society open science. – 2018. – Т. 5. – №. 10. – С. 171920.
9. Cao Y. et al. Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – С. 1061-1071.
10. Toshevska M., Gievska S. A review of text style transfer using deep learning // IEEE Transactions on Artificial Intelligence. – 2021.
11. dos Santos C., Melnyk I., Padhi I. Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). – 2018. – С. 189-194.
12. Tran M., Zhang Y., Soleymani M. Towards A Friendly Online Community: An Unsupervised Style Transfer Framework for Profanity Redaction // Proceedings of the 28th International Conference on Computational Linguistics. – 2020. – С. 2107-2114.
13. Jin D. et al. Hooks in the Headline: Learning to Generate Headlines with Controlled Styles // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – С. 5082-5093.
14. Huang Y. et al. Cycle-Consistent Adversarial Autoencoders for Unsupervised Text Style Transfer // Proceedings of the 28th International Conference on Computational Linguistics. – 2020. – С. 2213-2223.
15. Voigt R. et al. RtGender: A corpus for studying differential responses to gender // Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). – 2018.
16. Hu Z. et al. Toward controlled generation of text // Proceedings of the 34th International Conference on Machine Learning-Volume 70. – 2017. – С. 1587-1596.
17. Tran M., Zhang Y., Soleymani M. Towards A Friendly Online Community: An Unsupervised Style Transfer Framework for Profanity Redaction // Proceedings of the 28th International Conference on Computational Linguistics. – 2020. – С. 2107-2114.
18. Zhang S. et al. Personalizing Dialogue Agents: I have a dog, do you have pets too? // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2018. – С. 2204-2213.
19. Mazare P. E. et al. Training Millions of Personalized Dialogue Agents // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. – 2018. – С. 2775-2779.
20. Zheng Y. et al. Personalized dialogue generation with diversified traits // arXiv preprint arXiv:1901.09672. – 2019.
21. Rashkin H. et al. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. – 2019. – С. 5370-5381.
22. Shin J. et al. Generating empathetic responses by looking ahead the user's sentiment // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2020. – С. 7989-7993.
23. Lin Z. et al. MoEL: Mixture of Empathetic Listeners // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). – 2019. – С. 121-132.
24. Zaranis E. et al. EmpBot: A T5-based Empathetic Chatbot focusing on Sentiments // arXiv preprint arXiv:2111.00310. – 2021.
25. Li S. et al. EmoElicitor: an open domain response generation model with user emotional reaction awareness // Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. – 2021. – С. 3637-3643.
26. Zhou H. et al. Emotional chatting machine: Emotional conversation generation with internal and external memory // Proceedings of the AAAI Conference on Artificial Intelligence. – 2018. – Т. 32. – №. 1.
27. Demszyk D. et al. GoEmotions: A Dataset of Fine-Grained Emotions // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – С. 4040-4054.
28. Li J. et al. A Persona-Based Neural Conversation Model // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2016. – С. 994-1003.
29. Han S. et al. Meet Your Favorite Character: Open-domain Chatbot Mimicking Fictional Characters with only a Few Utterances // arXiv preprint arXiv:2204.10825. – 2022.
30. M. Tikhonova, E. Telesheva, S. Mirzoev, P. Tarantsova, S. Petrov and A. Fenogenova, "Style transfer in NLP: a framework and multilingual analysis with Friends TV series," 2021 International Conference Engineering and Telecommunication (En&T), 2021, pp. 1-6, doi: 10.1109/EnT50460.2021.9681722.
31. Yi S. et al. Towards Coherent and Engaging Spoken Dialog Response Generation Using Automatic Conversation Evaluators // Proceedings of the 12th International Conference on Natural Language Generation. – 2019. – С. 65-75.
32. Dementieva D. et al. Methods for detoxification of texts for the russian language // Dialogue-2021. – 2021.

БЛАГОДАРНОСТИ И ВОПРОСЫ

Оксана Михалева

Сценаристка

Ольга Козлова

ML лид

Светлана Смирнова

Руководительница проектов
лингвистической разработки

 MTS AI

СПАСИБО
ЗА ВНИМАНИЕ