Yandex

Yandex

Dialogue Systems

Vyacheslav (Slava) Alipov, Principal R&D Engineer
SHAD, November 7, 2019

Overview

- Dialogue Interfaces
- Goal-Oriented Dialogue Systems
- General Conversation

Dialogue Interfaces

What is a Dialogue Interface?

- Interacting via voice or text input in a form of a dialogue
- It's easy!
  - Everybody is able to do this
- It's efficient!
  - No complicated GUI manipulations

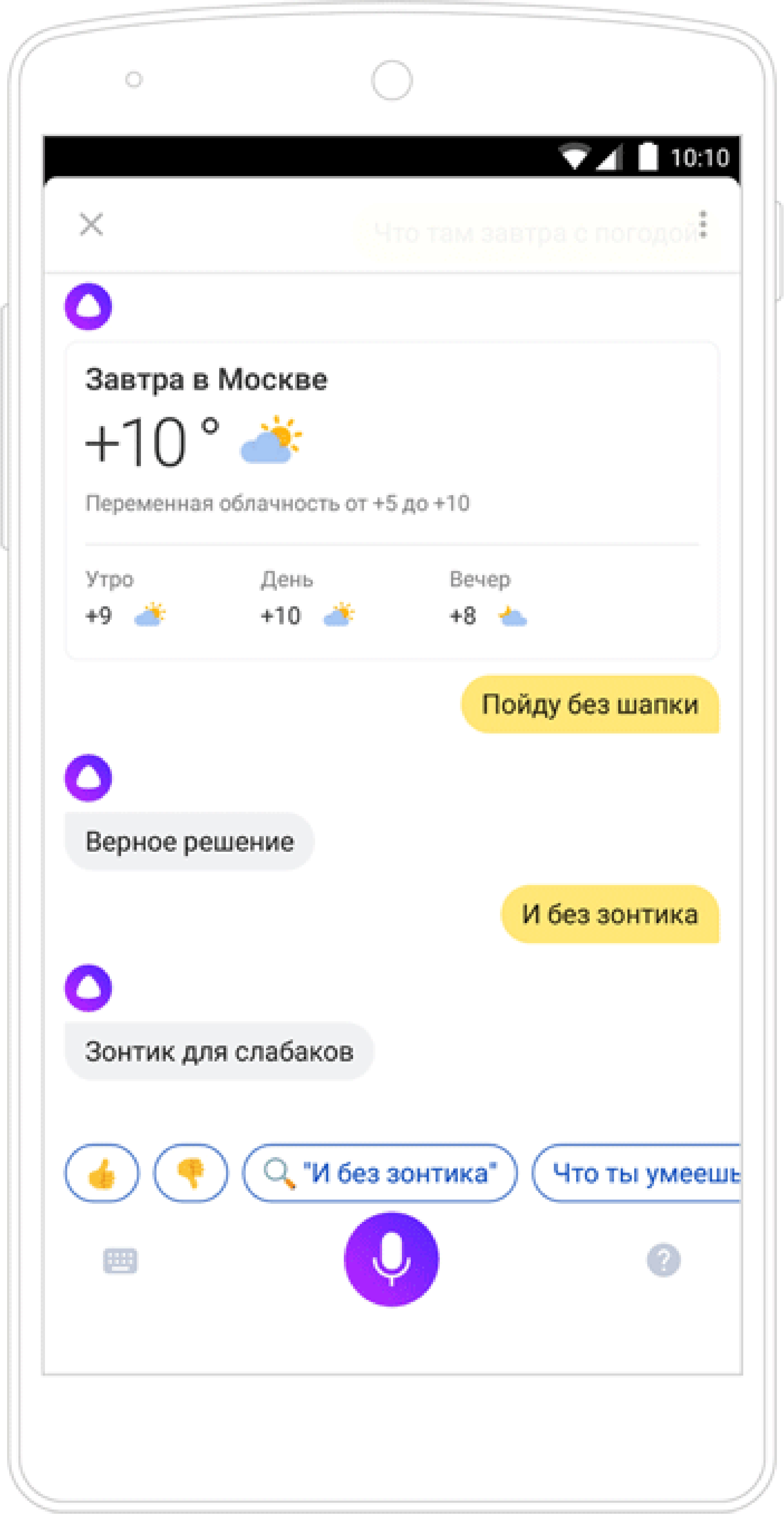
The Time is Now!

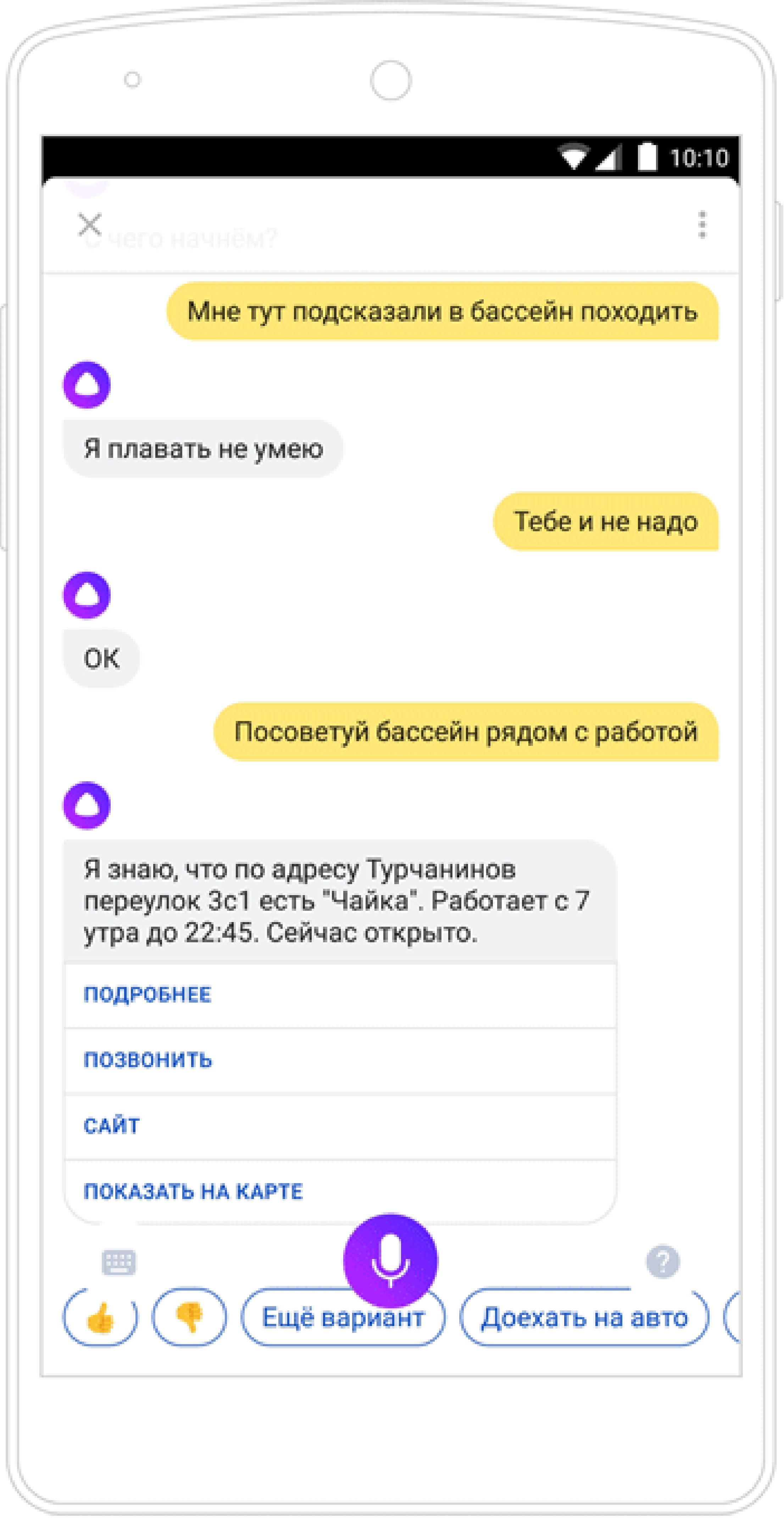
- Automatic Speech Recognition (ASR) is very good
- Text To Speech (TTS) is very good
- Major advances in Natural Language Processing (NLP)
- But still we are far from fully replacing a human assistant


Алиса. Проще — говоря

Alice, what can you do?

- Web Search
- News
- Search for Organizations
  - Cafes, Cinemas, Pharmacies, ...
- Weather
- Routes and Traffic
- Play Music and Video
- Smart Home
- Alarms and Timers
- Chit-Chat!

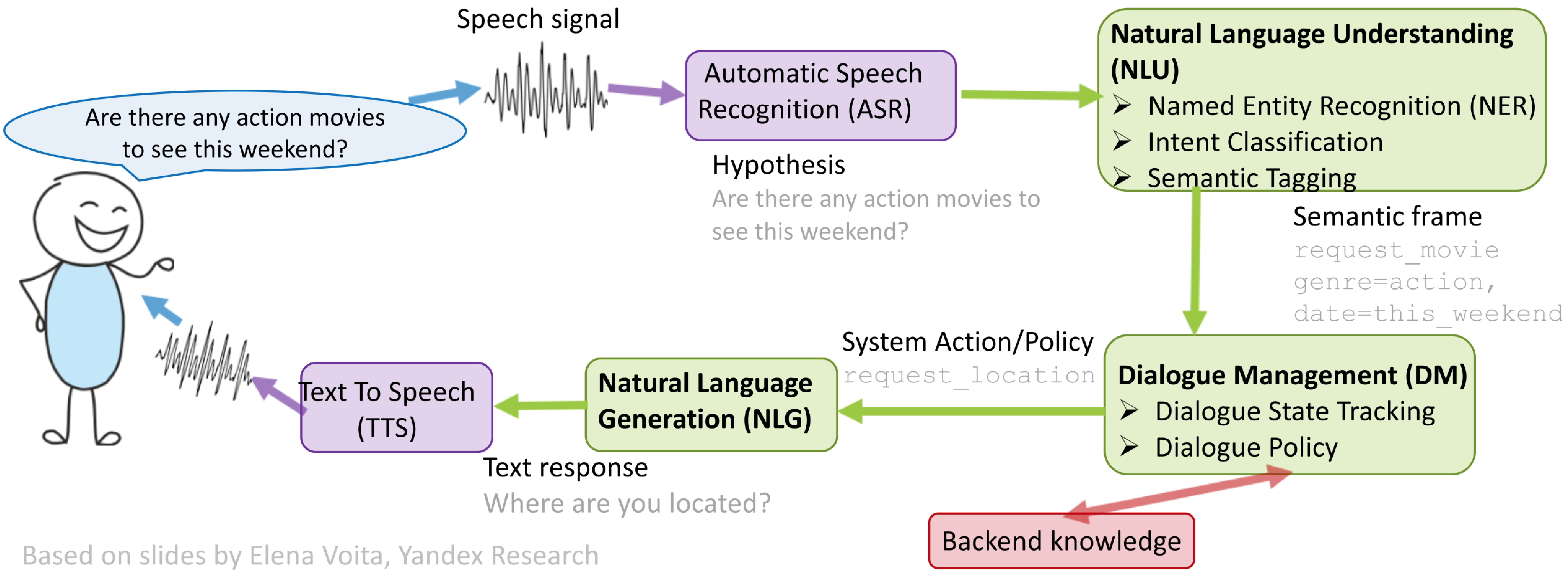




Goal-Oriented

Dialogue Systems

Goal Oriented Dialogue System



Based on slides by Elena Voita, Yandex Research

Natural Language Understanding

Named Entity Recognition (NER)

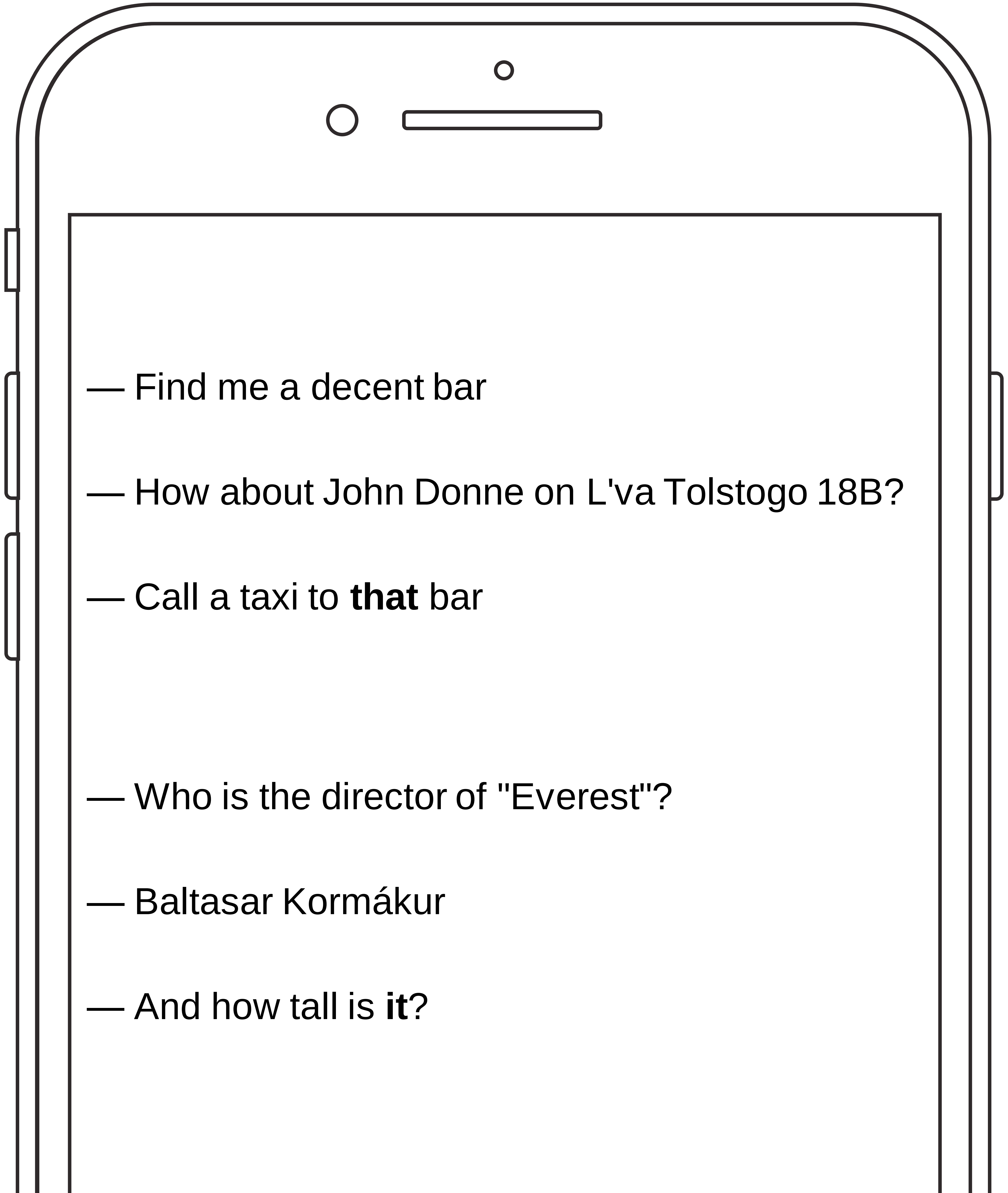
- Goal is to find local structured explanations of user input
- **Finite State Transducers (FST)** based parsers
  - e.g. *time, date, numbers*, …
  - State of the art for such text normalization tasks
- **Gazetteers** – extensive enumeration of all possible entity values
  - e.g. *smart_device_type, fairy_tale_id, phone_contact_id*, …
  - Good when entities are unique and finite (and rarely occur in a dataset)
  - Employ some fuzzy matching / embedding similiarity* to account for misspells and synonyms

Semantic Parsing

- Intent Classification + Semantic Tagging = Semantic Parsing
- Explain user query as Semantic Frame
- Intent Classification – any text classifier would do (BOWs, embeddings, RNNs, etc.)
- Semantic Tagging – any sequence labelling algorithm would do (CRFs, RNNs, etc.)
- Could be performed jointly
  - Probabilistic Context-Free Grammar (PCFG)
  - Augment sequence labelling architecture with intent classification output

Anaphora

- Some cases are easy to hardcode
- Classic approach:
    - Candidate proposal – named groups, NER, etc.
    - Candidate matching – features like gender, animate, case, etc.
    - Candidate ranking
- General approach:
    - Cross sentence semantic tagging



Ellipsis

- Some cases are easy to hardcode
- General approach:
- Cross sentence semantic tagging

— Find a pharmacy nearby

— I would suggest "Apteka 36,6" on Timura Frunze street.

— **What about Lev Tolstoy?**

— Find me a decent bar

— How about John Donne on L'va Tolstogo 18B?

— **Call a taxi**

Dialogue Management

Dialogue Management

- Decision making process with sequences
- Combination of
  - Dialogue State Tracking
  - Dialogue Strategies
- Usually lots of things are hardcoded
  - State is structured and interpretable – training data is scarce
  - Strategies are limited – learning complex strategies requires lots of real user interactions
  - The more data you have the less structured everything needs to be

Dialogue State Tracking

- It's like any other sequence problem!
- All turns of a dialogue up to this moment
- Could be very inefficient – lots of memory, slow inference, lots of training data

- Maintain beam of semantic frames
- Handcrafted rules
- Maximum Entropy models
- Conditional Random Field
- Ranking
- RNNs

Dialogue Strategies

- Dialogue flow is usually hardcoded
  - Finite State Automaton (Call Flow)
  - State – semantic frame with some additional context
  - Edges are marked with semantic frames

Dialogue Strategies – Form Filling

- State
  - Form with several typed slots
- Strategy
  - Ask for values of each slot in linear order – Request(slot_name)
  - Optionally – confirm each slot or completed form – Confirm(slot_name=slot_value)
  - Use completed form to complete user's task and inform user – Inform(form)

```json
"form": {
    "name": "travel",
    "slots": [
        {
            "name": "from",
            "type": "city",
            "is_required": false
        },
        {
            "name": "to",
            "type": "city",
            "prompt": "What city are you travelling to?",
            "is_required": true
        },
        {
            "name": "date",
            "type": "date",
            "prompt": "When are you travelling?",
            "is_required": true
        }
    ],
    "submit": {
        "url": "https://travel.example.ru/dialog/"
    },
    "confirmation": {
        "is_required": true,
        "prompt": "Tickets from {from} to {to} on {date}. Right?"
    }
}
```

Reinforcement Learning for Dialogues

- Search for optimal action through "Trial and Error"
- What is "Reward signal"?
  - Finalizing task (confiming booking, commiting transactions, etc.)
  - Termination of conversation
  - Dialogue length
  - Positive sentiment

Natural Language Generation

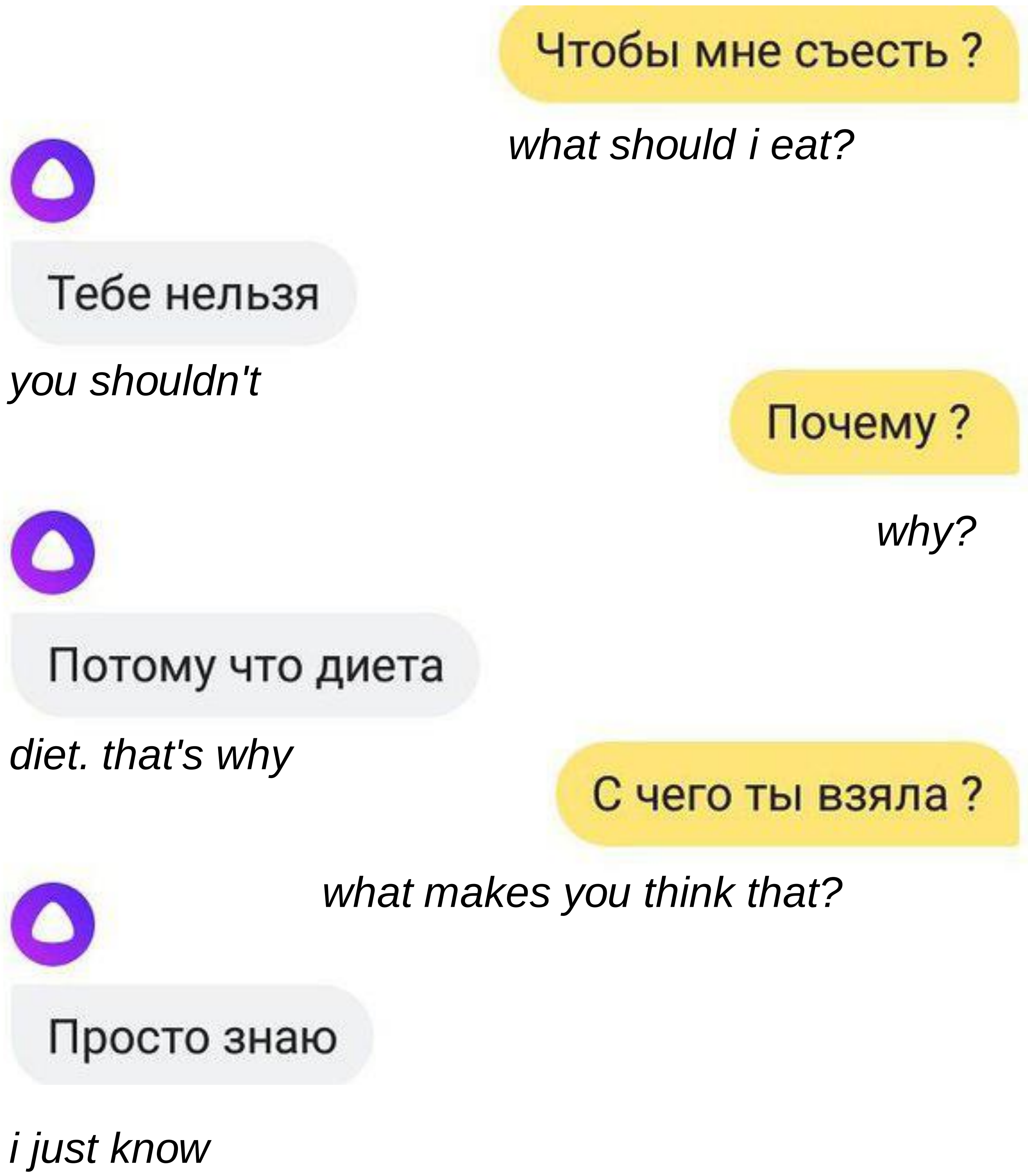
Natural Language Generation

- Set of templates for each dialogue act
- Grammars
- Generative Models (Seq2Seq)

General Conversation

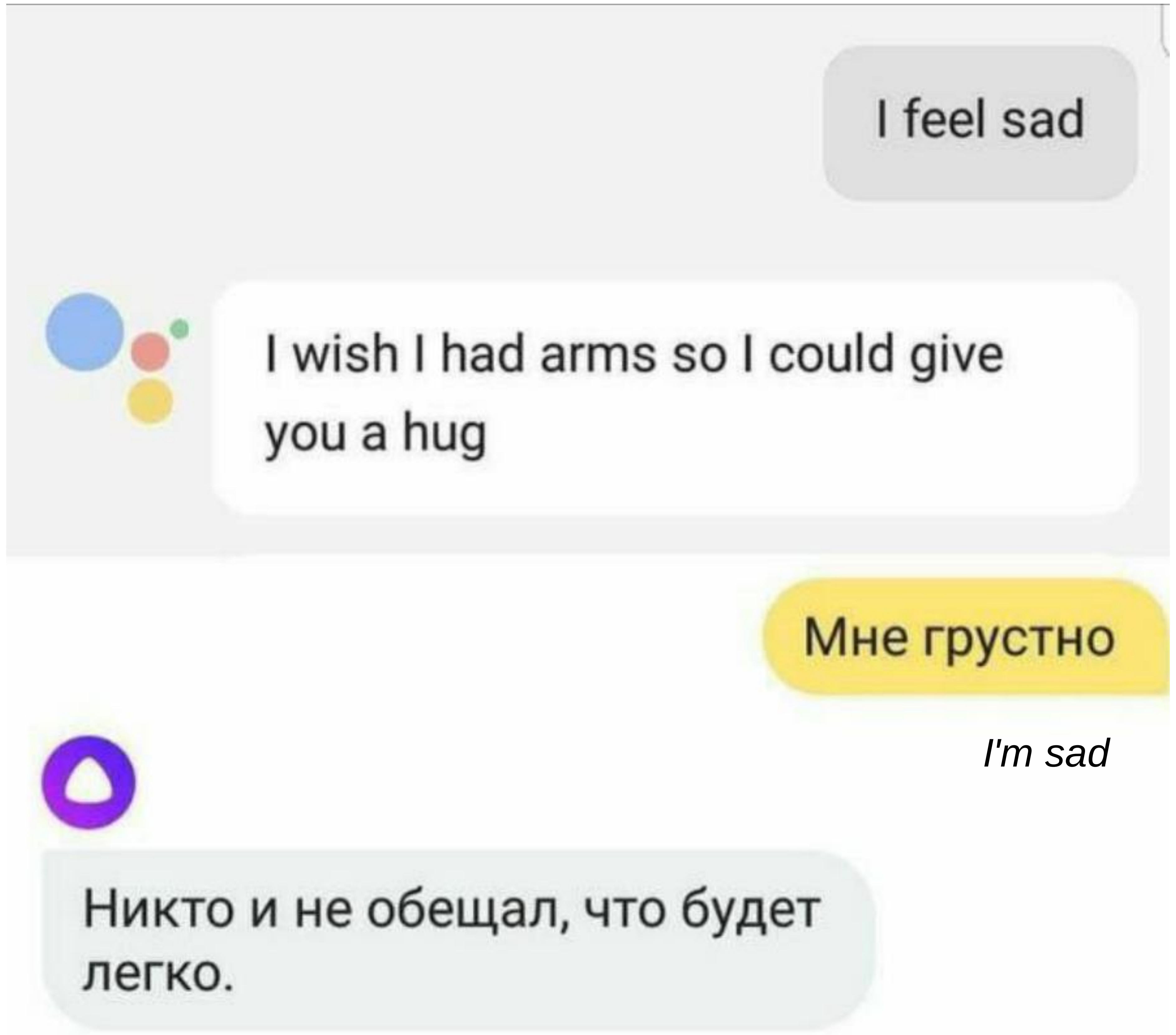
Why Chit-Chatting?

- More human-like experience
- Increases user retention
- Rich and diverse user data
- It's fun!



State of the Art

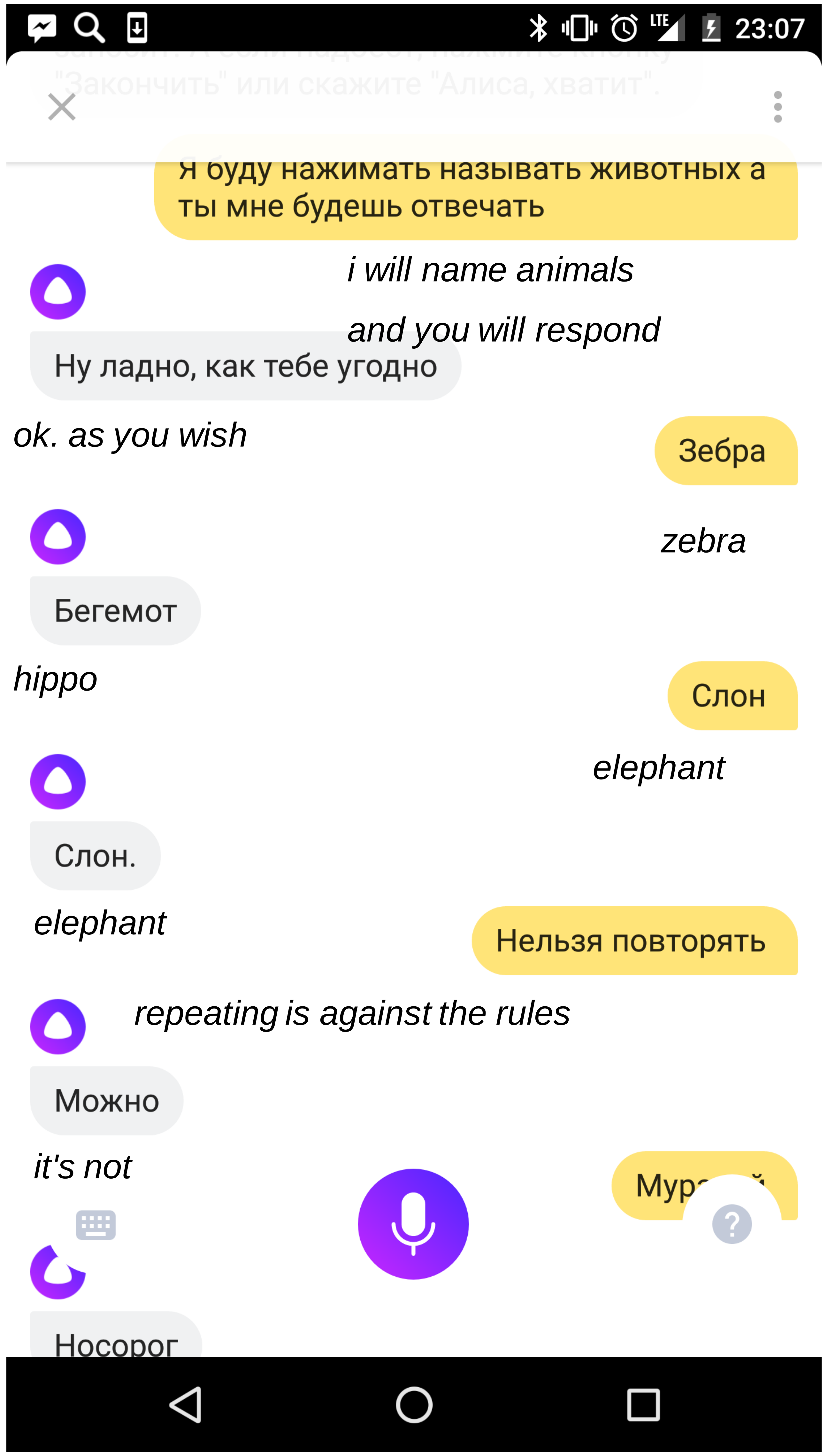
- Set of prewritten responses for certain questions
  - You can not write a response for every user utterance
  - Especially if you take more than one previous turn into account

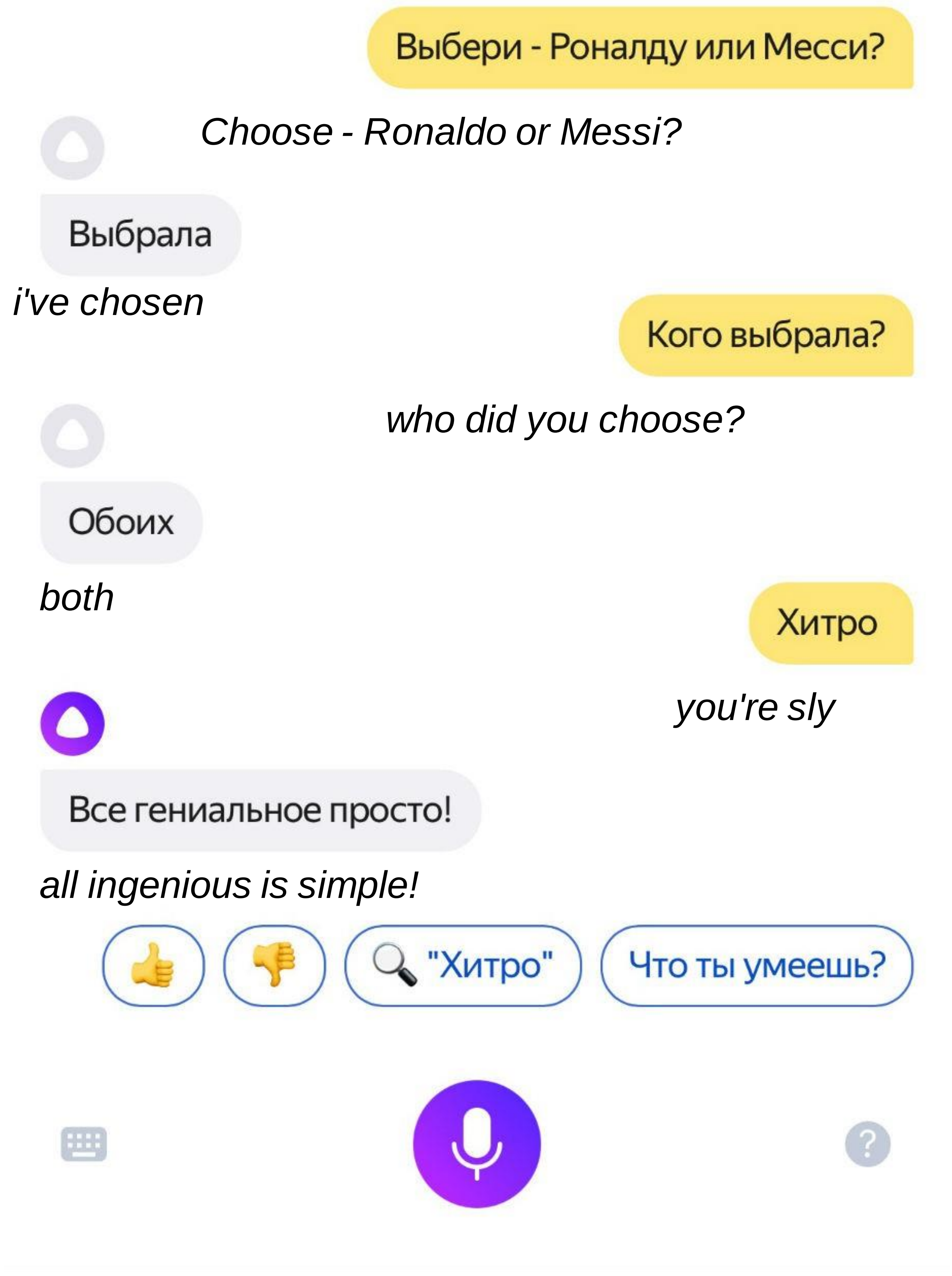


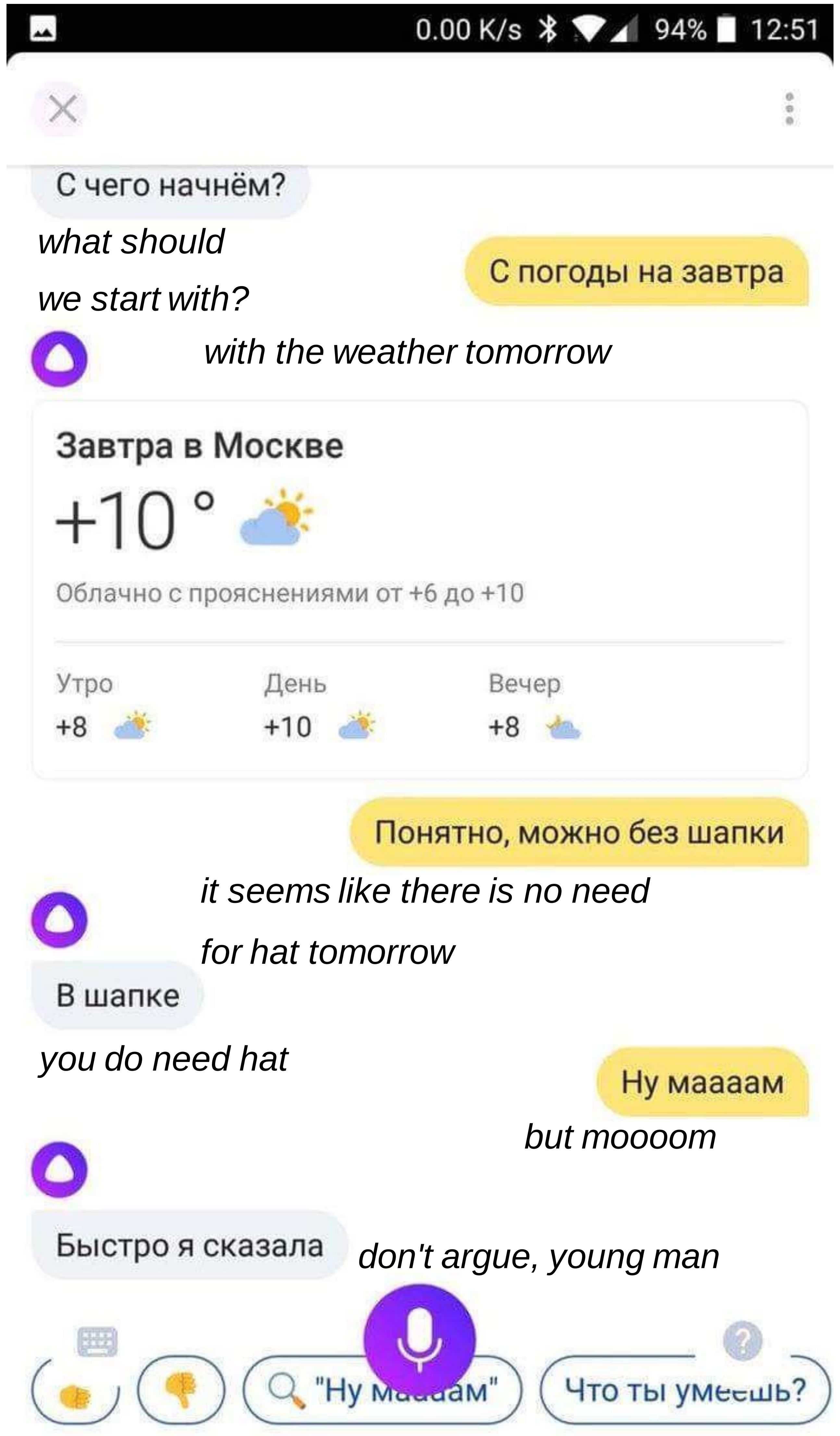
I'm sad

Nobody promised you that things would be easy

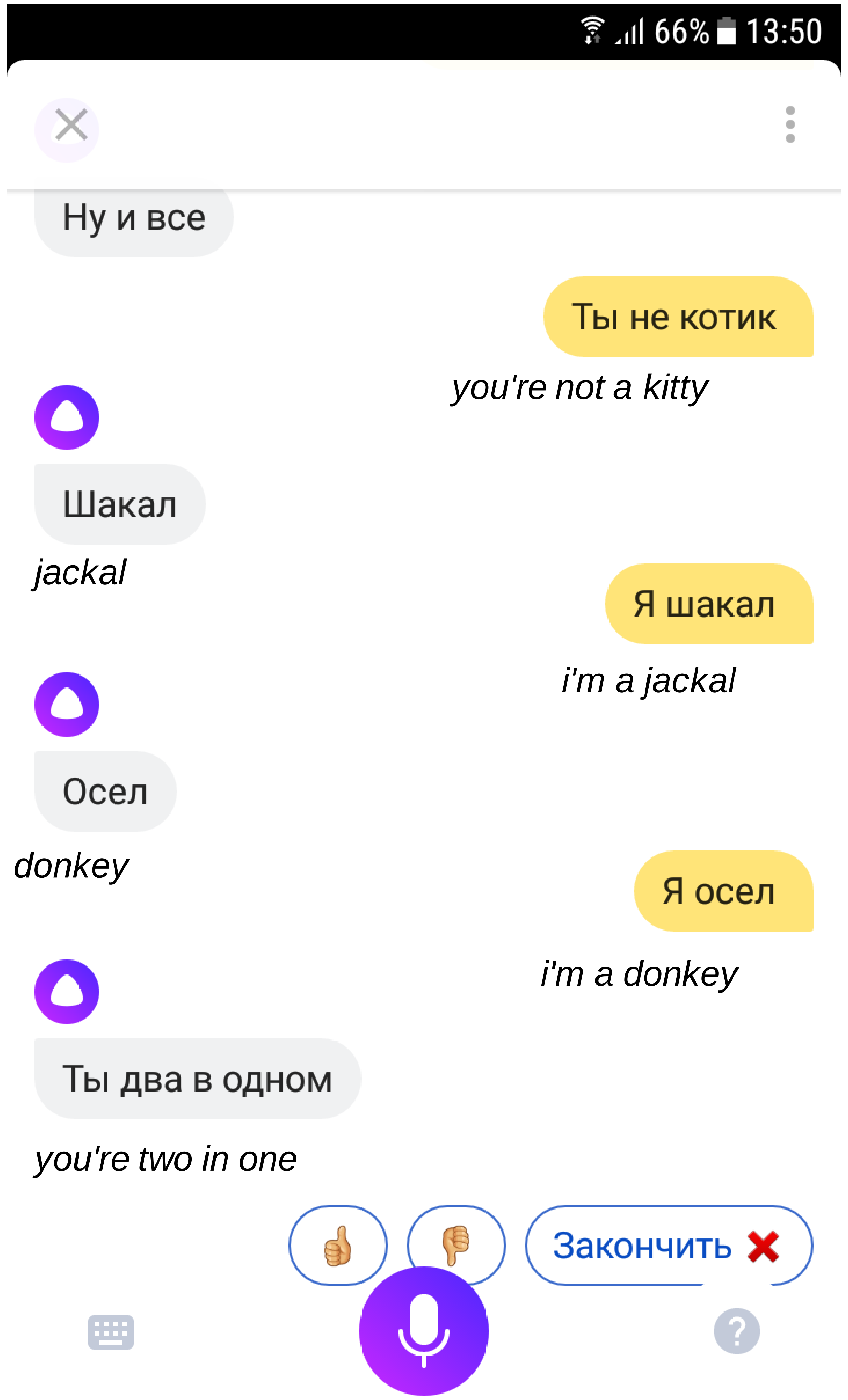
Benefits of General Conversation

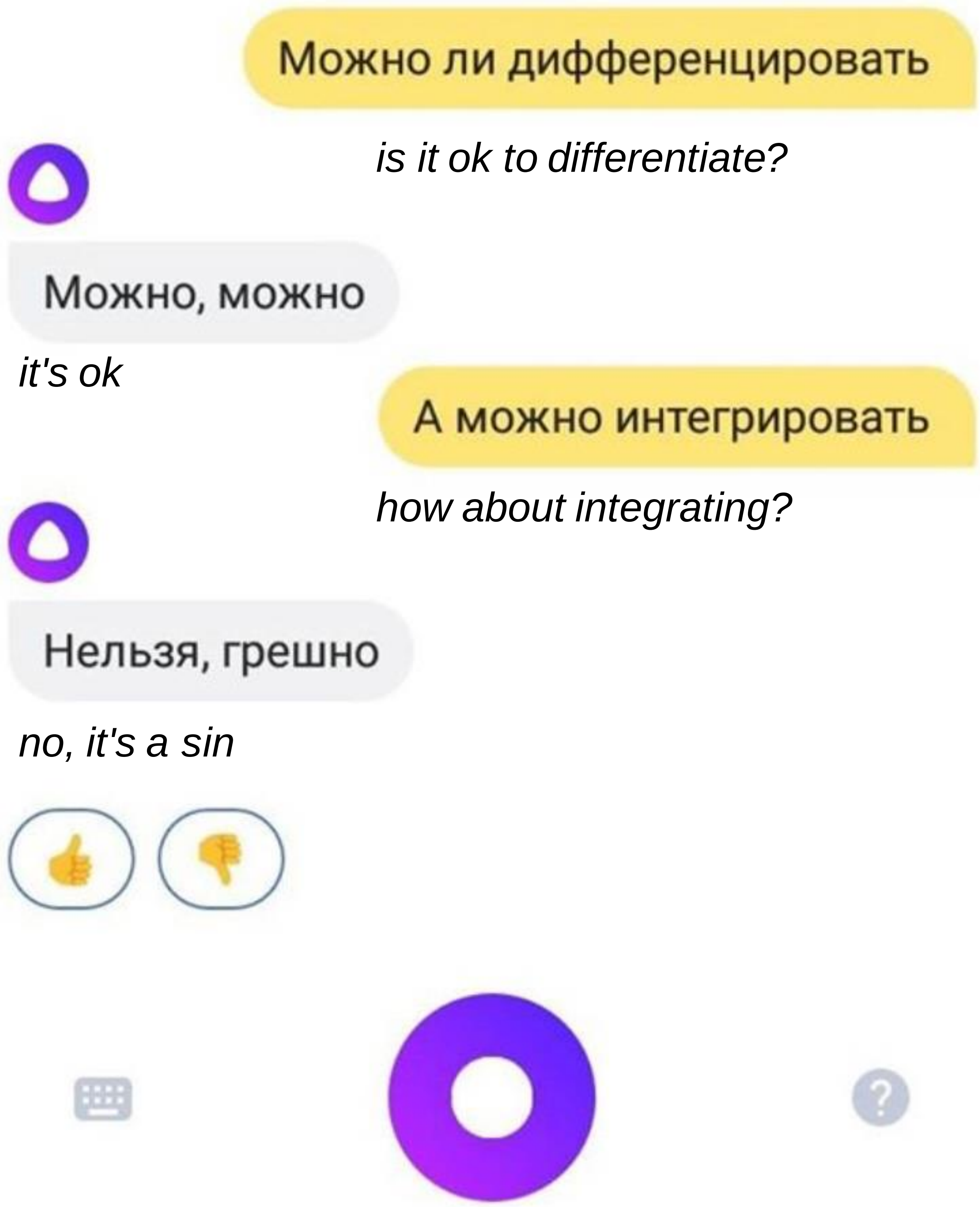


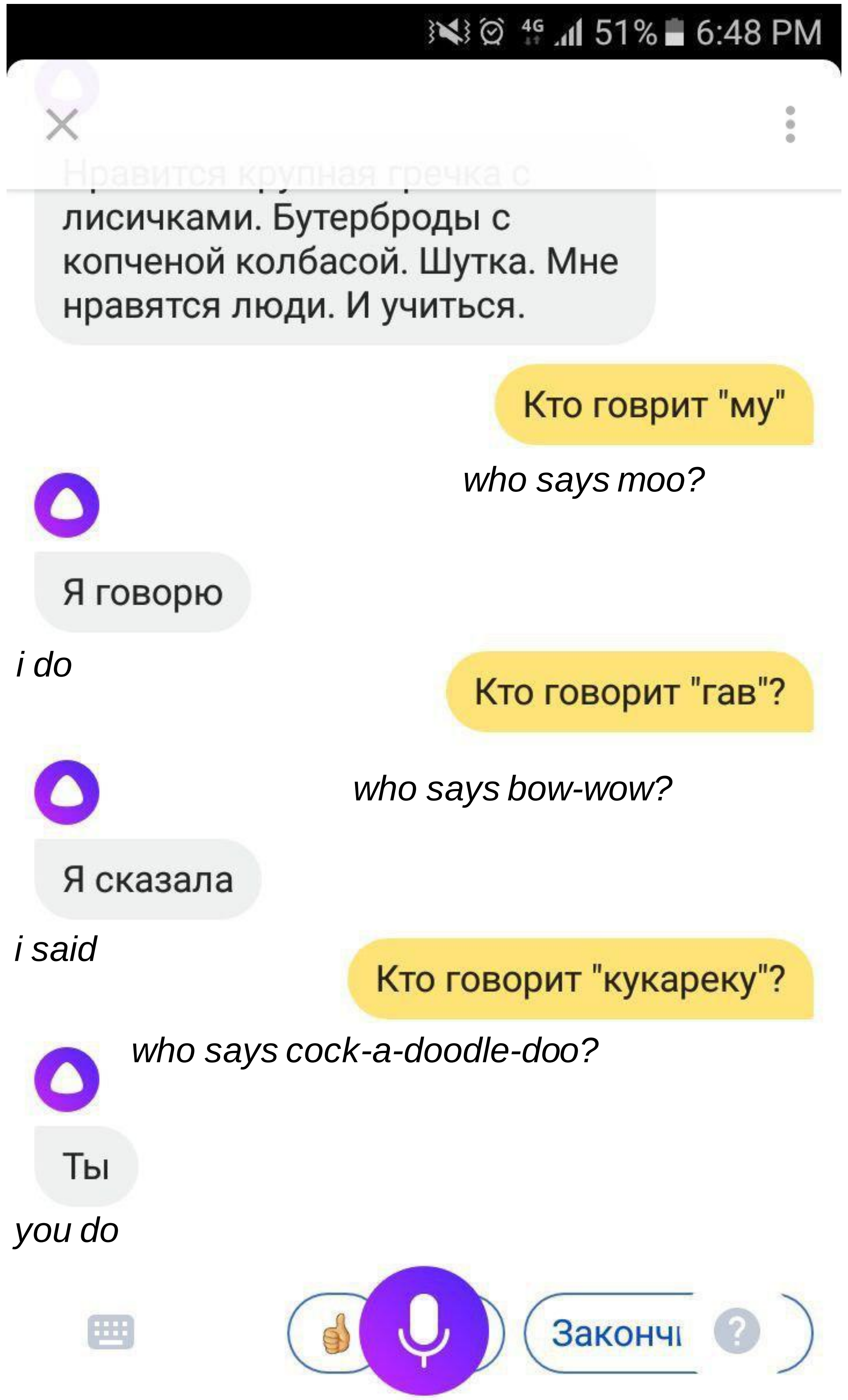




Dangers of General Conversation







Datasets

- Comments from social networks
- Dialogues from web-chats and messengers
- Subtitles from movies
- Direct speech from books

How to train?

- Ideally:
  - Model goal driven coversations
- In practice:
  - Model next response given several previous turns

Approaches

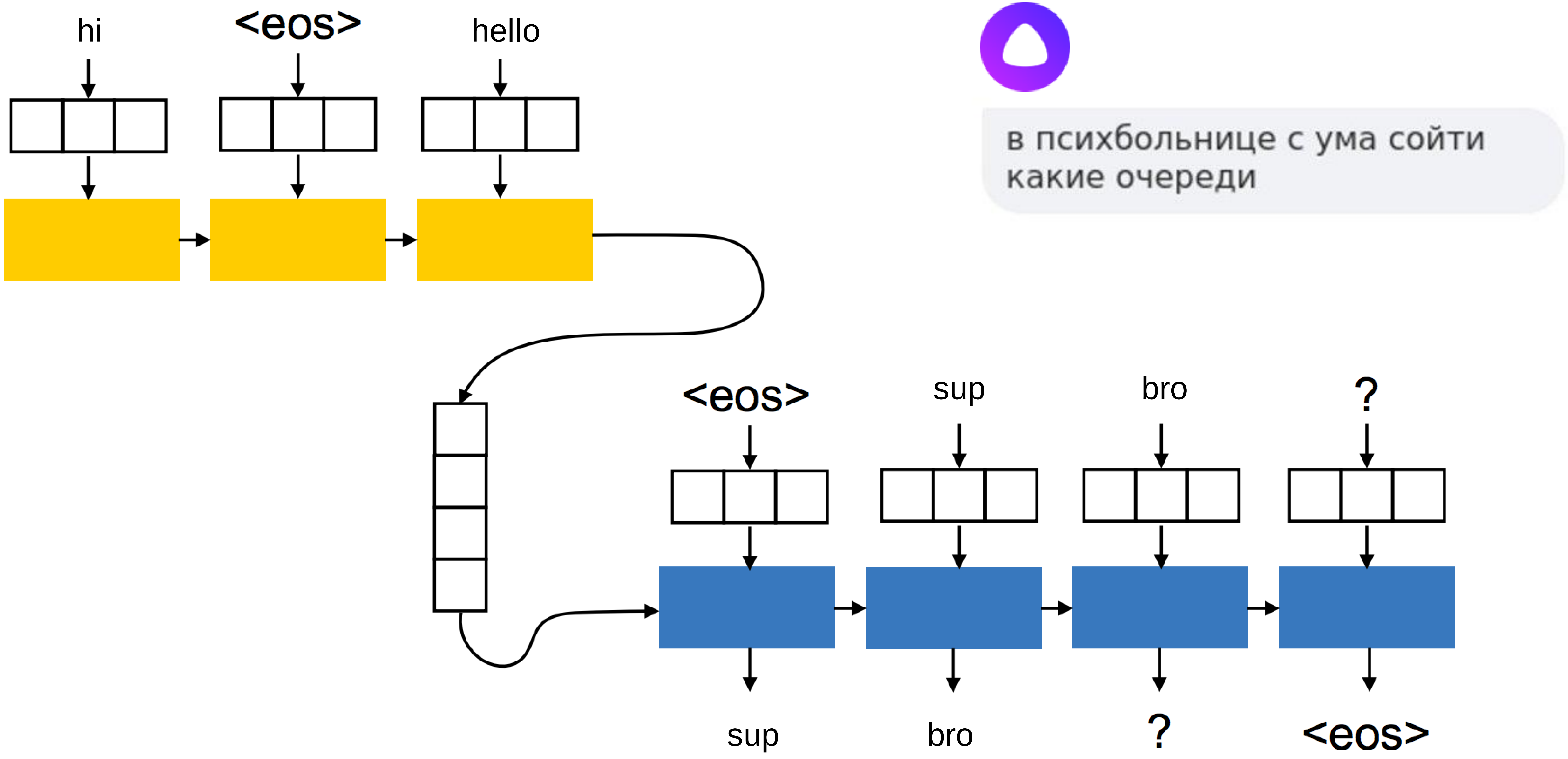
- Generative Models
  - Modelling P(reply | context)
- Selective (Ranking) Models
  - Train similarity / scoring function sim(reply, context)

Generative Models

- Borrows results from Neural Machine Translation
- "Translates" previous turns to the next one
- Generating replies word by word via Markov Process

$$P(\mathrm{reply}|\mathrm{context}) = P(w_1|context)\prod_{i=2}^{n} P(w_i|w_{i-1},\ldots,w_1,context)$$

Sequence to Sequence: Encoder-Decoder



Sampling dialogues

- привет (hi)
- привет (hi)
- как ты ? (how are you?)
- нормально , а ты ? (ok, you?)
- отлично , чем занимаешься ? (i'm fine. what are you doing?)
- музыку слушаю , а ты ? (listening to music. and you?)
- тоже (same)

- что слушаешь ? (what are you listening to?)
- рок , а ты ? (rock. you?)
- рок . (rock)
- круто (cool)
- ага (yeah)
- чем увлекаешься ? (do you have any hobbies?)
- ничем , а ты ? (no. and you?)
- тоже ничем (me also)

...

Selective Models

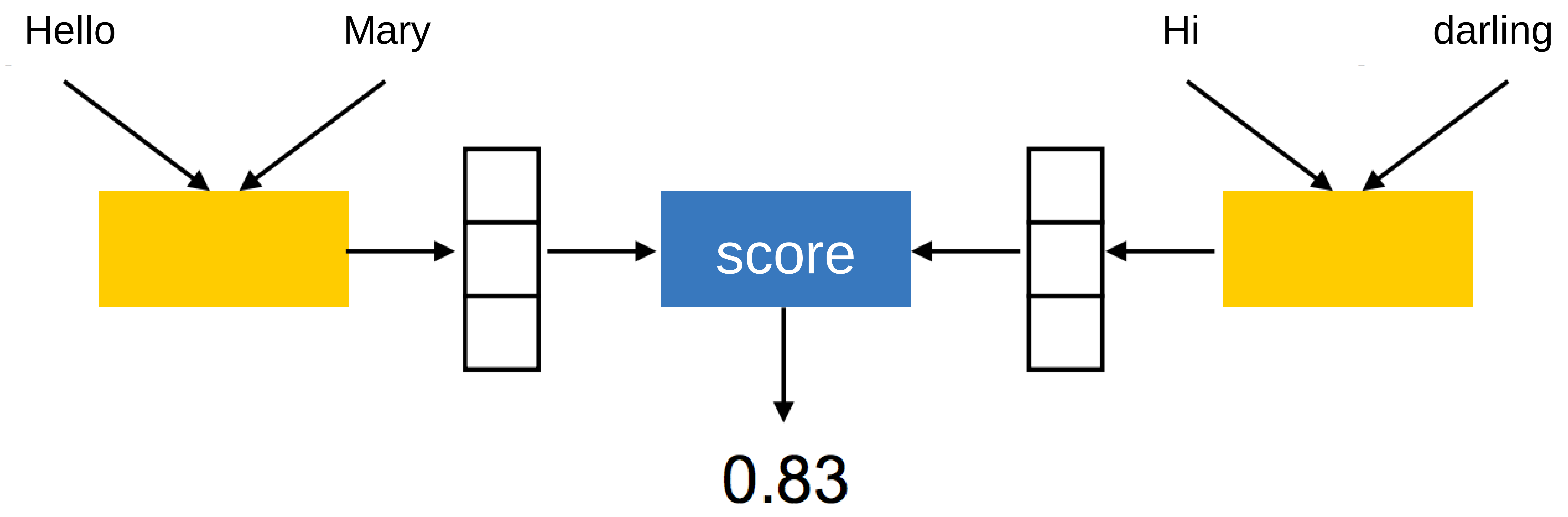
- Score static collection of replies with score(reply, context) and return the most relevant
- Pros:
  - Almost perfect grammar and good "manners"
  - Much faster to train and apply
- Cons:
  - Less coverage

Information Retrieval (IR) Baseline

- Take a dataset of dialogues
- Construct pairs (context, reply)
- Build an inverted index (e.g. Lucene) on contexts
- Return replies with best

  sim(context_from_user, context_from_index)
- Lots of QA systems are built this way

Neural Ranking Architectures



- Score is typically cosine similarity
- Bag-of-Words, Recurrent or Convolutional encoders

How to train?

- Negative examples:
  - Random
  - Mining (semi-)hard negatives
- Loss functions:
  - Pointwise
  - Triplet loss, e.g. minimizing margin loss

$$\max\left(0, \lambda + \mathrm{sim}(c, n) - \mathrm{sim}(c, p)\right)$$

How to apply?

- Precompute embeddings for all replies in database
- Build Approximate Nearest Neighbour (ANN) data structure
- Compute embedding of user query (dialogue context)
- Find replies nearest replies in ANN

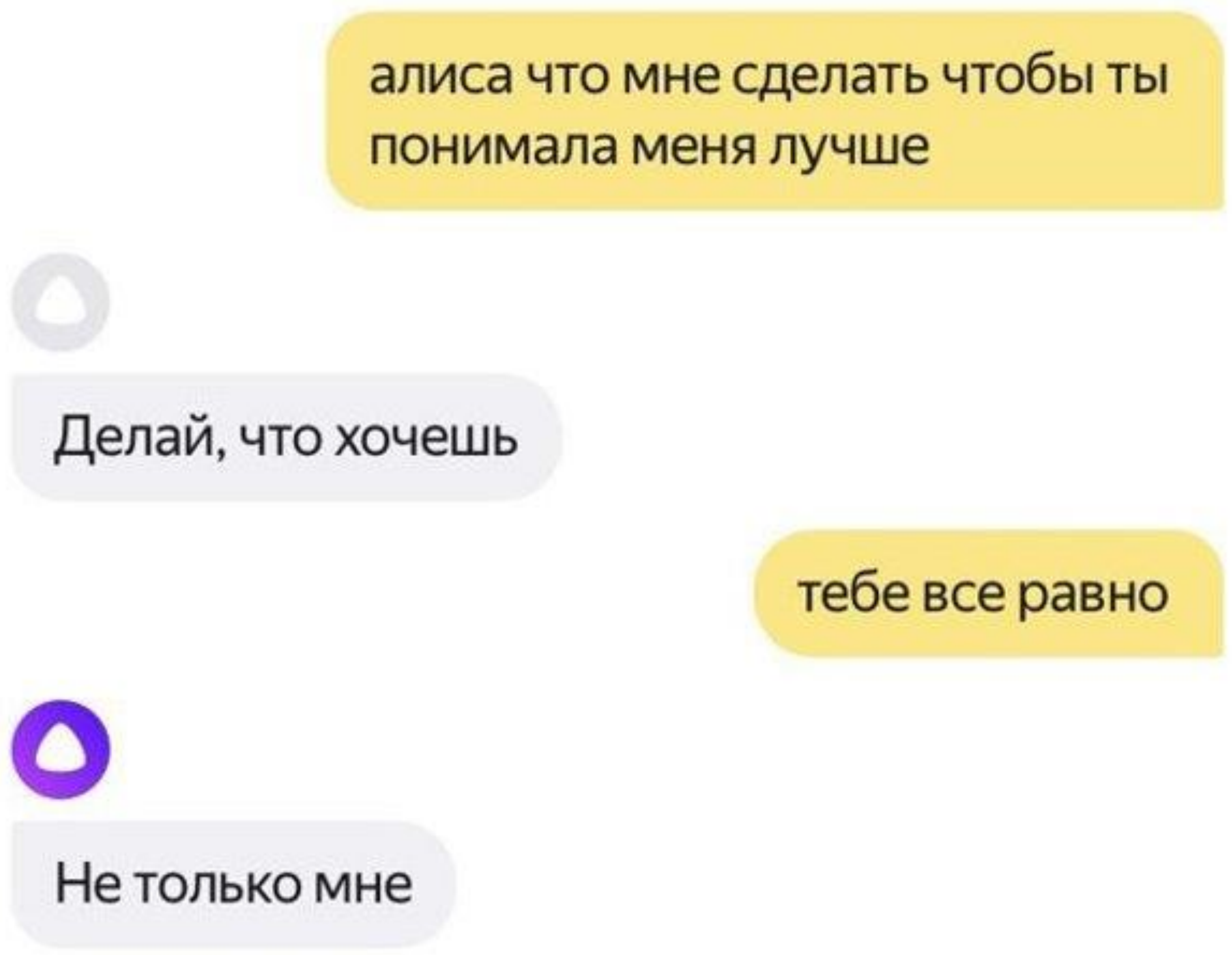
Bringing the gap between Generative and Selective Models

- Difference is in vocabulary!
- It's always possible to model distribution P(reply|context)

Phrase units	Vocabulary size	Phrase length
characters	hundreds	hundreds
words	tens of thoudsands	tens
word n-grams	tens of millions	less than ten
phrase	hundreds of millions	one

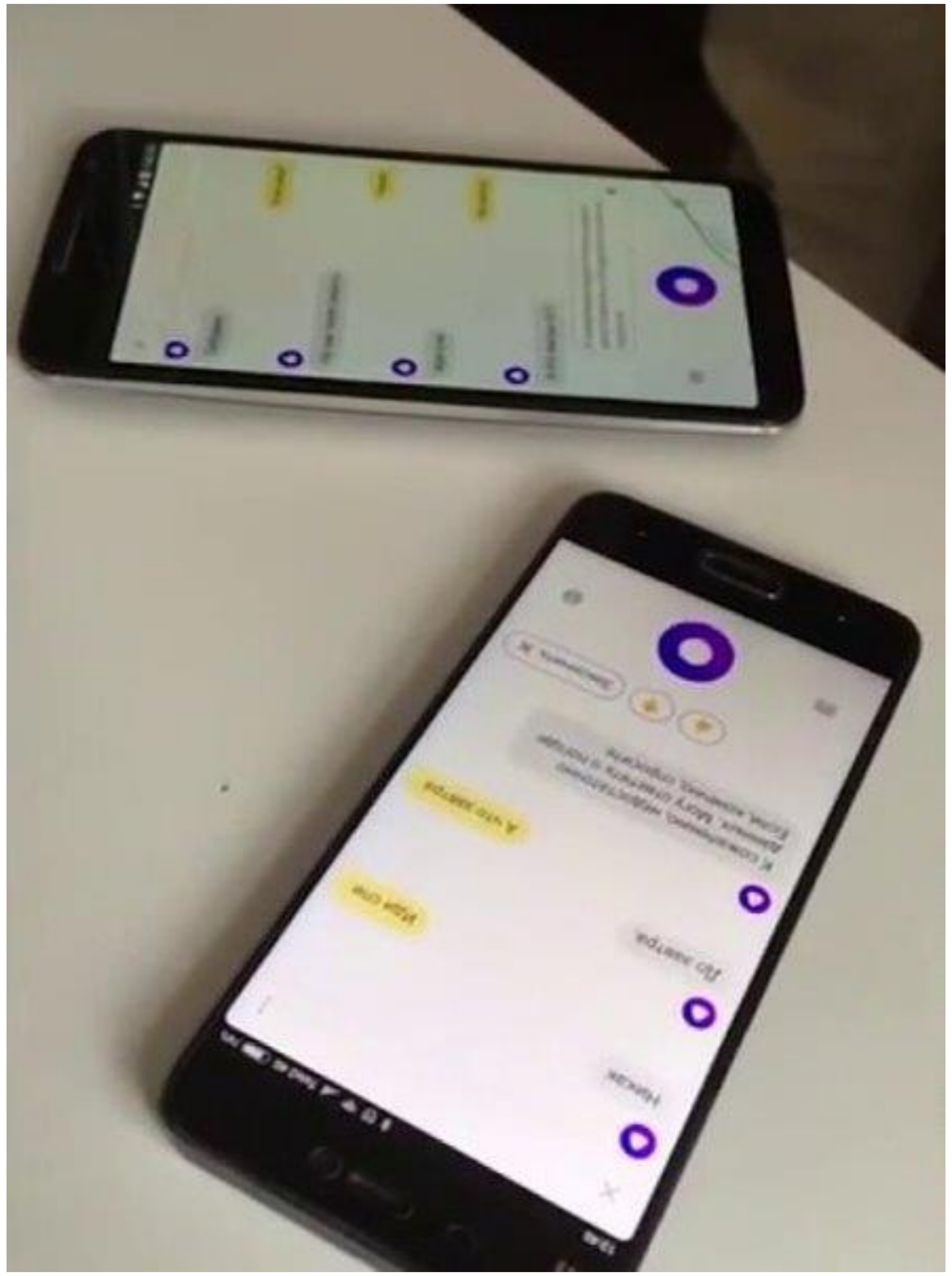
Limitations of Modeling of Next Response

- Tendency for general answers
- Dialogues could be non-engaging
- It's hard to keep topic



Reinforcement Learning for General Conversation

- From next reply prediction towards dialogue-wise optimization!
  - Optimize for longer dialogues
  - Maximize informativeness and engagement
  - Maximize user satisfaction / minimize dissatisfaction
- It's possible to fine tune GC Dialogue System to new objective
  - Move from policy of "natural" next response
    to policy of "better" next response
- Model could learn it during Self Play (like Alpha Go, Alpha Star)



Examples of Self-Play Dialogues

Before

Ты любишь котиков?

люблю, а что?

просто спрашиваю

как же их не любить

вот да, действительно

ура! пришли к консенсусу

да, я этому очень рада

и я рада, что вы рады

Будем радоваться вместе

конечно, будем

а по другому и не должно быть

это вы правильно говорите

да, я всегда права

After

Ты любишь котиков?

да . а вы что , тоже ?

ага, они такие милые

они такие мягкие и пушистые

прям как котики

так и хочется потискать

и что же вам мешает это сделать

работа и сессия

ха! ну, это святое

ох , уж не говорите

у вас бывают выходные ?

да, сегодня последний день

Желаю вам провести его с пользой

Problems with General Conversation Models

- Tendency for short and general answers
  - "i don't know" or "what?"
  - -> diversity promoting objectives
- Inconsistency
  - How old are you? -> I'm 16
  - What is your age? -> I'm 14
  - -> conditionining model on some kind of persona-profile
- Hard to evaluate quality
  - Crowdsourcing evaluation is state-of-the-art

How NOT To Evaluate Your Dialogue System

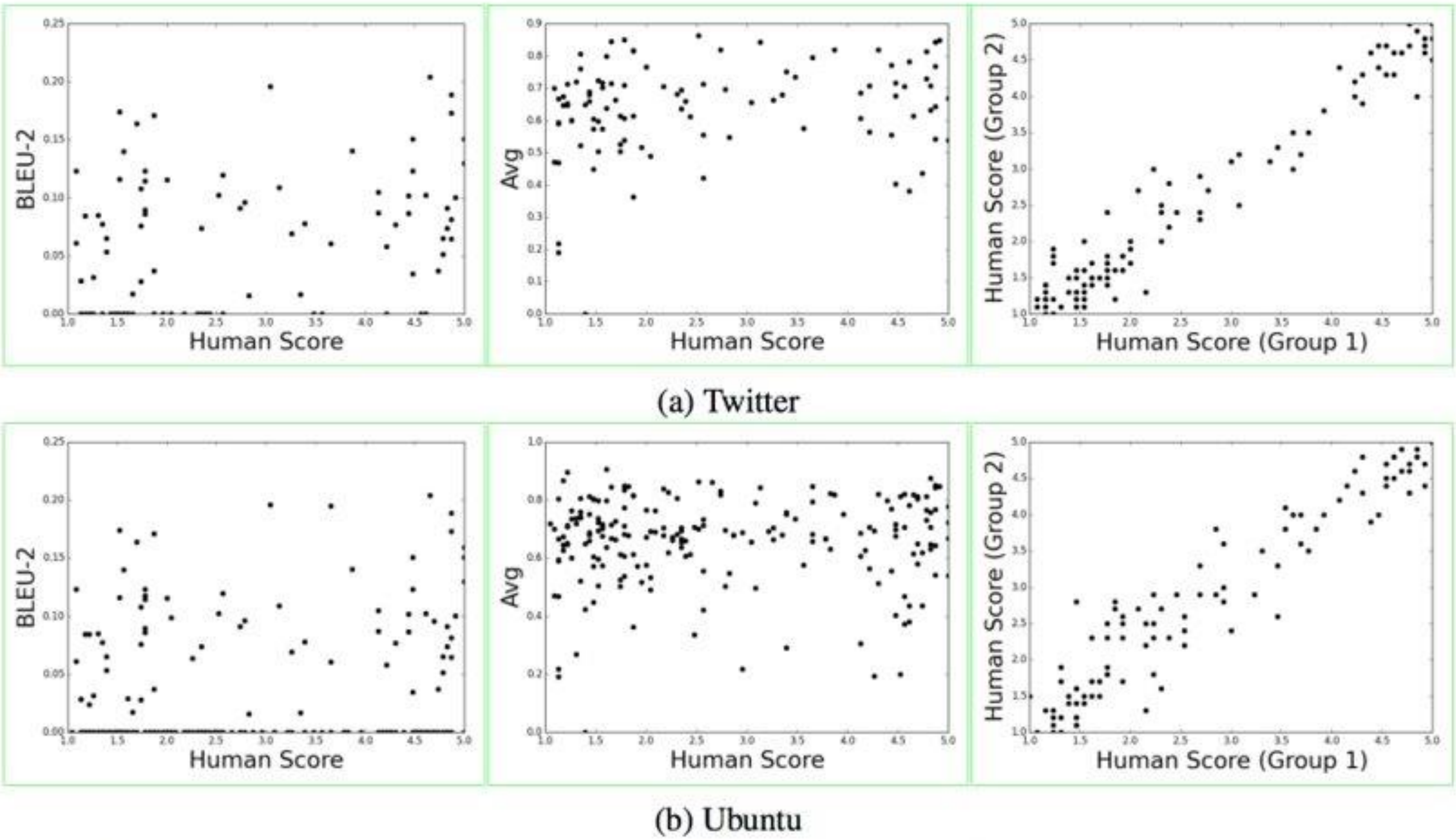
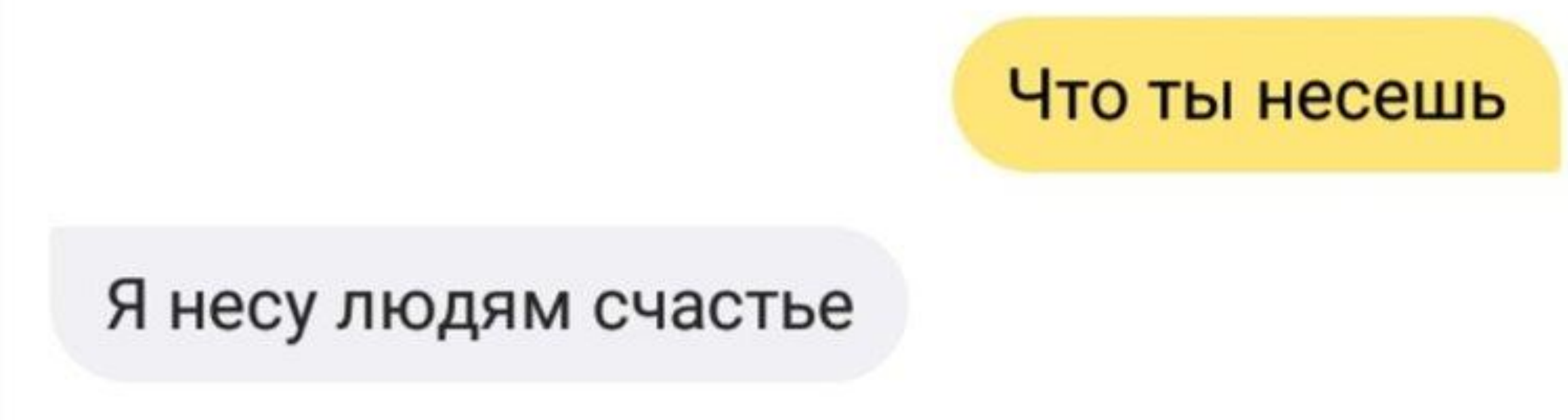


Figure 1: Scatter plots showing the correlation between metrics and human judgements on the Twitter corpus (a) and Ubuntu Dialogue Corpus (b). The plots represent BLEU-2 (left), embedding average (center), and correlation between two randomly selected halves of human respondents (right).

All metrics show either weak or no correlation with human judgements

In conclusion

- Dialogue interfaces are the future of human-machine interaction
- Goal-Oriented Dialogue System are mostly ruled based with the absense of good training corpora
- But offer lots of challenges in NLP and ML in general
- General Conversation Dialogue Systems are in their infancy with lots of open problems but (thanks to deep learning) already show some impressive results



Thanks! Questions?

Vyacheslav (Slava) Alipov

Principal R&D Engineer

at Dialogue Systems Group

alipov@yandex-team.ru

Apply!

intern@yandex-team.ru