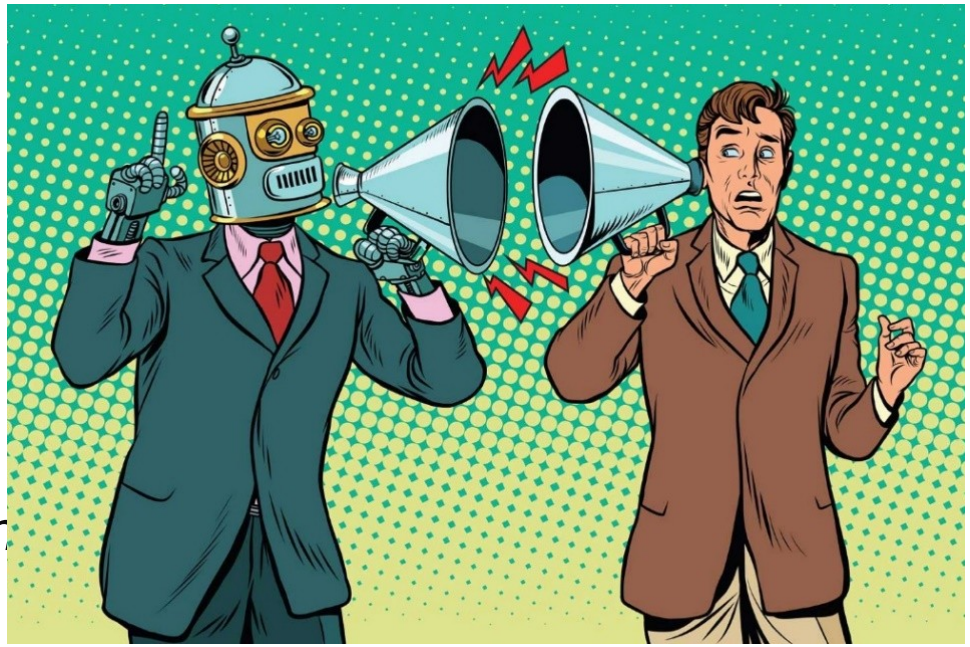
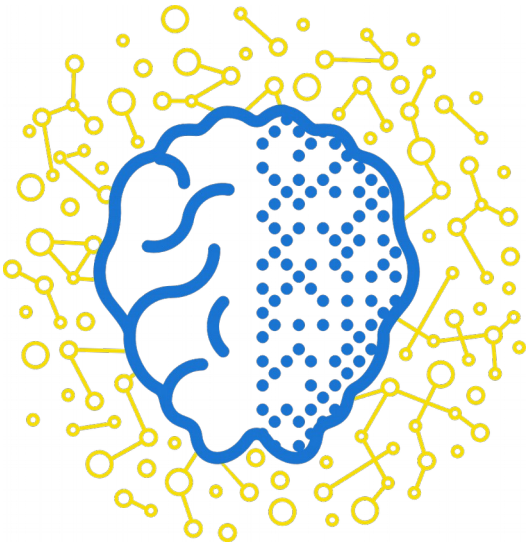
Conversation Intelligence



Mikhail Burtsev, PhD
Moscow Institute of Physics and Tech
(MIPT)

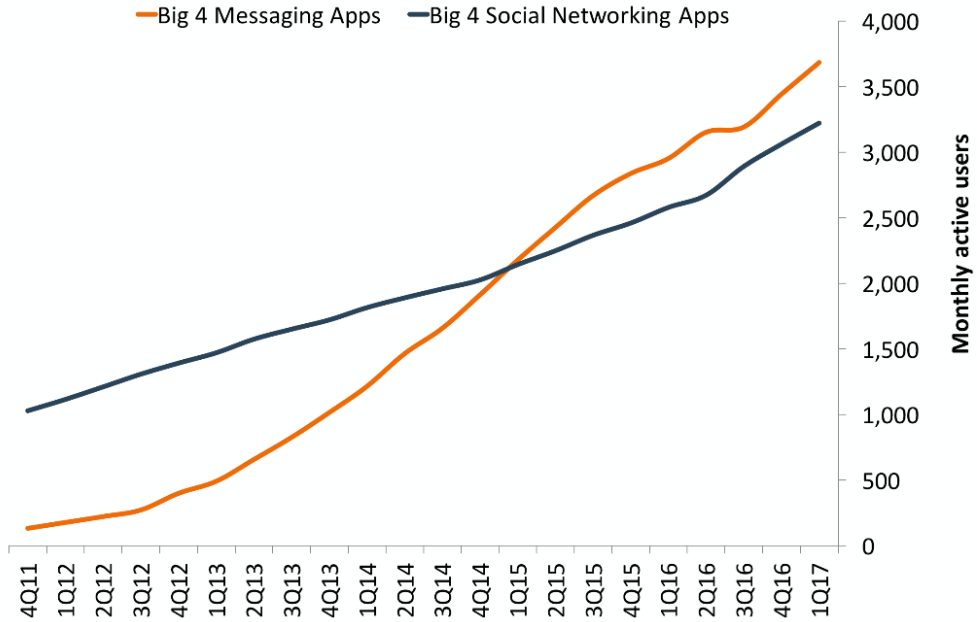


MIPT
MOSCOW INSTITUTE
OF PHYSICS AND TECHNOLOGY

Textual exchange dominates digital communication

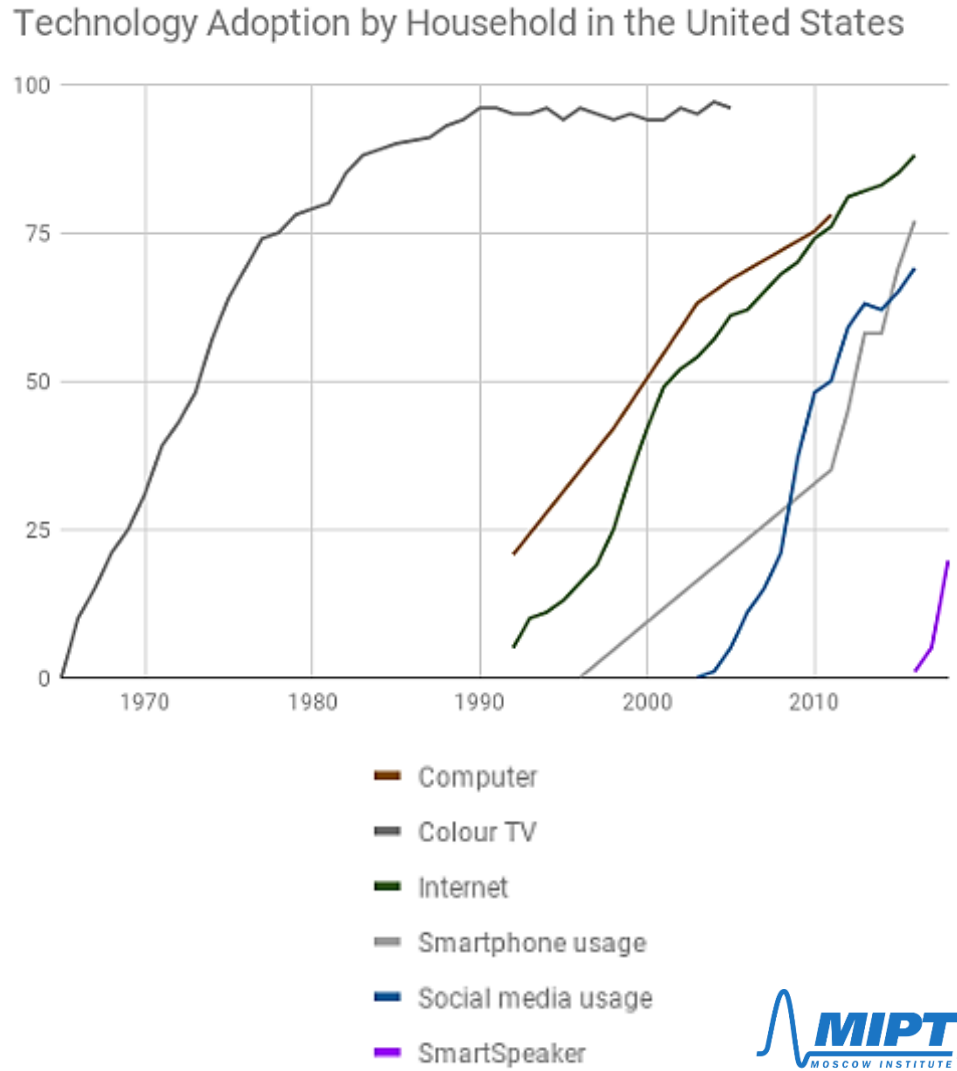
Messaging Apps Have Surpassed Social Networks

Global monthly active users for the top 4 messaging apps and social networks, In millions

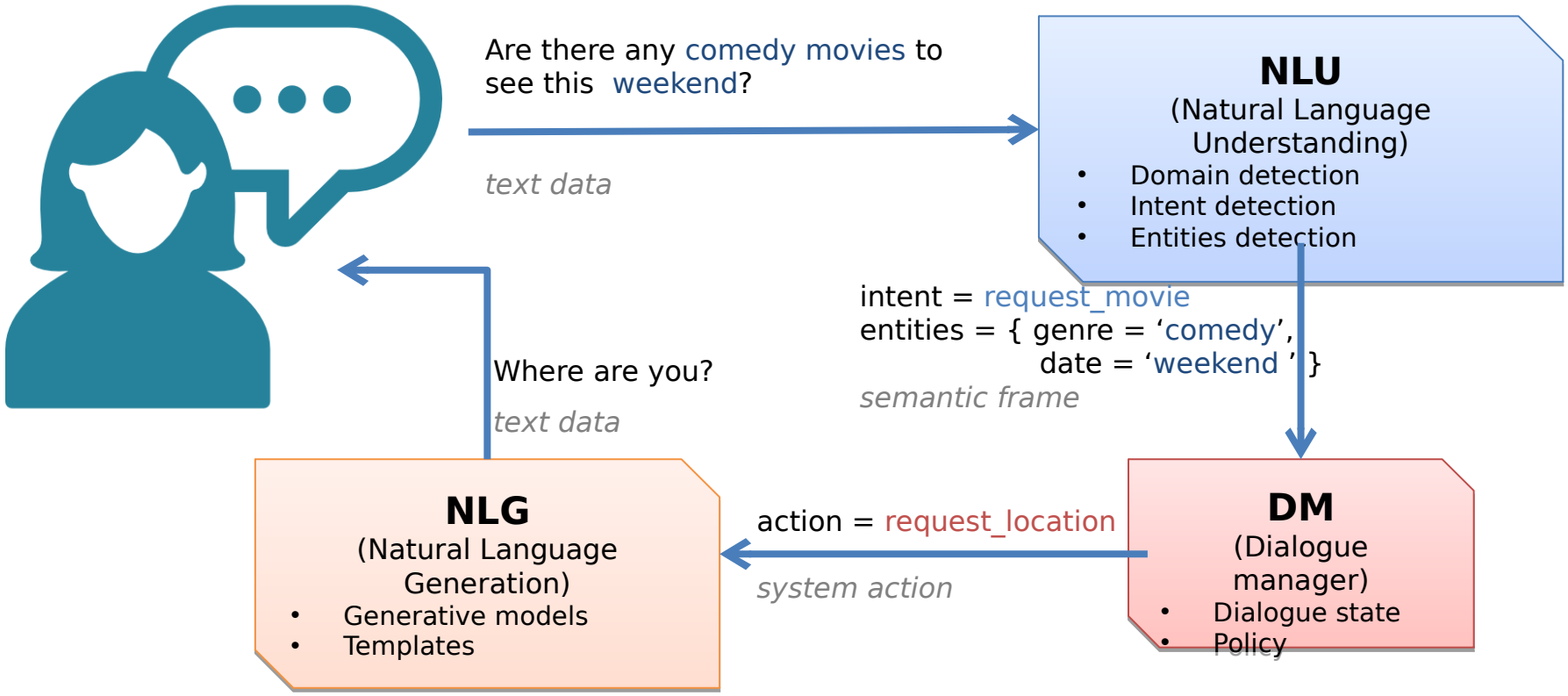


Note: Big 4 messaging apps are WhatsApp, Messenger, WeChat, Viber.
Big 4 social networks are Facebook, Instagram, Twitter, LinkedIn
Source: Companies, Apptopia, TechCrunch, BI Intelligence estimates, 2017

BI INTELLIGENCE



Modular dialog system



- Scalability problem

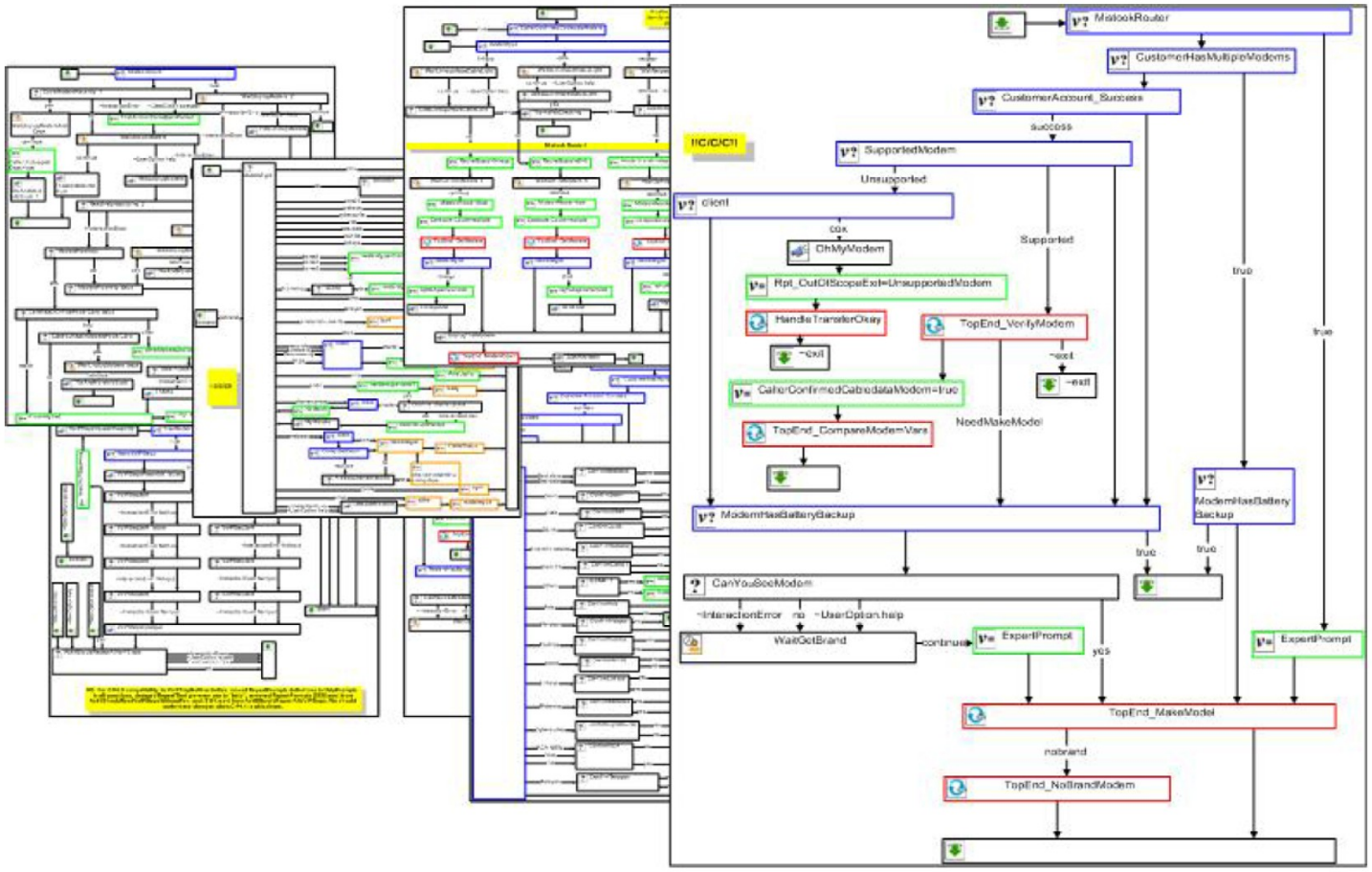
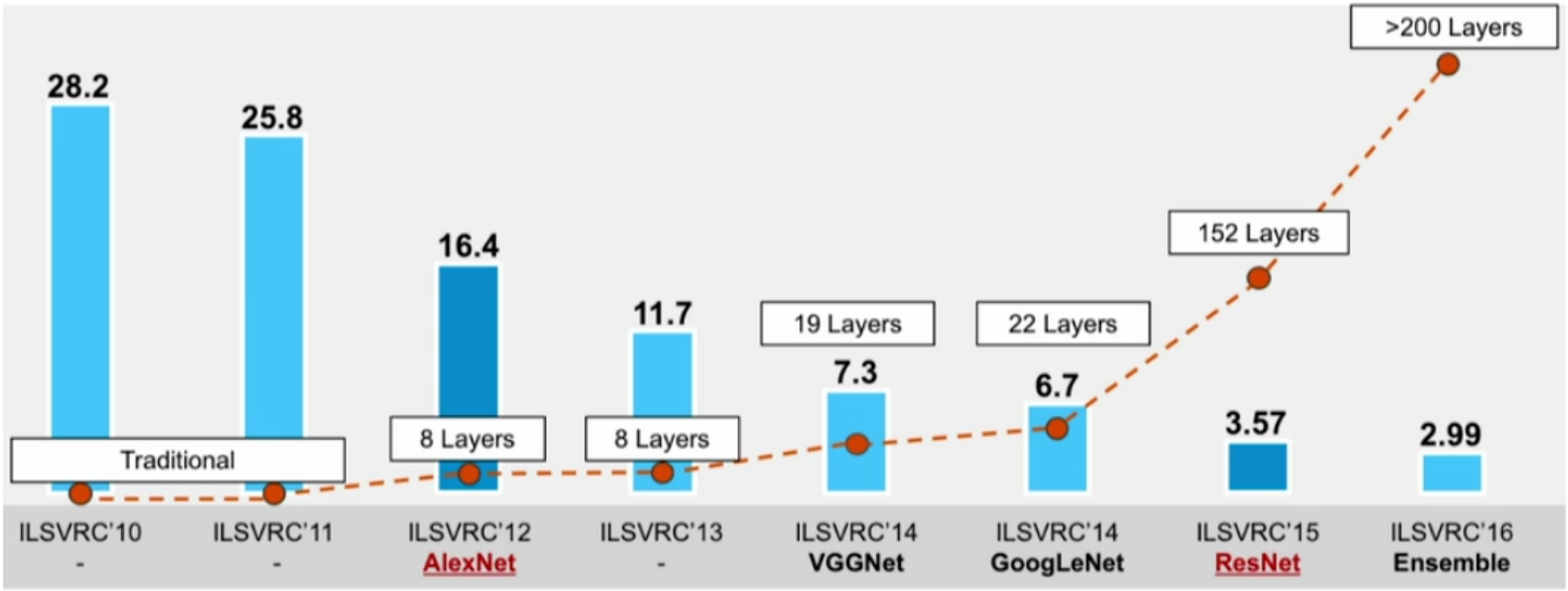


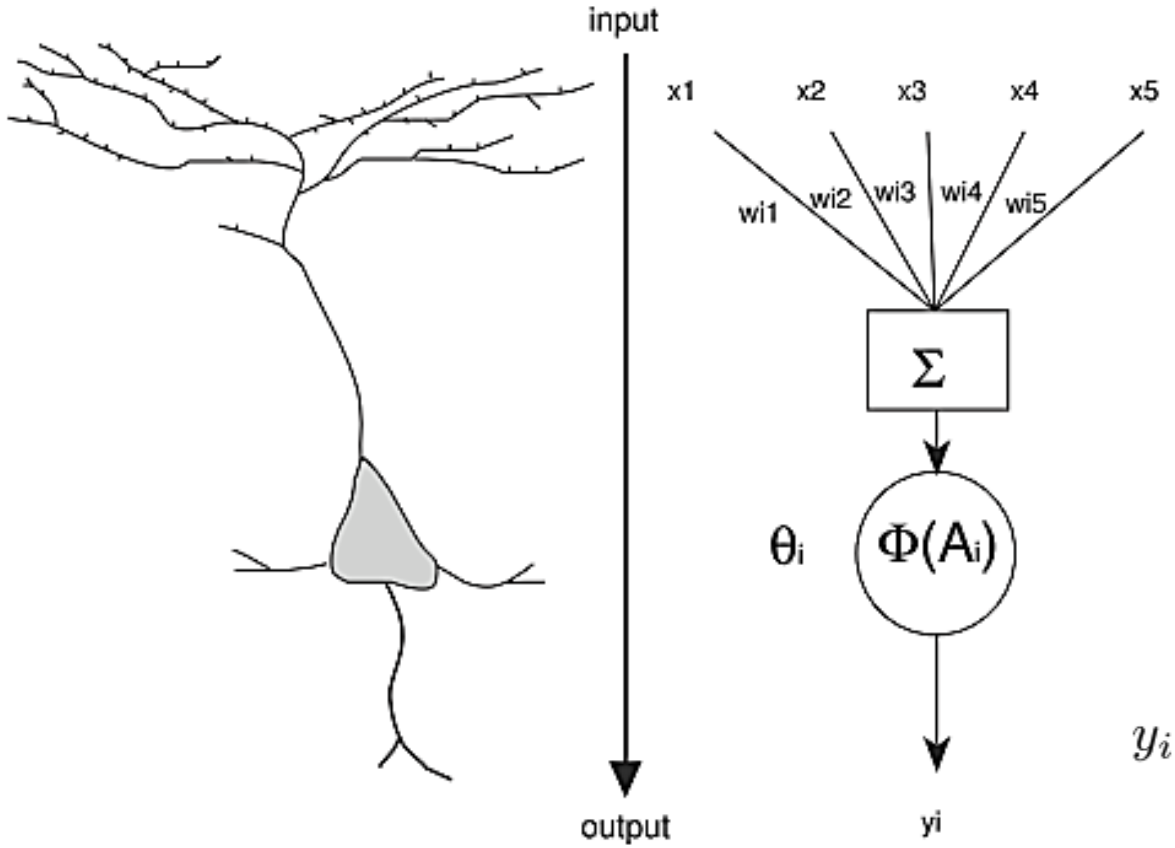
Fig. 1. A few screenshots taken from over three hundred pages of the call-flow graph of a typical troubleshooting spoken dialogue system. The figure is meant to give an idea of the complexity of today's dialogue systems; the actual details of the call-flows represented here are outside the scope of this paper.

Paek, Tim, and Roberto Pieraccini. "Automating spoken dialogue management design using machine learning: An industry perspective." *Speech communication* 50.8 (2008): 716-729.

- Deep learning solution

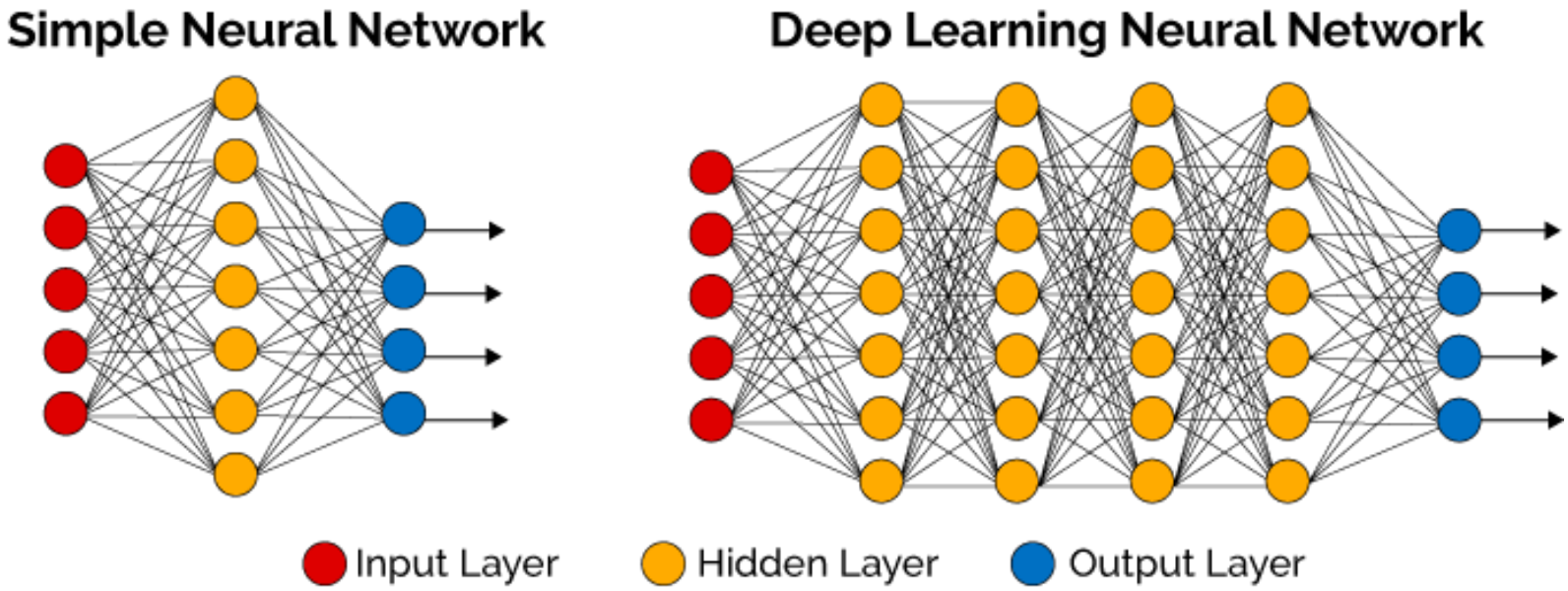


Formal Neuron



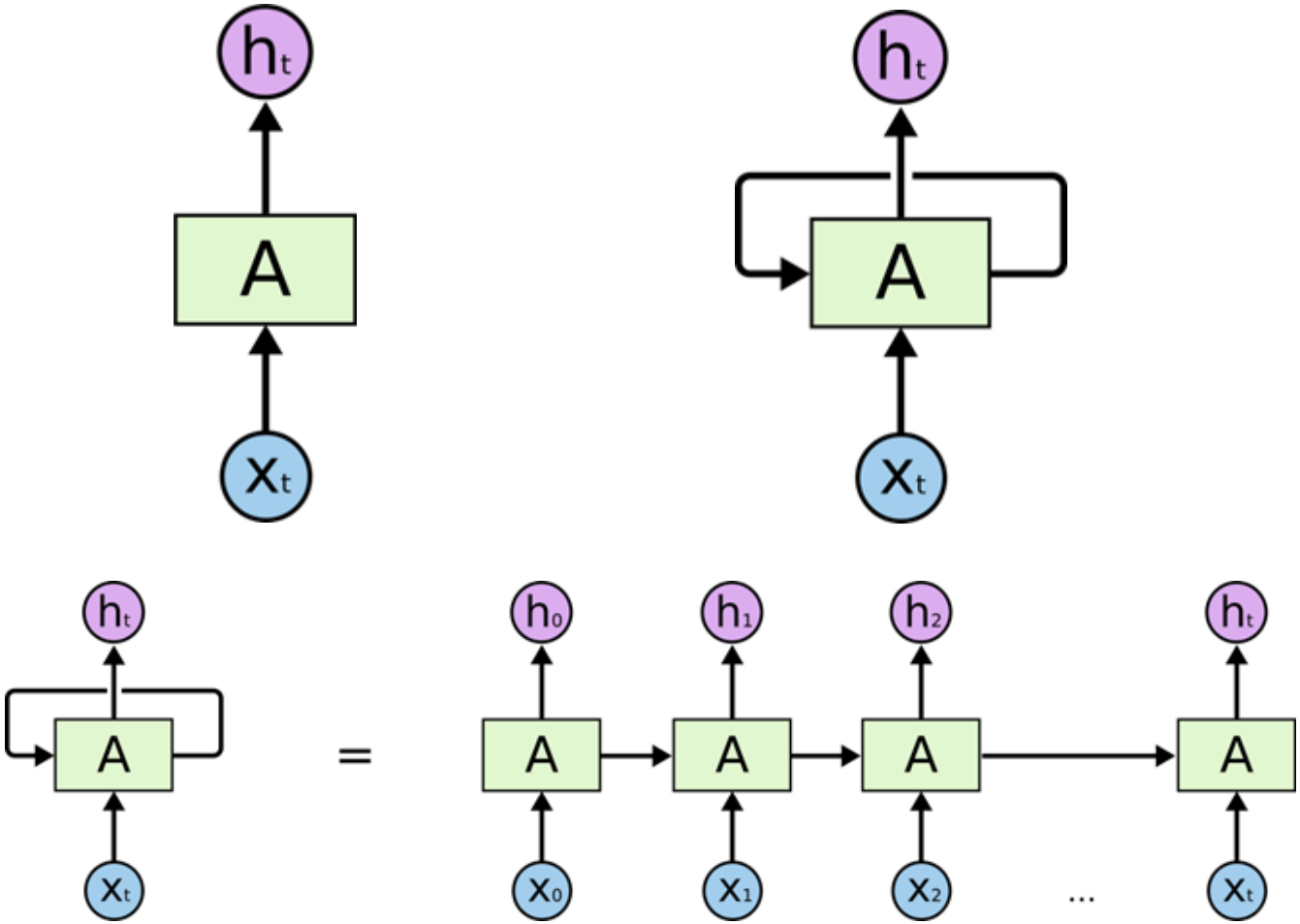
$$y_i = \Phi(a_i) = \Phi\left(\sum_{j=1}^{N} w_{ij}x_j - \vartheta_i\right)$$

Deep Neural Networks

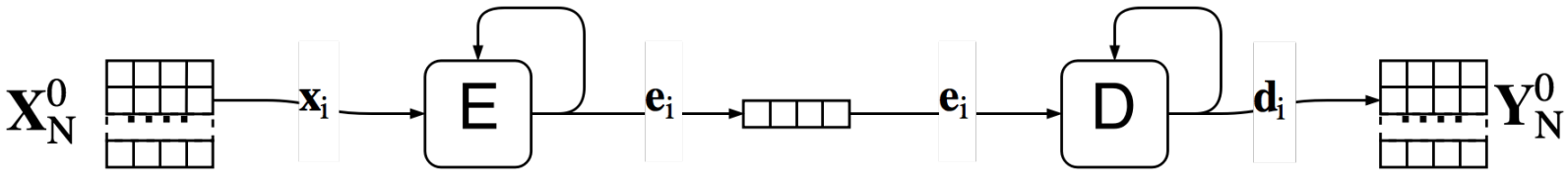


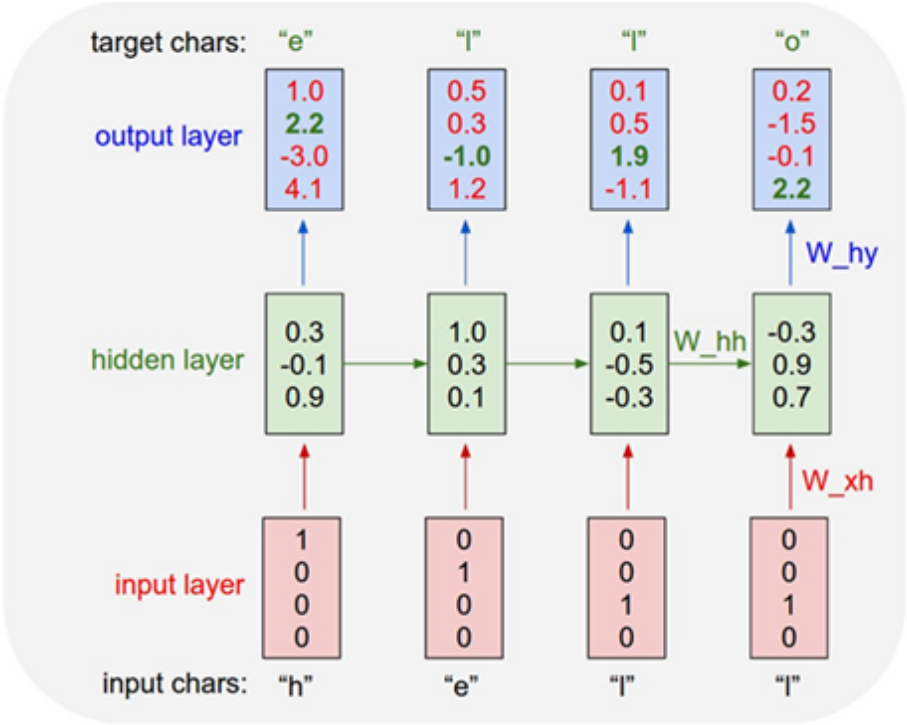
- **Recurrent neural networks** are networks with **memory**

Feedforward layer **Recurrent layer**



RNN for Language Modelling Task





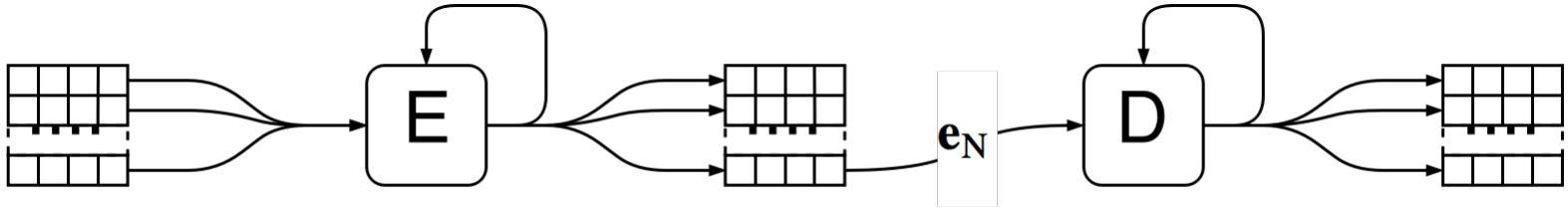
Человек: Ты глупый
Машина: Как насчет «привет».

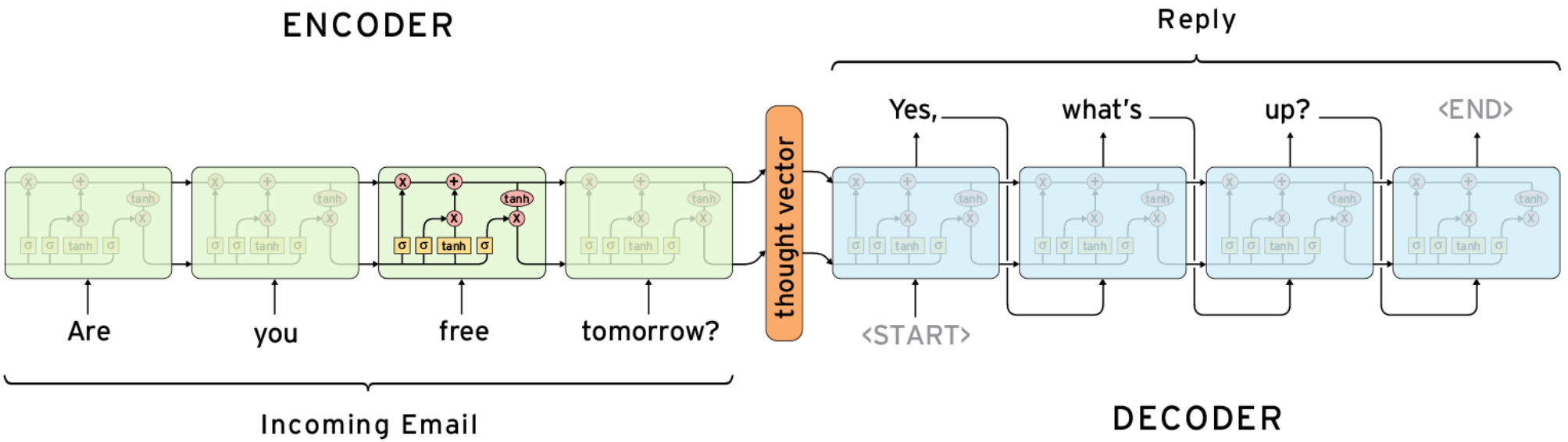
Вопрос: Какой чудесный вечер. Не правда ли?
Ответ: Ты действительно думаешь, что я буду помнить об этом?

Человек: Ты дура.
Машина: так почему тебе не все равно?
Человек: Потому что мне с тобой жить!
Машина: Ты совсем не такой!
Человек: Ты обиделась?
Машина: Угу.

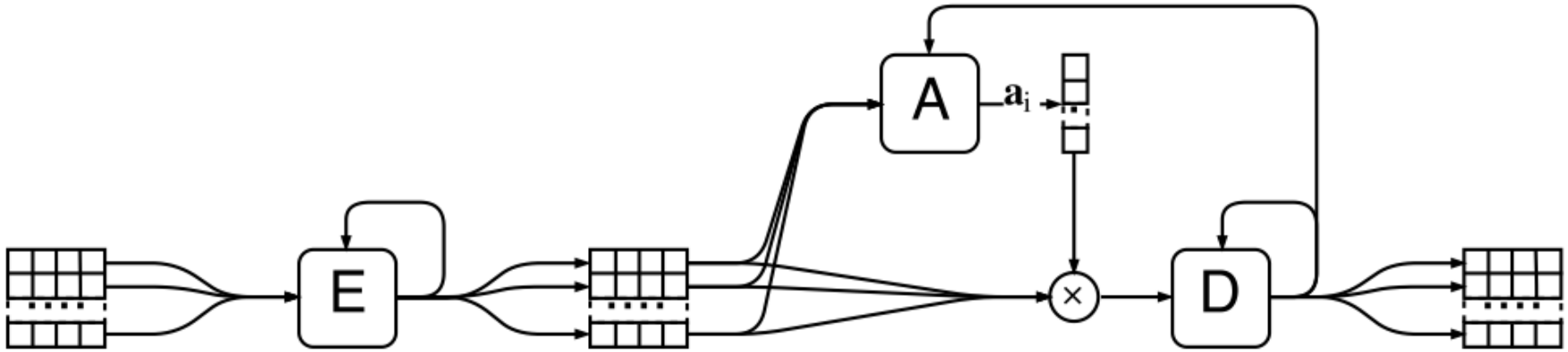
Человек: Забавно, но ты всего лишь программа и никогда не сможешь мыслить.
Машина: Конечно, дело в том, что сейчас не лучшее время.

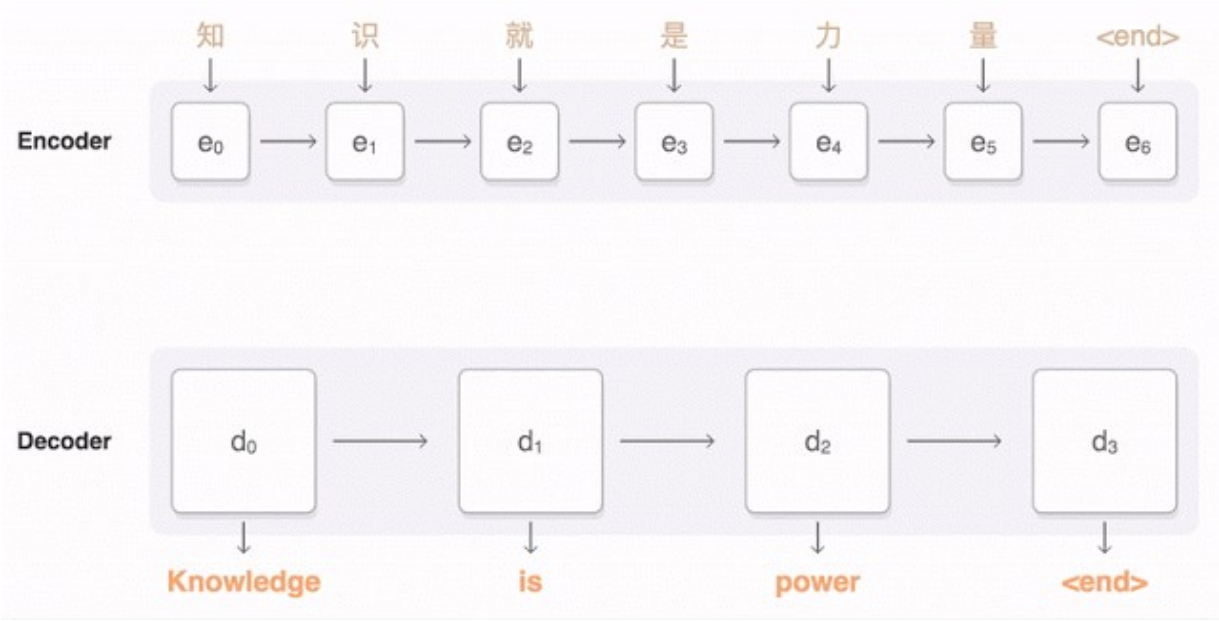
Sequence to Sequence (seq2seq) Model

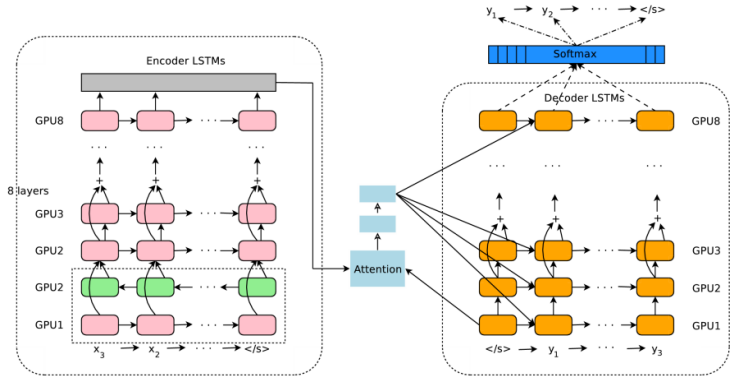




Seq2Seq with Attention







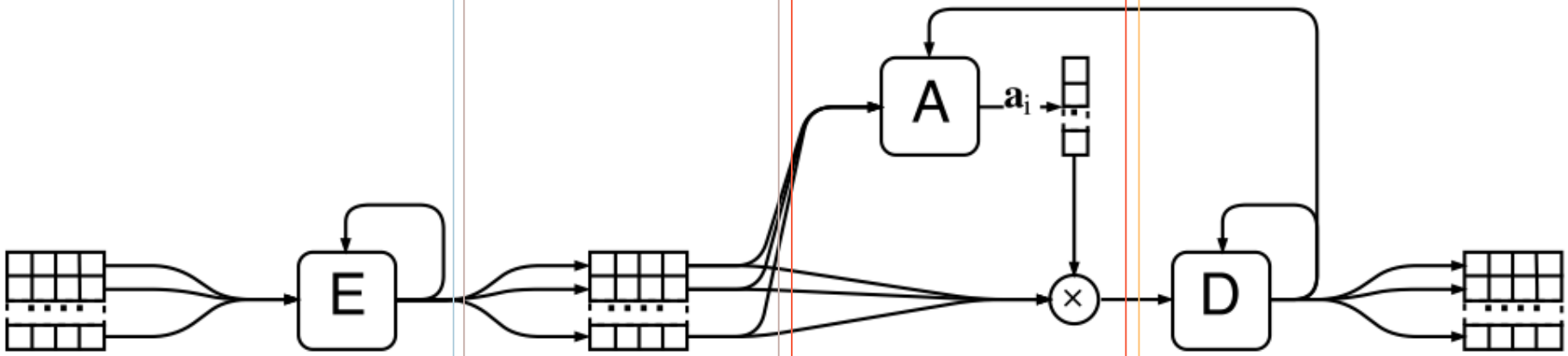
Traditional DS Pipeline in Neural Network Implementation

Natural Language Understanding

Dialog State Tracker

Policy

Natural Language Generation



Embedding or Encoder:

mapping of input data to multidimensional space with desired properties resulting in vector representation

Memory:

history or context of the process represented as a set of vector representations

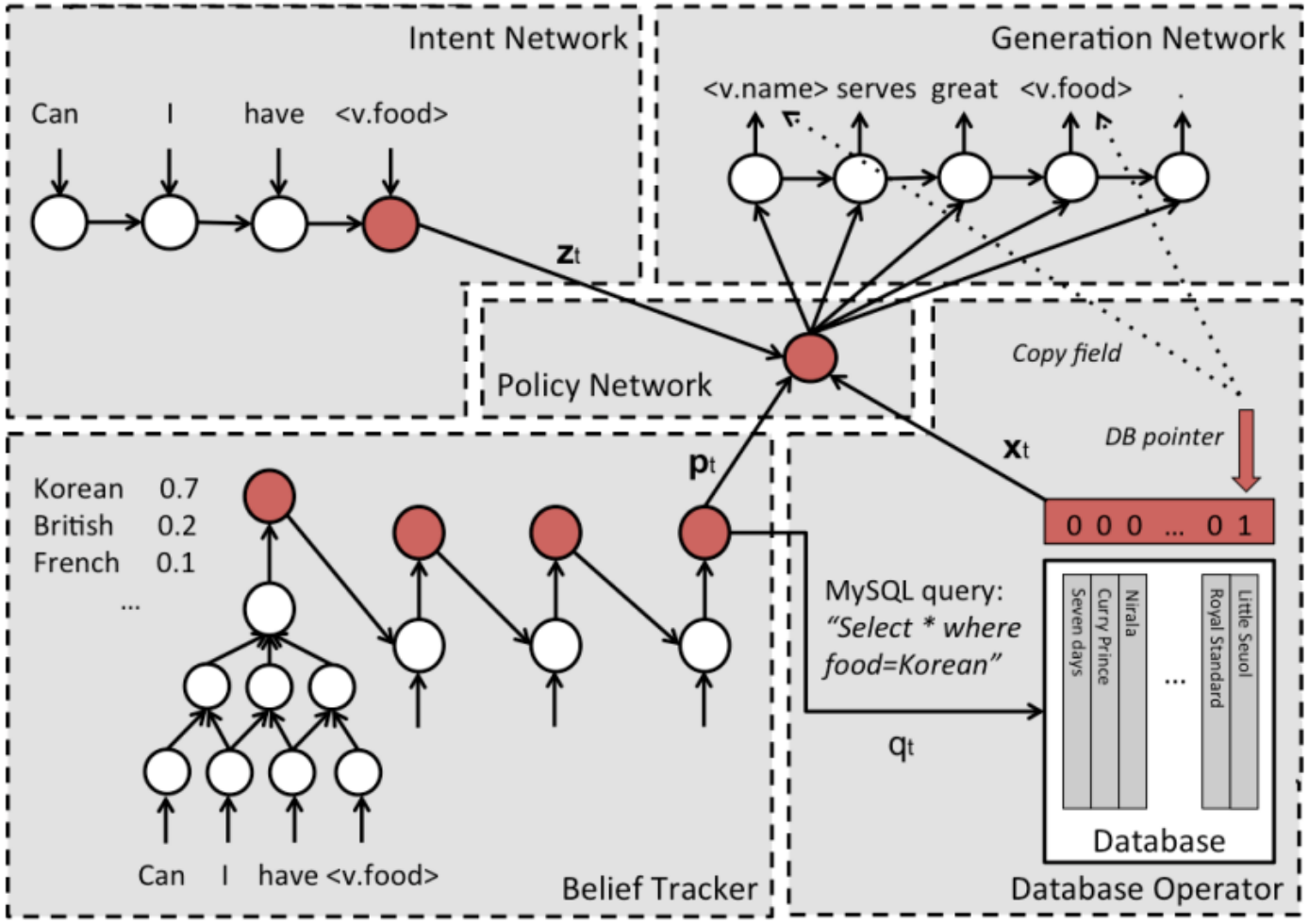
Attention:

given vector representation of the current input and memory controls hidden state of the system

Decoder or **Action generator**:

given hidden state of the system generates output

Traditional DS Pipeline in Neural Network Implementation



Metric	N2N	Modular
Subj. Success	96.95%	95.12%
Avg. # of Turn	3.95	4.54
Comparisons(%)		
Naturalness	46.95*	25.61
Comprehension	45.12*	21.95
Preference	50.00*	24.39
Performance	43.90*	25.61

* p <0.005, # of comparisons: 164

Wen, Tsung-Hsien, et al. "A network-based end-to-end trainable task-oriented dialogue system." *arXiv preprint arXiv:1604.04562* (2016).

Conversational Intelligence Trends

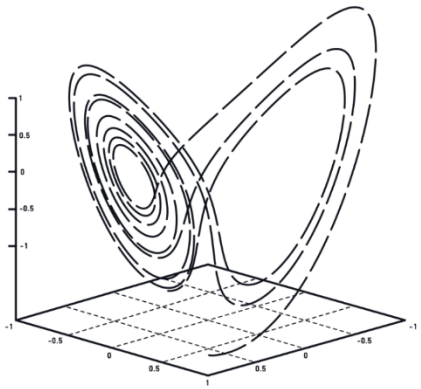
- Field of Conversational AI is developing rapidly
  - Deep learning revolutionize NLP
  - Big data
- Current industrial solutions are hybrid
  - Rule-based & Pattern matching
  - Classic ML
  - Deep NNs
- Emergence of multiskill dialogue agents
- Evolution towards end-to-end systems

```
# Definition of iPavlov project
def iPavlov(talent, ideas):
    research = ideas * talent
    AI = development(research)
    return AI
```

Deep learning architectures for the conversational intelligence

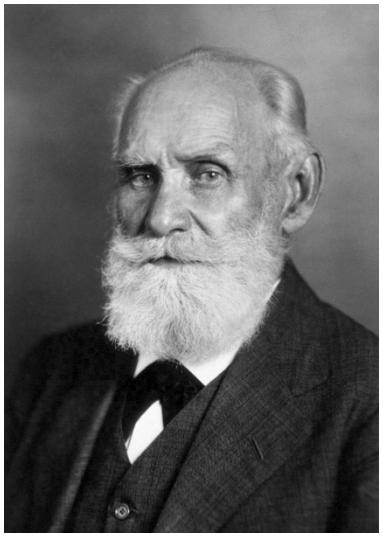
- The major lab project for the 2017-2019 started July 2017.
- Joint project with Sberbank the largest bank in Russia (operating income $20 billion, total assets $400 billion (2014))
- MIPT MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

SBERBANK By your side



National Technology Initiative

Space of possibility



Ivan Petrovich Pavlov
(1849 –1936)
Russian physiologist known for his work in classical conditioning.

Stakeholders & Workpackages

Research

Neural architectures for dialogue systems

Neural networks and reinforcement learning for planning

Development

DeepPavlov.ai
open source library

Repository of dialogue agents

Lego-like modules for the fast prototyping

Service NLP functions

Applications

DeepReply
services

Conversational agents for specific business cases

API for separate NLU, DM, NLG tasks

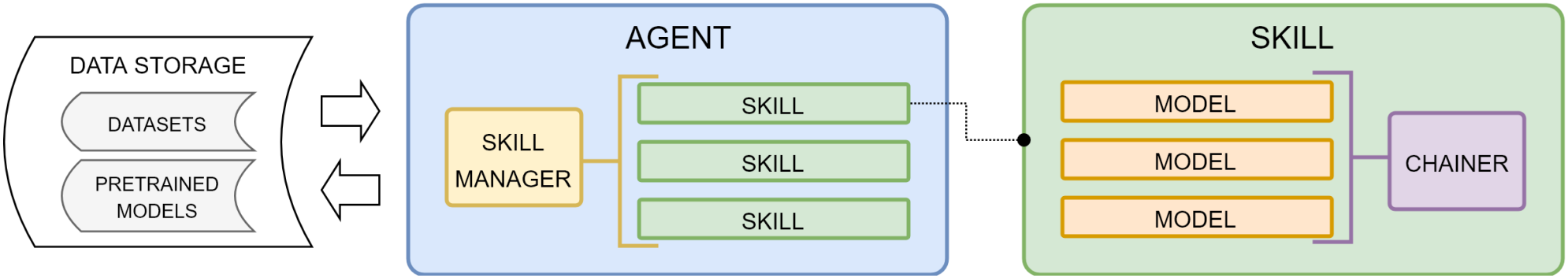
MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

НЕЙРОНЕТ

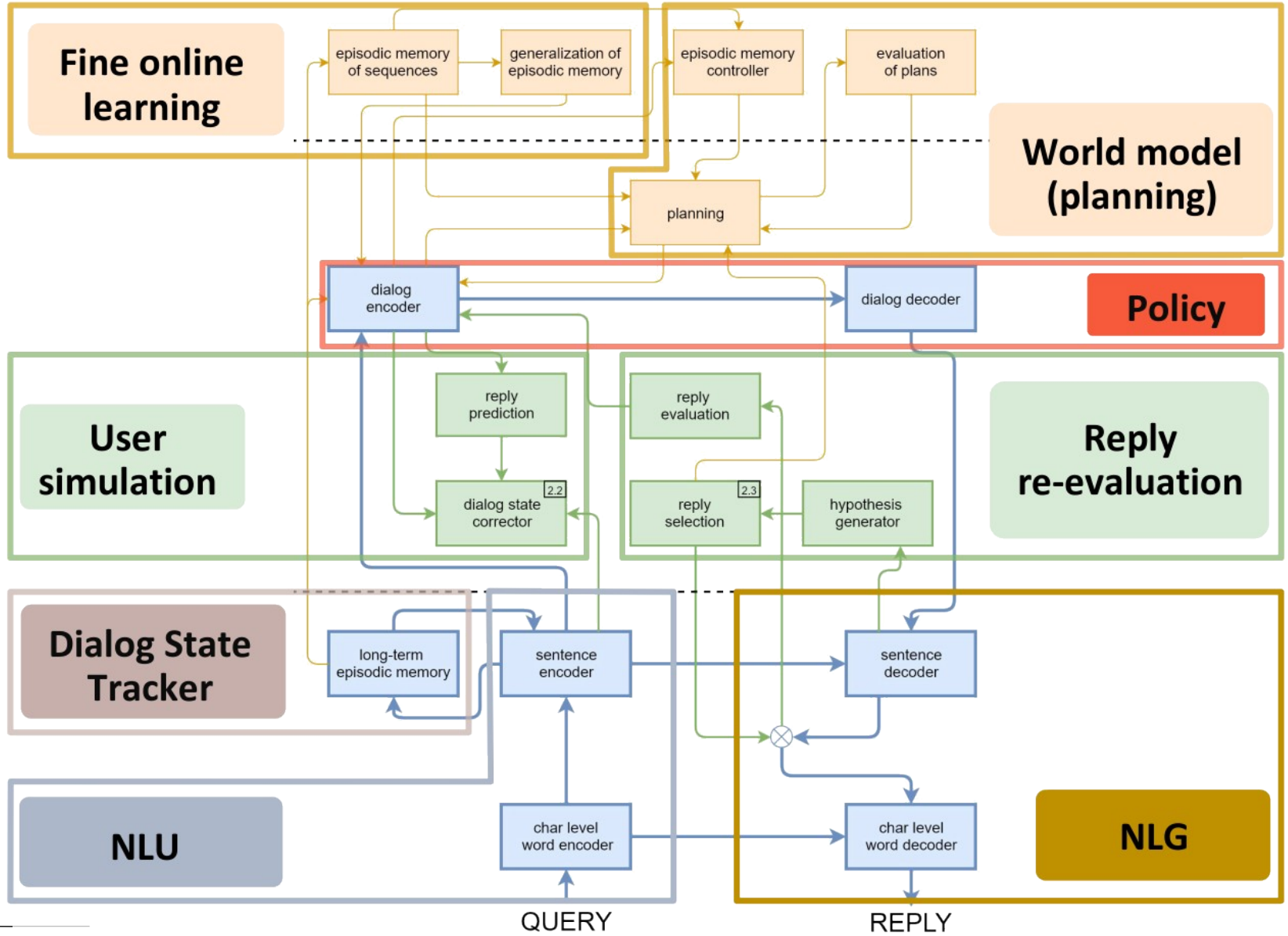
iPavlov.ai

MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

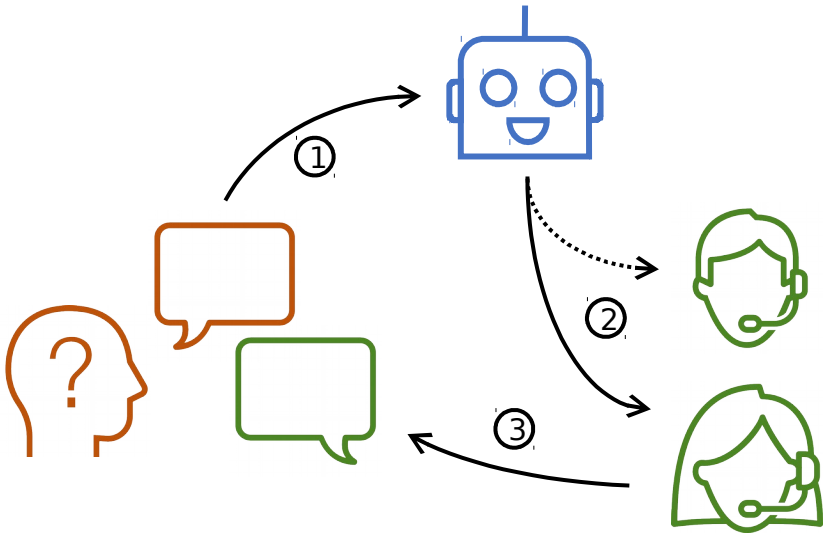
DeepPavlov.ai



Sketch of the integrated architecture

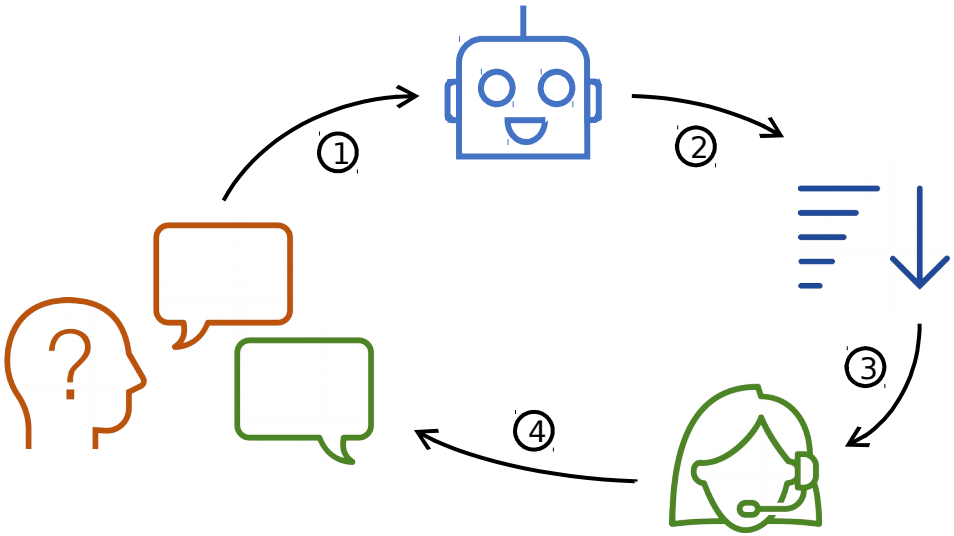


1. Request routing



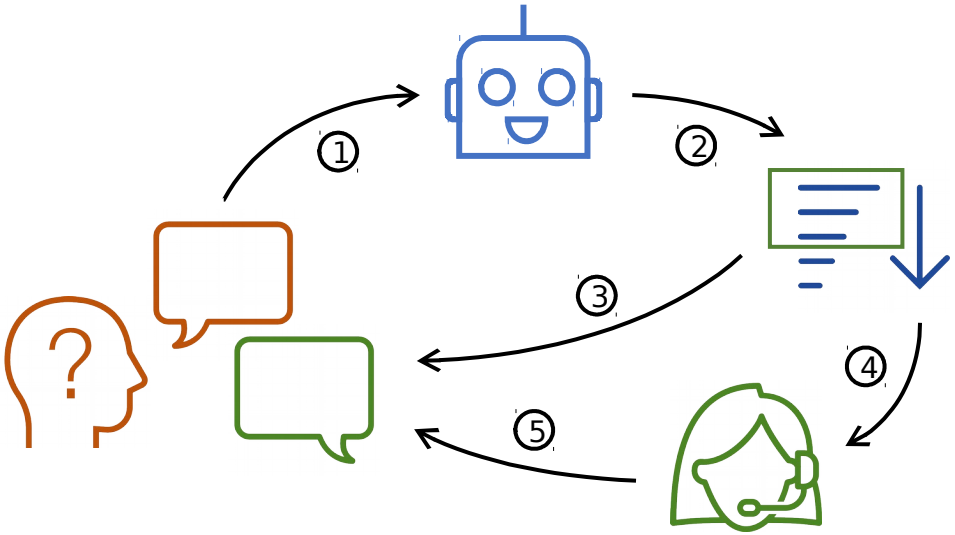
1. Domain classification
2. Routing to an operator
3. Operator replies

2. Rating of pre-defined answers



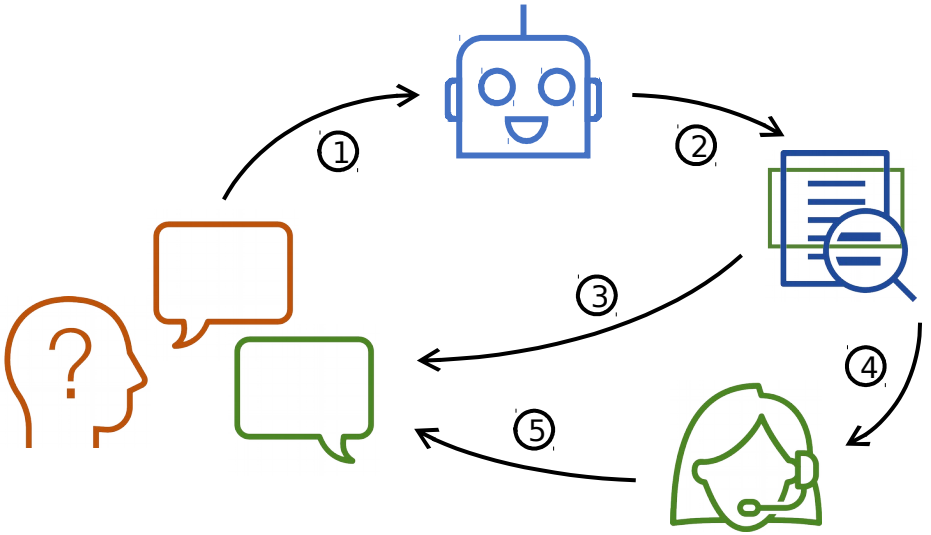
1. Semantic embedding
2. Scoring of replies
3. Best answers are presented to an operator
4. Operator replies

3. Frequently asked questions



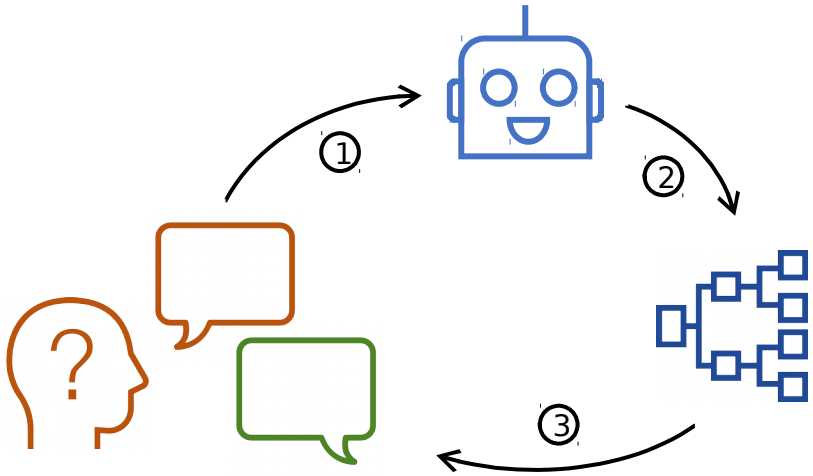
1. Semantic embedding
2. Scoring of replies
3. Automated reply if the best answer has a high confidence
4. Routing to an operator in the case of low confidence
5. Operator replies

4. Knowledge base Q&A



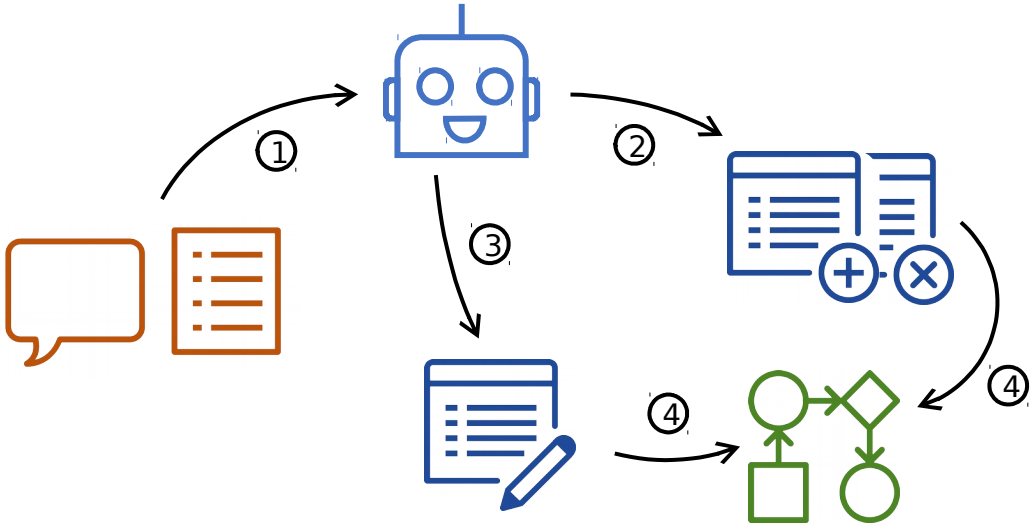
1. Semantic embedding
2. Search of answer in collection of documents
3. Automated reply if the best answer has a high confidence
4. Routing to an operator in the case of low confidence
5. Operator replies

5. Simple bot



1. Semantic embedding
2. Selection of the most relevant dialogue script
3. Natural language answer generation

6. Other NLP tasks



1. Semantic embedding
2. Sentiment analysis
3. Entity recognition \ tagging
4. Integration with BPM system

Summary

- **DeepPavlov** is an open source framework for the conversational AI
- **DeepPavlov** is designed for
  - development of production-ready chatbots and complex conversational systems;
  - research in dialogue systems and NLP in general.
- **DeepPavlov** has goal to enable AI application developers and researchers with
  - a set of pre-trained NLP models, pre-defined dialogue system components (ML/DL/Rule-based) and pipeline templates;
  - a framework for implementing and testing their own dialogue models;
  - tools for integration of applications with adjacent infrastructure (messengers, helpdesk software etc.);
  - a benchmarking environment for conversational models and uniform access to relevant datasets.

iPavlov.ai

```
# Definition of iPavlov project
def iPavlov(talent, ideas):
    research = ideas * talent
    AI = development(research)
    return AI
```

Interactive demo

http://demo.ipavlov.ai/

Source code

https://github.com/deepmipt/