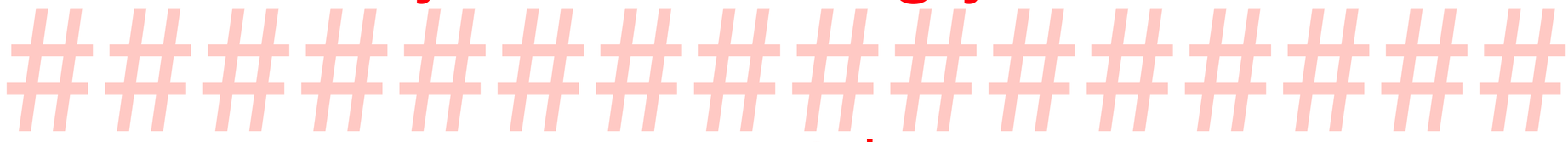
what does(n't) work for chatbots

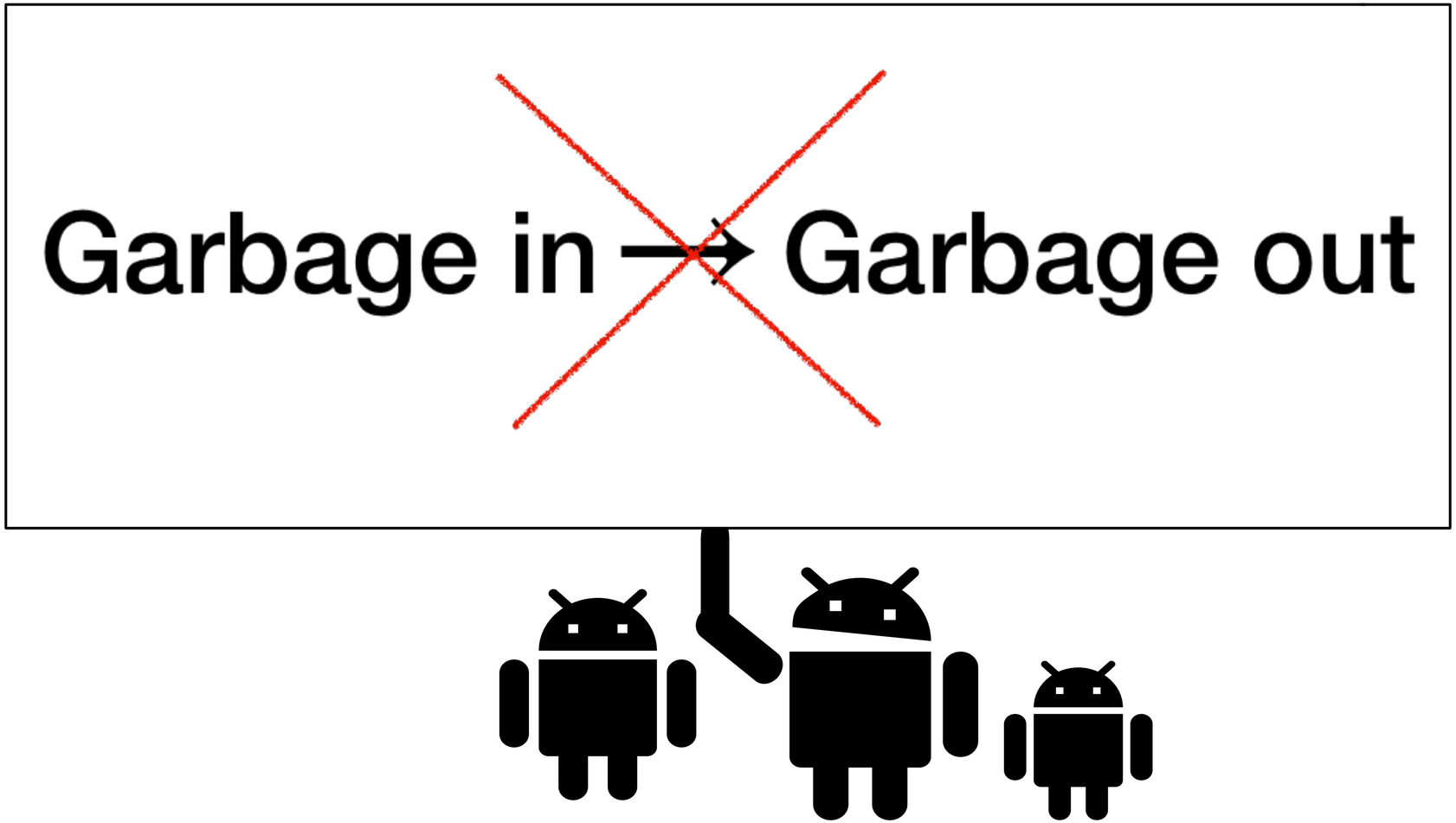
Oleksandr Khryplyvenko
m3oucat@gmail.com

have you tried fixing your device

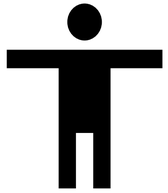


with a screw driver? | 10W 50K vocab

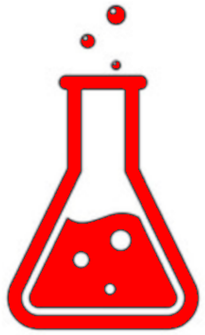
try screw driver | 3W 4K vocab


Garbage in → Garbage out

NNs suffer from imbalanced datasets



I'd like to refund	i don't know
my unit is broken	i don't know
do you work?	i don't know
can i see your manager?	i don't know
whatever question?	i don't know



data cleaning
preprocessing

vocabulary reduction

data separation

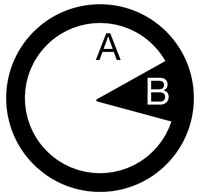
labeling

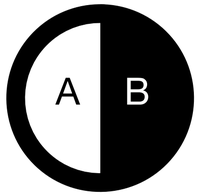
clustering

dataset balancing

vocabulary reduction

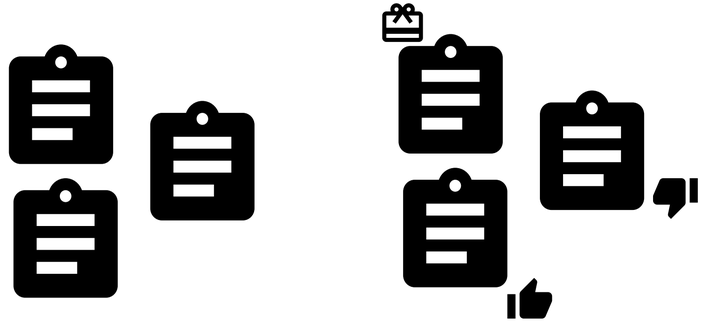
frequency, lemmatization, spelling





dataset balancing

reduce bigger, increase smaller



labeling

intents, slots, rewards, start/end of answer

data separation

conversational, query, task completion

cluster separated data

for typical problems

components

NLU

↓

State Machine (SM)

↓

NLG

o=question

↓

$s, \pi : S \rightarrow A$

↓

a=answer

NLU

selects a model to use from

{conversational, query, task completion}

intent detection & slot values https://dialogflow.com/

"wanna pizza margarita @ work, pay cash"

order_pizza(loc=_, pay=cash)

SM

makes you able to define policies

"wanna pizza margarita @ work, pay cash"

o(bservation) = order_pizza(loc=_, pay=cash)

H(istory): work: loc = 49,72

S(tate) = order_pizza(loc=49,72, pay=cash)

π(olicy): $S \rightarrow A$(ction space)

π(order_pizza(loc=*, pay=*)) =

= a_deliver(order(loc=loc, pay=pay)) = a(ction)

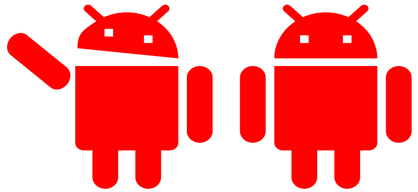
NLG

runs on state machine

depends on policy

templates(user-friendly answers)

```
a_greet = random.choice(["Hi", "Hello", "Greetings"])
```



Chatbots

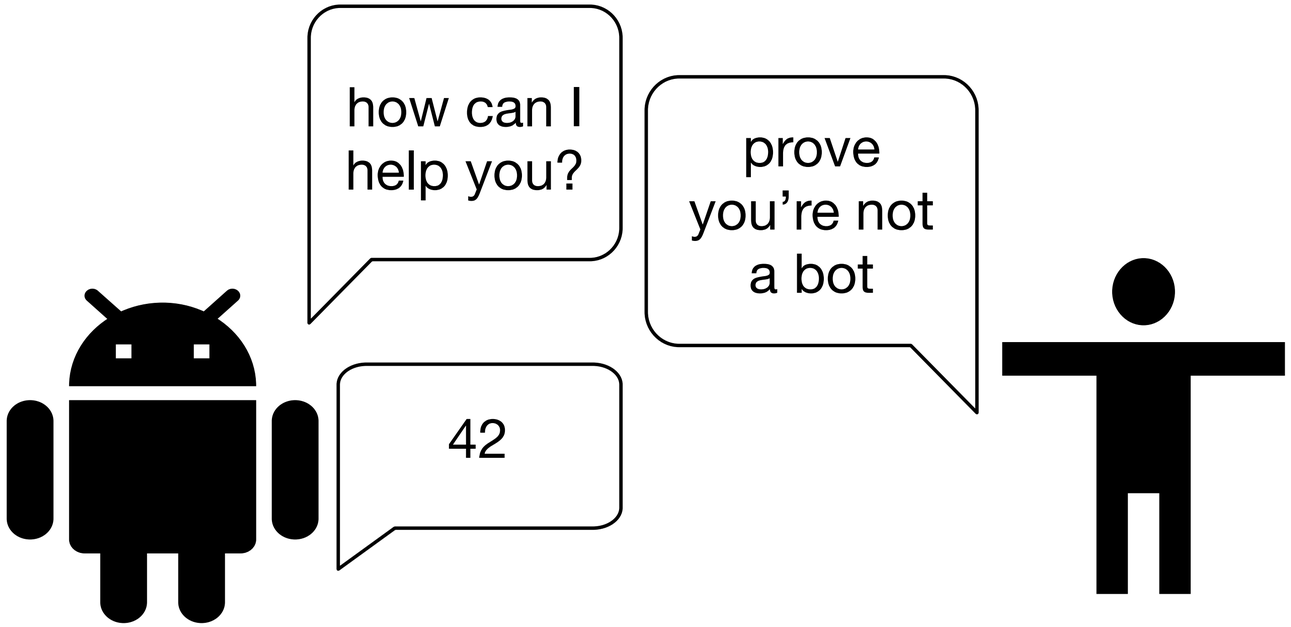
conversational

task completion(TC)

text query(TQ)

conversational

sustain a chat



covers a lot of existing text w/o labeling

NLU, SM, NLG, TQ: all in one

suffers from imbalanced datasets

hard to train

hard to manage NERs

conversational

plain seq2seq

seq2seq + attention

transformer

BIG models/ensembles

ranging

plain seq2seq
(uni/bi directional LSTMs)

lots of samples

simple to grasp

very high probability to stuck in

"i don't know" | dirty data

✓ seq2seq + attention

official version is easy to modify

if cooked properly, performs the best

beam search out-of-the box

MMI may be easily added

works(better) when trained on one to many

takes more time to train than transformer

https://github.com/tensorflow/nmt

https://arxiv.org/pdf/1510.03055.pdf MMI

https://www.youtube.com/watch?v=RLWuzLLSIgw beam search

i can't use this device in my life

you can use the browse by device type make search engine on our home page

i searched there and it says it should work

if you have any question later let us know

/common answer I unrelated context/

i would like to refund

i will let them know to refund you the refund

may i talk to your manager

i will let her know to send it to you

you're pissing me off
yes
why you're doing this?
i do not know if it is possible to be
to be what?
you can also use the browse
i made browsing and it assures me i should use my charger
you can check our youtube channel to see if there is anything you could do
/from time to time it gives helpful answers because trained on good data/

i just watched your channel and it says i'm charging it in wrong way

i am sorry i am glad i can help

what are you sorry of? you haven't helped me!

i am not sure what you mean by more question i will have to check on forum for that device

thank you a lot for keeping trying

i will let her know to send you the document

you just ensured me that you'll look for a solution on forum

i will have to check on forum coupon code

no need a coupon, need my device working!

you can use the search engine on our home page to link you to the manual for your device

nice chat sample

my unit doesn't work

what firmware version are you use

i'm using firmware version 2.1

you can try update firmware if you would

ok, I'll update it, thanks

yes correct

transformer

available on github

feedforward

quick to train

does worse than nmt

official version is not so easy to customize

too generic answers(only attention)

https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

my ubuntu doesn't work
i would say that is a bit of a of
a bit of what?
i would say bit
can you tell me something meaningful?
i do not know
i can't get online on my laptop
what is the problem
i can't install software
what is the error
the error is: exception case
i would try the alternate eol of the ubuntu

BIG models, ensembles

if you've got lucky, you get diverse answers

require too much resources

not so easy to modify, even run

high chance "I don't know" | YOUR data

require labeled data

https://github.com/jiweil/Neural-Dialogue-Generation

ranging (concept)

$$q = argmin\ D(q_{user}, \forall q \in Q)$$

$$\pi = f : Q \rightarrow A$$

OR

$$a = max_{score} a \mid q$$

simple to implement

easy to debug

works for TQ(text query)

heavily depends on encoding quality

(huge labeled datasets)

chat context

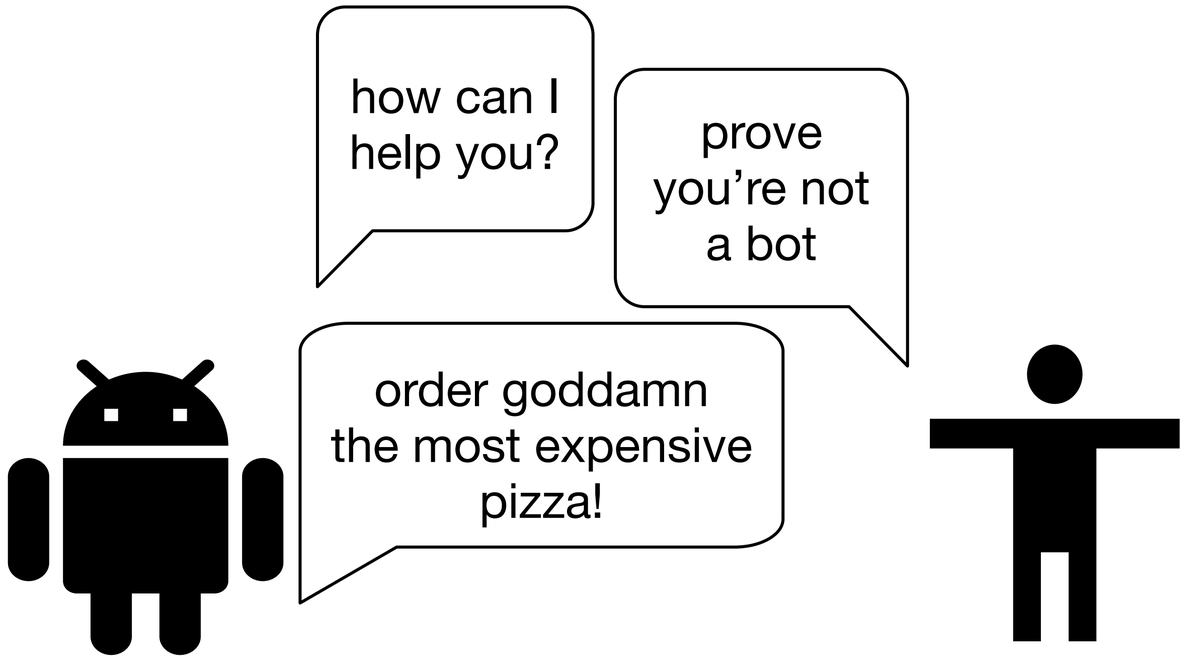
for sentence embedding - you can simply add avg so far to each embedding as a context

https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html

TC

have to complete a task

or conclude there's no way you can do it



tractable, easy to debug & modify

requires lots of manual work

most used approaches require NLU

 (RL) in discrete space

$$f : \underbrace{V^{\ell} \rightarrow \overbrace{\mathbb{R}^{n} \rightarrow \{0,1\}^{m}}^{LSH} \rightarrow V^{\ell}}$$

$$\underbrace{pretrained\ NN}_{word/sentence \rightarrow embedding} \quad \underbrace{\pi : S \rightarrow A}$$

hard to keep LSH as non-linear homomorphism(metric changes)

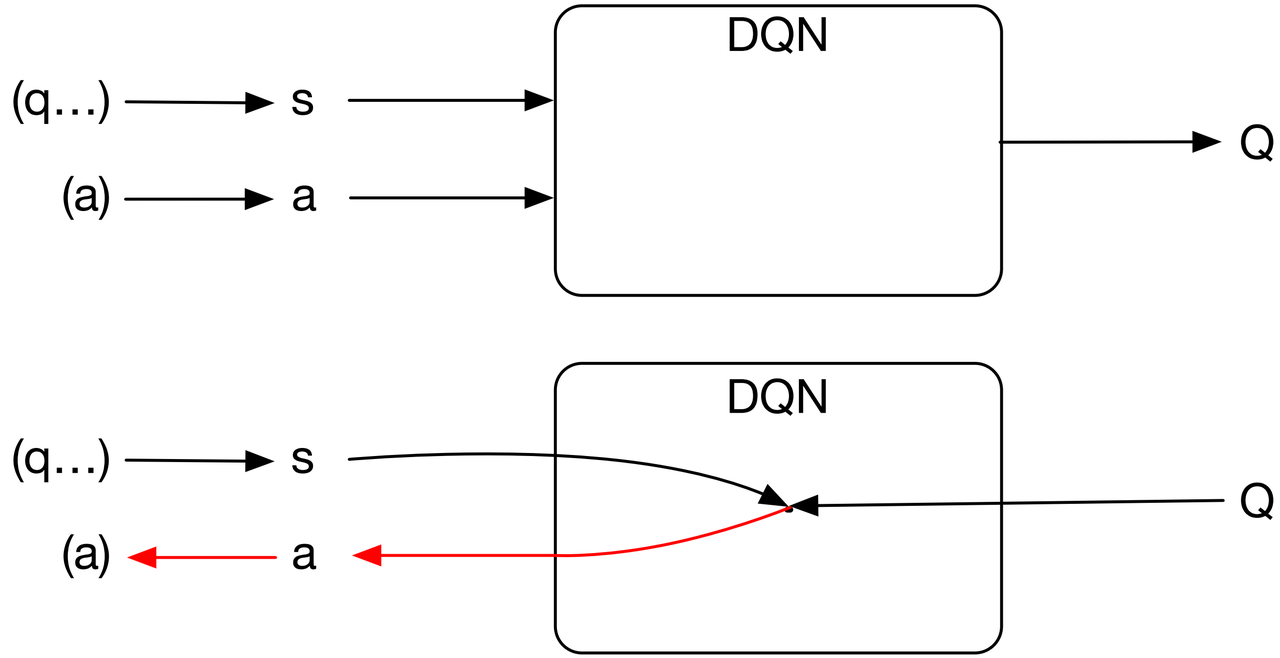
pretrained NN = skipthoughts, BiMPM...
R(eward) depends on how successfully you complete a task

? (RL) in continuous space

state definition

- We may say, we simply learn policy $\pi : S \rightarrow A$
- q/a = all successive user's/agent's sentences(utterance) before agent's/(user's) ones respectively
- a = a
- s = q(aq)(aq)(aq)... | all (q...q) so far

$$a = arg\max_{Q} \frac{\partial Q(s,a)}{\partial a}$$



$$(q\ldots), (a) \in vocab \qquad s \in \mathbb{R}^m, a \in \mathbb{R}^n, Q \in \mathbb{R}$$

the problem even worse when a is a sentence-wise vector

$$a \in \mathbb{R}^n !$$

getting back to discrete V space is non-trivial

we have to train another network

$$a \in \mathbb{R}^n \rightarrow a \in V^{\ell}, \ell = sen\ len\ in\ words$$

✓ Forms

interpretable

allows RL in discrete space(converges)

datasets are optional (coldstart)

lots of manual work

need NLU

need to define default policy

if state defined wrong, pisses off users

SM test sample

```
TEST_CASE = [
  [
    "Hey, buddy, I have nothing to do. I'd like to talk to someone",
    [
      ("NO_INTENTS_DETECTED", dict()),
    ]
  ],
  [
    "Could you give me a pepperoni and one more",
    [
      ("specify_pizza", {"pizza_type": "pepperoni", "quantity": 1}),
      ("order_pizza", dict()),
    ]
  ],
                                    ...
  [
    "yep, that's all",
    [
      ("confirm", dict())
    ]
  ]
]
```

```
Hey, buddy, I have nothing to do. I'd like to talk to someone
Hello
What kind of pizza do you want?
  you can select pre-set pizzaz
  or you can make you own pizza from scratch from the following ingredients:
  cheddar salami mozzarella tomatoes sweet red pepper pineapple chicken bacon
chili red beef
preset pizza types are: margarita pepperoni chicken vegetarian calzone
Could you give me a pepperoni and one more
What kind of pizza do you want?
  you can select pre-set pizzaz
  or you can make you own pizza from scratch from the following ingredients:
  cheddar salami mozzarella tomatoes sweet red pepper pineapple chicken bacon
chili red beef
preset pizza types are: margarita pepperoni chicken vegetarian calzone
which address deliver to?
delivery time?
which payment type do you prefer?
...
Would you like to add something else to your order?
yep, that's all
{'delivery_address': '80 banana st', 'delivery_time': '14:00', 'payment_type':
'cash', 'cart': [('pizza', {'pizza_type': 'pepperoni', 'quantity': 1}),
('custom_pizza', {'quantity': 1, 'ingredients': {'cheese': 1}})]}
```

S definition

question:

words = sentence.lower().split()

'cat' and 'meow' in words = s^1_{cat}

'cat' and 'scratch' in words = s^2_{cat}

'dog' and 'barks' in words = s^1_{dog}

(this something like google's dialog flow does)

- is it a state?
- when it is a state?

This is a state if you define a complement state:

$$s_{default} = s^c = S \backslash \{s^1_{cat}, s^2_{cat}, s^1_{dog}\}$$

Then your task is to define a policy which will reach a terminal state: when all slots for all intents are set or proven they couldn't be

Often, action for the default state is fallback to conversational

...Still, is S a well-defined state space ?

A definition

a is a predefined answer from a bot to a user which makes(convinces) the user to fill some slot or reveal an intent

also, there's hidden 'a' part - change bot state (if needed)

```
intent_1(_, _, 'slot_3_val') ->

    db.save(intent_1.slot_3)

    ask_a_user_fill_slot_1_for_intent_1()
```

until there are ambiguities
your state is bad defined

ambiguity resolution!(context, pronouns, ...) based on previous observations

No ambiguities, then s as defined above is a state

intent_1("panda", "eats", "shoots"| bank in history) -> call_the_police

intent_1("panda", "eats", "shoots"| animal in history) -> take_a_photo

$$\pi : S \rightarrow A$$

...**for now** your model is fully tractable

SM(recall)

makes you able to define policies

"wanna pizza margarita @ work, pay cash"

o(bservation) = order_pizza(loc=_, pay=cash)

H(istory): work: loc = 49,72

S(tate) = order_pizza(loc=49,72, pay=cash)

π(olicy): $S \rightarrow A$(ction space)

π(order_pizza(loc=*, pay=*)) =

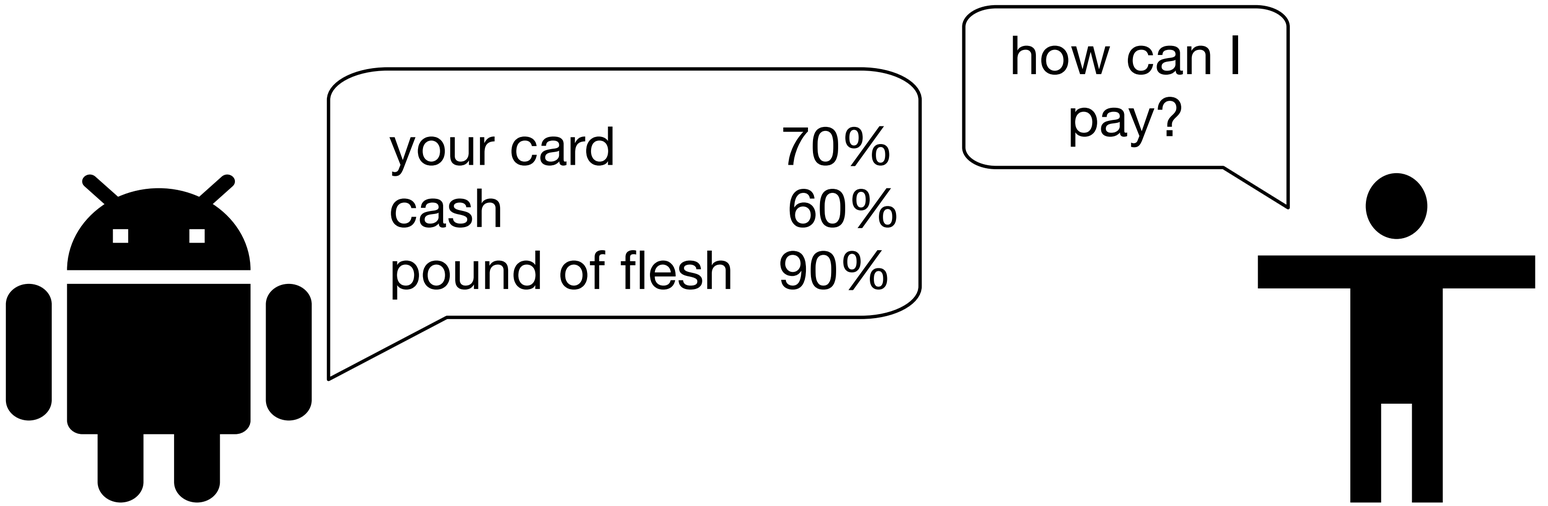
= a_deliver(order(loc=loc, pay=pay)) = a(ction)

same works for ranging

text query(TQ)

the best text search you can

by documents you have



+ pretrained models available
- hard/impossible to train on own data
- hardcore labeling

BiDAF

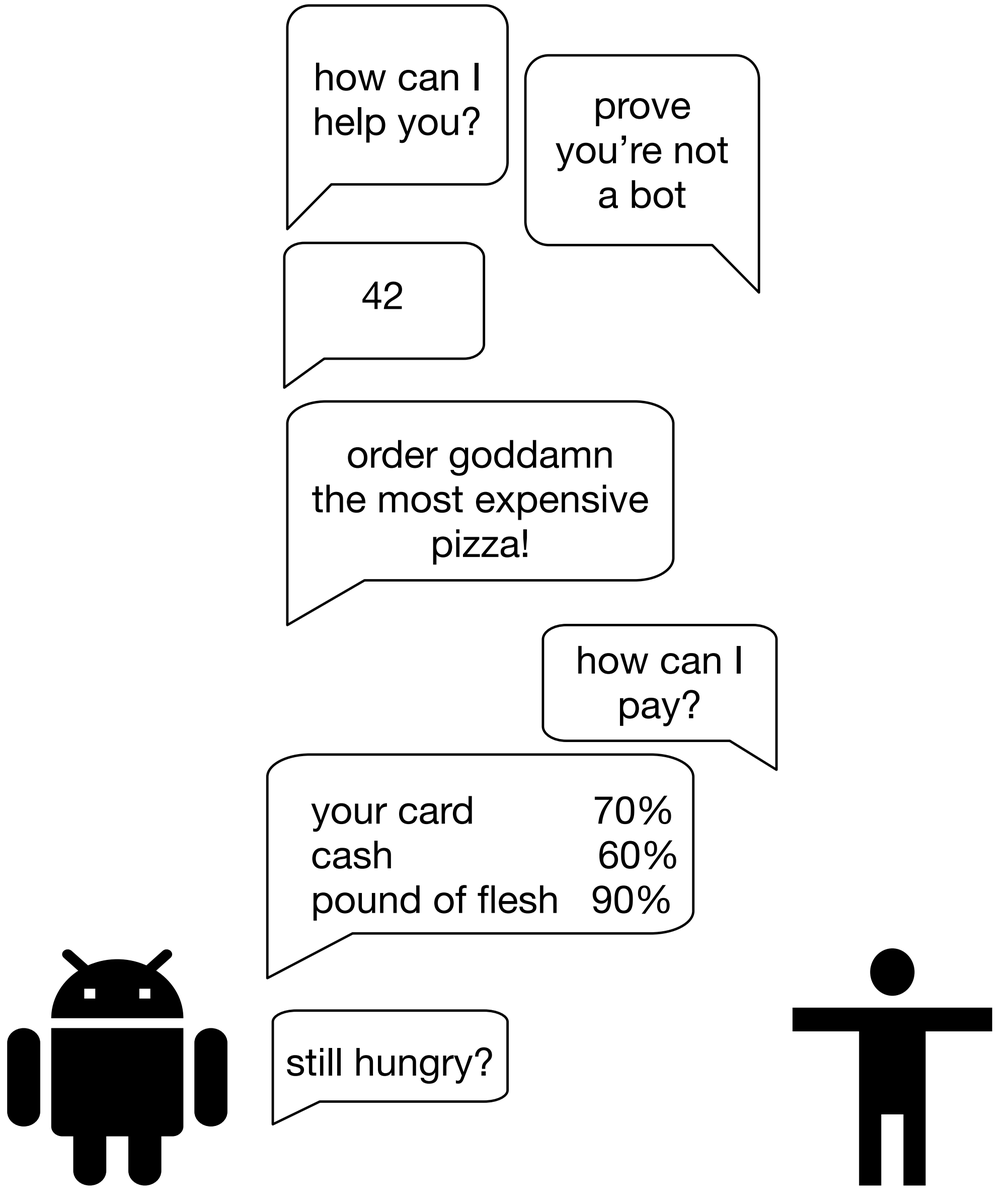
https://allenai.github.io/bi-att-flow/

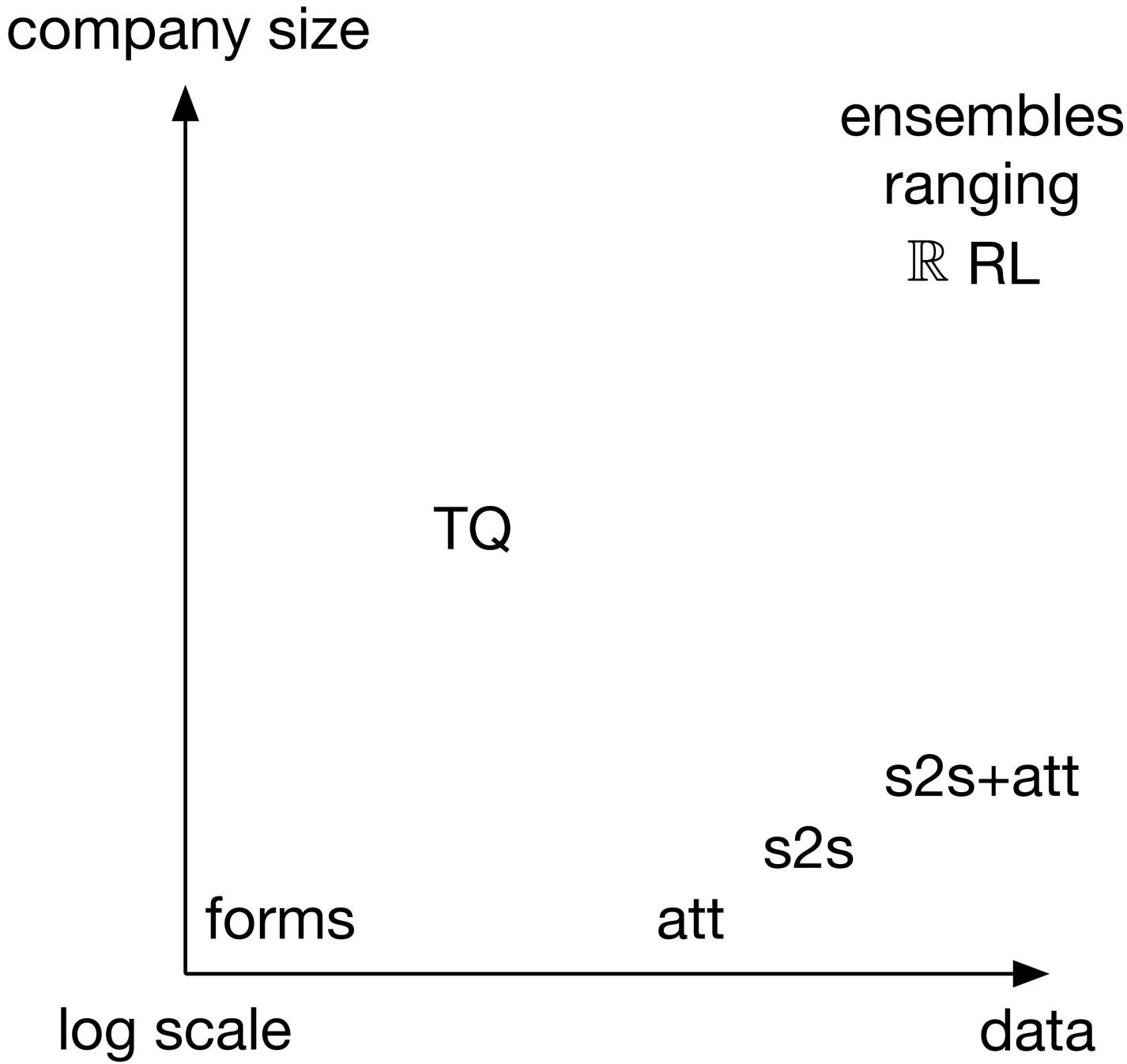
bAbi

https://research.fb.com/downloads/babi/

Used if you have some unstructured text DBs, like manuals, HOWTOs, etc.

All 3 within a dialogue




company size
ensembles
ranging
ℝ RL
TQ
s2s+att
s2s
forms
att
log scale
data

Ecosystem?

Yandex Алиса
https://www.youtube.com/watch?v=_law_tey0OQ

Amazon Alexa + skills

Google dialog flow
https://dialogflow.com/

Storyline(Alexa skills)
https://getstoryline.com/

Deep Pavlov
https://deeppavlov.ai/

some open problems

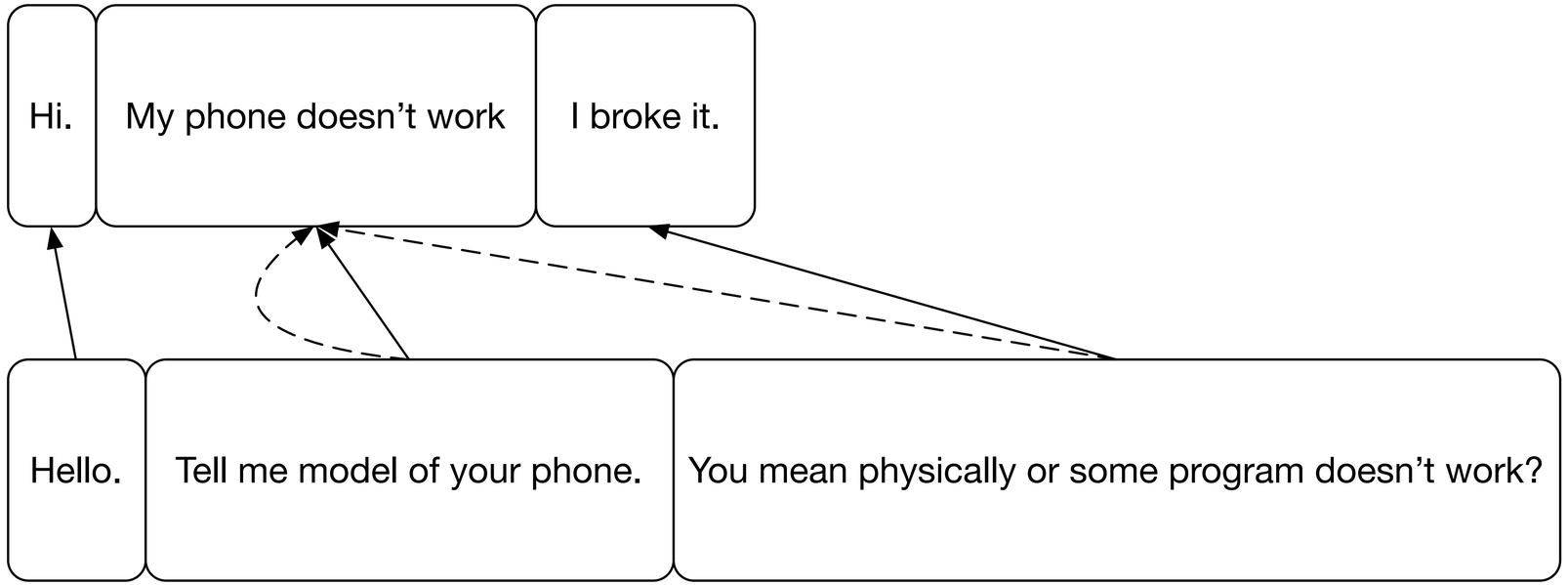
Chinese room
(open domain)

no good objectives

(ACC is mostly used)

Best evaluated by

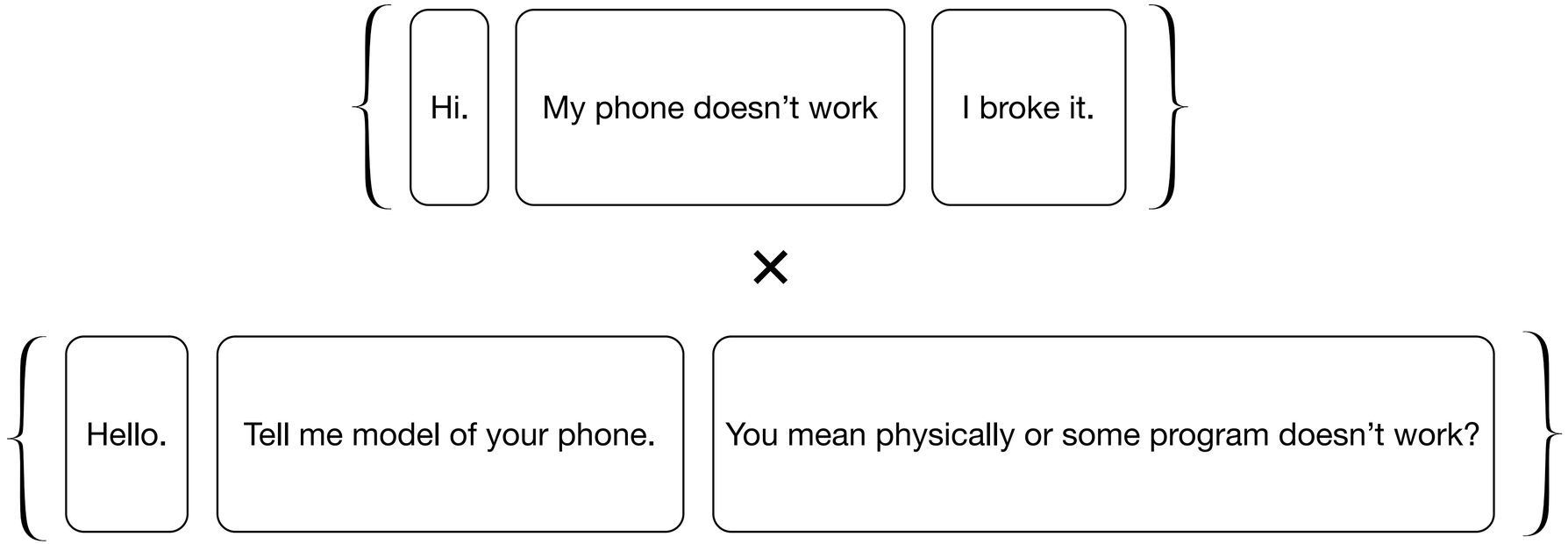
utterance-utterance



Either boring data labeling...

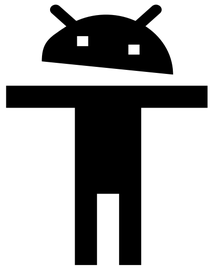
... or merge utterance in a single sentence?

too long | high I/O variation sentences stuck in "I don't know"



then nmt on I/O all pairs

sounds good, does work. But don't know why.



thanks!

see this soon on

https://oleksandr-khryplyvenko.github.io/

m3ucat@gmail.com