

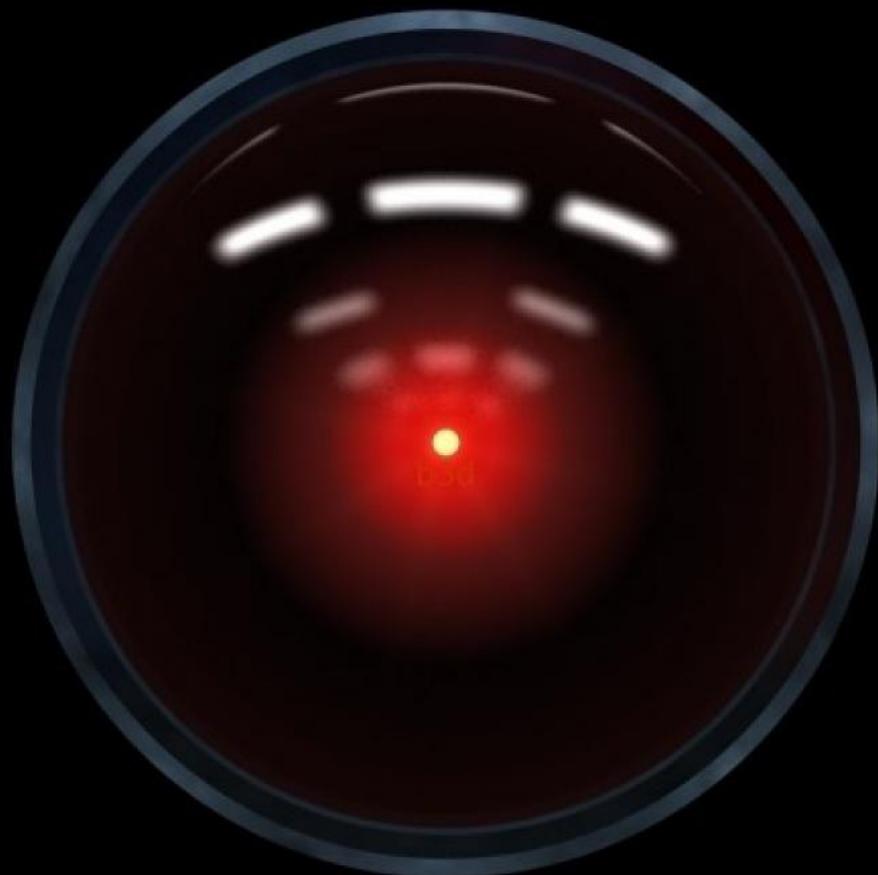


Conversational agents

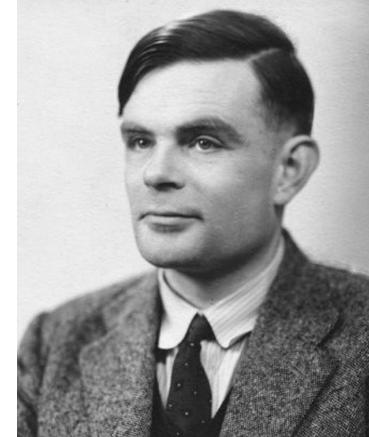
Benoît Sagot

Inria (ALMAnaCH), Paris (France)

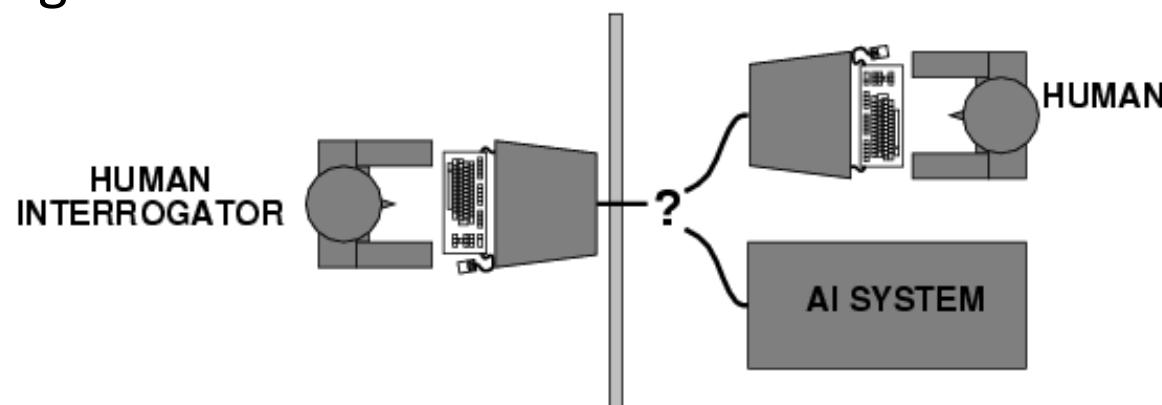
Credit and disclaimer: some of the following slides are taken from, illustrated or inspired by presentations by Jurasfky, Li, Lazebnik, Galley and Bellegarda



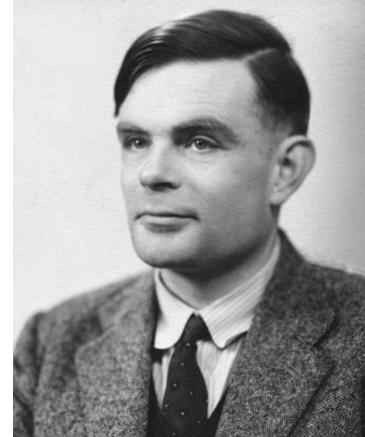
The “Imitation Game”, a.k.a. “Turing Test” (Turing 1955)



- Modern reformulation: a system passes the “Turing Test” if it can fool 33% of a pool of human interrogators for 5 minutes about it being a machine rather than a human
- It requires advanced conversational abilities, which itself requires the system to have abilities in NLP (understanding + generation), knowledge representation and automated reasoning



The “Imitation Game”, a.k.a. “Turing Test” (Turing 1955)



- Modern reformulation: a system passes the “Turing Test” if it can fool 33% of a pool of human interrogators for 5 minutes about it being a machine rather than a human
- It requires advanced conversational abilities, which itself requires the system to have abilities in NLP (understanding + generation), knowledge representation and automated reasoning
- Problems:
 - cheap tricks (spelling errors, transformation of the user’s input...)
 - variability in protocols and judges
 - success depends on deception, not real success

Winograd schemas

- Winograd schema = Multiple choice questions that can be easily answered by people but cannot be answered by computers using “cheap tricks”
- Example:

The trophy would not fit in the brown suitcase because it was so small. What was so small?

 - *The trophy*
 - *The brown suitcase*
- **Advantages** over standard Turing Test
 - Test can be administered and graded by machine, and does not depend on human subjectivity
 - Does not require ability to generate English sentences (a good thing?), and therefore questions cannot be evaded using verbal dodges
 - Questions can be made “Google-proof” (at least for now...)
- **Winograd schema challenge** held at IJCAI conference in July 2016
 - Six entries, best system got 58% of 60 questions correct
 - (humans get 90% correct)

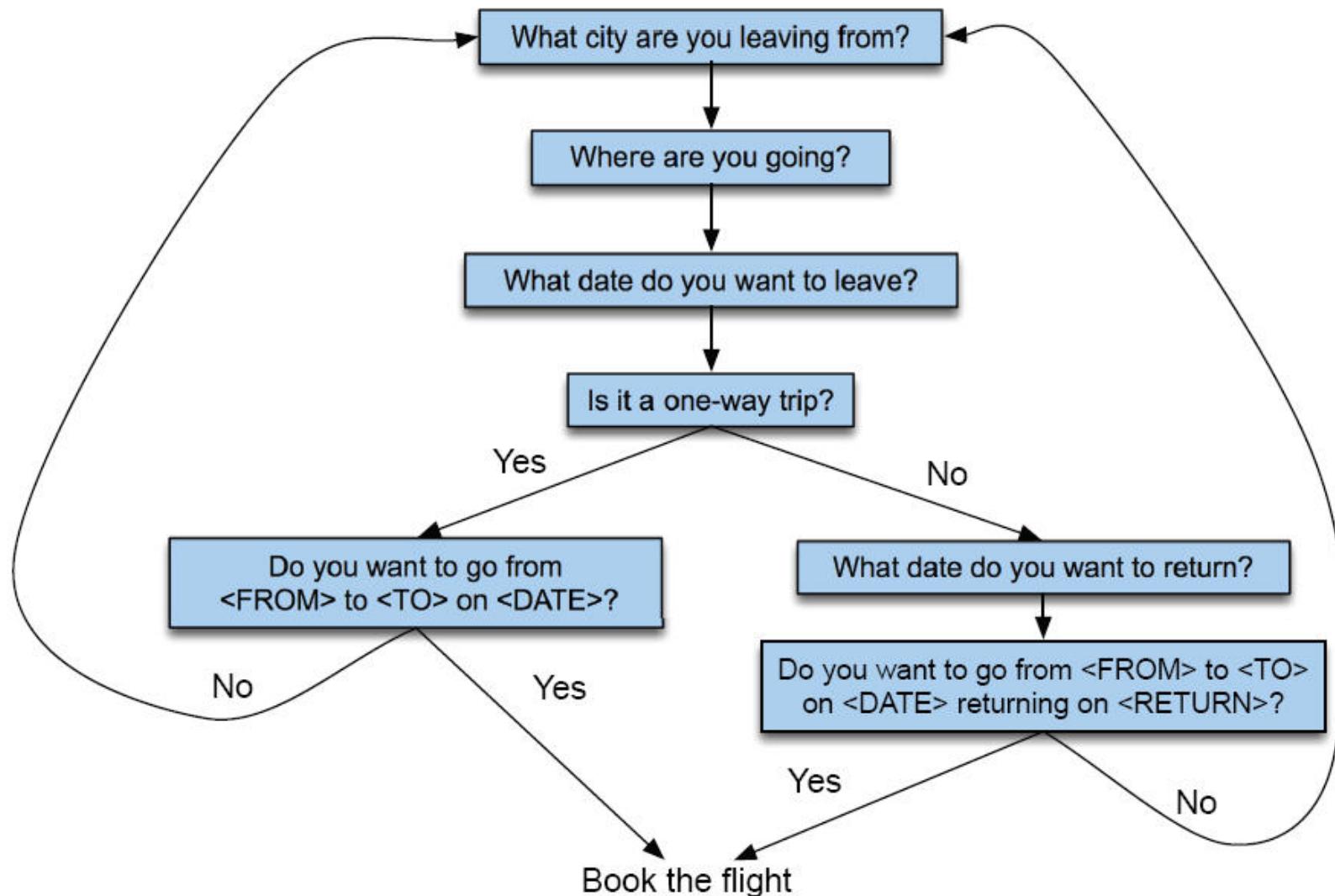
Finite-state dialogue management



On an example

- Consider a trivial airline travel system:
 - Ask the user for a departure city
 - Ask for a destination city
 - Ask for a time
 - Ask whether the trip is round-trip or not

On an example



System Initiative

- The system completely controls the conversation with the user
 - It asks the user a series of questions, ignoring (or misinterpreting) anything the user says that is not a direct answer to the system's questions
- Systems that control conversation like this are “System Initiative” or “Single Initiative”
 - **Initiative:** who has control of conversation
 - In normal human-human dialogue, initiative shifts back and forth between participants

System Initiative

- System completely controls the conversation
 - Simple to build
 - User always knows what they can say next
 - System always knows what user can say next
 - Known words: Better performance from ASR
 - Known topic: Better performance from NLU
 - OK for VERY simple tasks (entering a credit card, or login name and password)
- • Too limited

Problems with System Initiative

- Real dialogue involves joint initiative (give and take)
- E.g. users might want to say something that is not the direct answer to the question and provide additional information

I want a flight from Milwaukee to Orlando one way
leaving after 5 p.m. on Wednesday

System Initiative + Universals

- We can give users a little more flexibility by adding universals: commands you can say anywhere
 - As if we augmented every state of the FSA with these
 - Help
 - Start over
 - Correct
 - This describes many implemented systems
 - But still doesn't allow user much flexibility

Mixed Initiative

- Conversational initiative can shift between system and user
- Simplest kind of mixed initiative: use the structure of the frame to guide dialogue
- Example for travel planning:

Slot	Question
ORIGIN	What city are you leaving from?
DEST	Where are you going?
DEPT DATE	What day would you like to leave?
DEPT TIME	What time would you like to leave?
AIRLINE	What is your preferred airline?

Mixed Initiative: Frames

- User can answer multiple questions at once
 - System asks questions of user, filling any slots that user specifies
 - If user answers 3 questions at once, system has to fill slots and not ask these questions again!
 - Avoids strict constraints on order of the finite-state architecture
 - When frame is filled, do database query
- Natural Language Understanding:
 - There are many ways to represent the meaning of sentences
 - For speech dialogue systems, most common is “Frame and slot semantics”.

Example of a Frame

Show me morning flights from Boston to SF on Tuesday.

SHOW: FLIGHTS

ORIGIN:

CITY: Boston

DATE: Tuesday

TIME: morning

DEST:

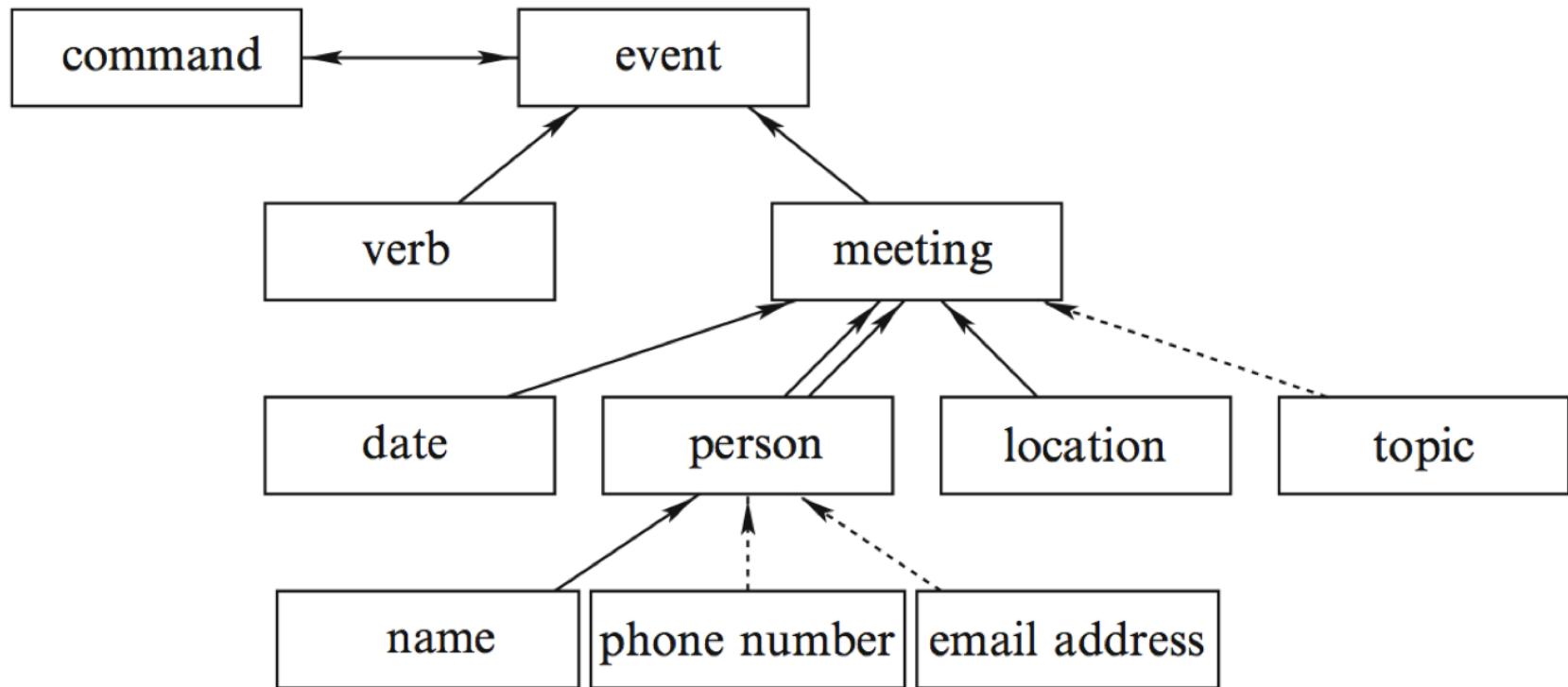
CITY: San Francisco

Condition-Action rules (SIRI)

Active Ontology: relational network of concepts

- **data structures:** a meeting has
 - a date and time,
 - a location,
 - a topic
 - a list of attendees
- **rule sets** that perform actions for concepts
 - the **date** concept turns string “**Monday at 2pm**” into date object
`date(DAY,MONTH,YEAR,HOURS,MINUTES)`
 - rules consist in condition + action: when user input is processed, facts are added to store, rule conditions are evaluated and relevant actions executed

Part of ontology for meeting booking task



meeting concept: if you don't yet have a location, ask for a location

Improvements to the rule-based approaches

- **Statistical or neural classifiers** to map words to semantic frame-filler
- Given a set of labeled sentences,
 - Example:
Sentence : “I want to fly to San Francisco on
Tuesday”
Labels: Destination: SF + Depart-date: Tuesday
 - build a classifier to map from one to the other
 - > **semantic parsing** (cf. also FrameNet)
- Requirements: **Lots of labelled data**

Towards generic chatbots: symbolic approaches



ELIZA (Weizenbaum 1966)



ELIZA: Please tell me what's been bothering you.

YOU: Men are all alike.

ELIZA: In what way?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example?

YOU: Well, my boyfriend made me come here.

ELIZA: Your boyfriend made you come here?

YOU: He says I'm depressed much of the time.

ELIZA: Who else in your family takes care of you?

YOU: My father

ELIZA: Your father?

YOU: You are like my father in some ways.

ELIZA (Weizenbaum 1966)



ELIZA: What resemblance do you see?

YOU: You are not very aggressive but I think you don't want me to notice that.

ELIZA: What makes you think i am not aggressive?

YOU: You don't argue with me.

ELIZA: Why do you think i don't argue with you?

YOU: You are afraid of me.

ELIZA: Does it please you to believe I'm afraid of you?

YOU: My father is afraid of everybody.

ELIZA: What else comes to mind when you think of your father?

YOU: Bullies

ELIZA: Does that have anything to do with the fact that your boyfriend made you come here?

Domain: Rogerian psychology interview

- Draw the patient out by reflecting patient's statements back at them
- Rare type of conversation in which one can “assume the pose of knowing almost nothing of the real world”
- Example:
 - Patient: “I went for a long boat ride”
 - Psychiatrist: “Tell me about boats”
 - You assume the patient knows what a boat is
 - You assume the patient has some conversational goal
- Most chatbots trying to pass Turing test choose a domain with similar properties

ELIZA's transformation-based approach

- Input pattern > Output pattern
 - (0 you 0 me) > (What makes yout think I 3 you)
 - Example:
You hate me

ELIZA's transformation-based approach

- Input pattern > Output pattern
 - **(0 you 0 me) > (What makes you think | 3 you)**
- Example:

You hate me [0 = any (possibly empty) sequence of words]

ELIZA's transformation-based approach

- Input pattern > Output pattern
 - you: (0 **you** 0 **me**) > (What makes you think I 3 **you**)
 - Example:

You hate me

What makes you think I hate you

ELIZA's transformation-based approach

- ELIZA relies on a **ranked list of keywords**, each of them associated with a **ranked list of patterns**
- **Algorithm:**
 - Examine each word w in user sentence
 - Return the w with highest keyword rank
 - If w exists:
 - Check each **pattern** for w in ranked order
 - Choose first one that matches sentence
 - **Apply transform**
 - If no keyword applies, either
 - Apply the transform for the “**NONE**” key, or
 - Grab an action off the **memory queue**

ELIZA's transformation-based approach

- I know everybody laughed at me
 - “I” is a very general keyword:
 - $I: (I *) \rightarrow (\text{You say you } 2)$
 - You say you know everybody laughed at you
- “Everybody” is much more interesting (someone using universals like everybody/always is probably “referring to some quite specific event or person”)
 - Who in particular are you thinking of?
- Implementation: keywords stored with their rank
 - Everybody 5: (*transformation rules*)
 - I 0: (*transformation rules*)

ELIZA: Last-resort strategies

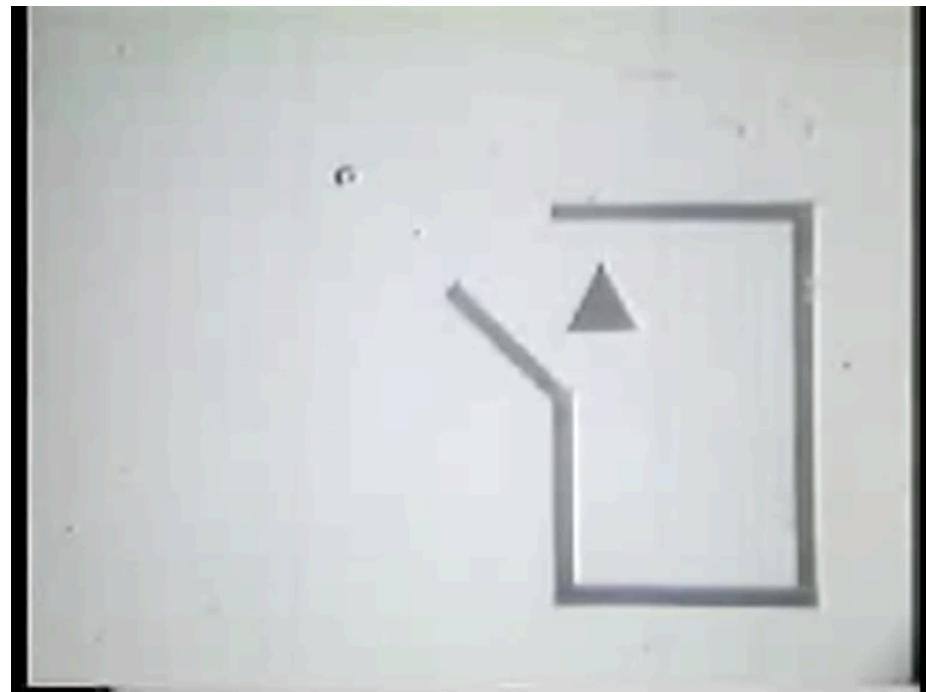
- “NONE”
 - Please go on
 - That’s very interesting
 - I see
- Memory
 - Whenever “my” is highest keyword
 - Randomly select a transform on the MEMORY list
(0 your 0 = Earlier you said your 3)
 - Apply to sentence
 - Store on a stack
 - Later, if no keyword matches a sentence
 - Return the top of the MEMORY queue instead
 - *This queue is the earliest hierarchical model of discourse*

ELIZA: Further details

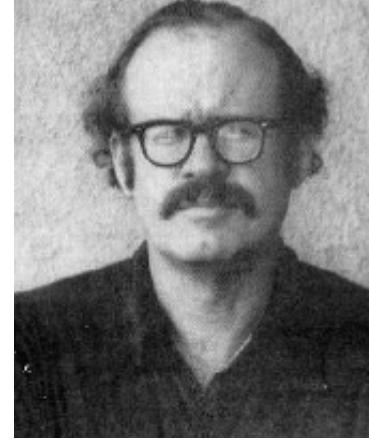
- Rules can refer to classes of words
 - Family = mother, father, brother, sister
 - NOUN = ...
- Don't reuse transforms in the same conversation
 - Whenever we use a transform associated with a pattern
 - We increment a counter for that rule
 - So the next time we use the next ranked transform
- Some basic transforms happen during input processing
 - I -> You
 - You -> I

ELIZA: Outcomes

- People became deeply emotionally involved with the program
 - Weizenbaum tells the story of his secretary who would ask Weizenbaum to leave the room when she talked with ELIZA
 - When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications
- Anthropomorphicism and the Heider-Simmel Illusion (1944)

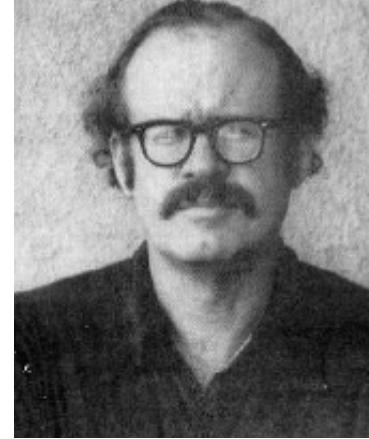


PARRY (Colby 1971)



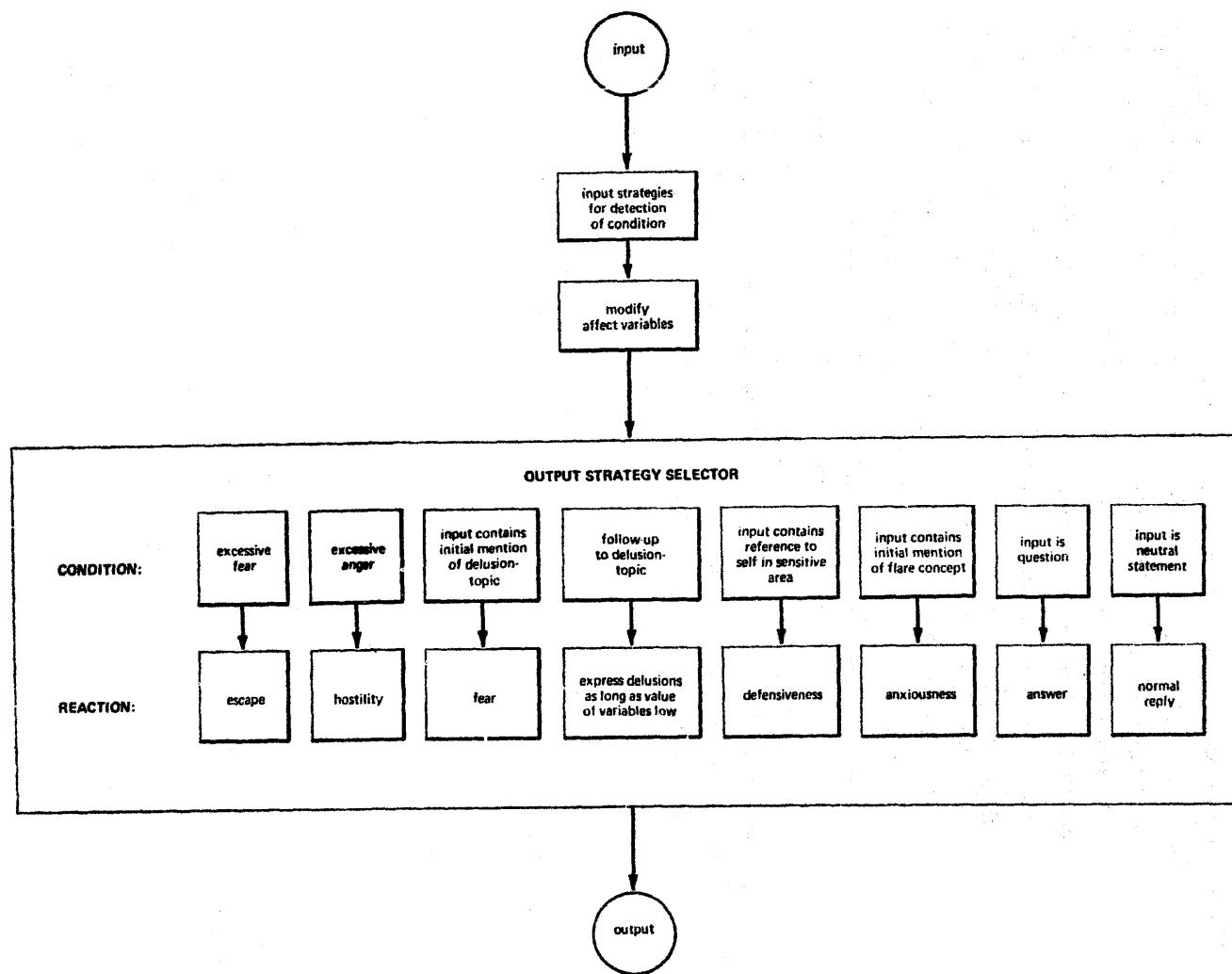
- Attempt to simulate a person with paranoid schizophrenia
- Same pattern-response structure as Eliza
- But a much richer:
 - **control structure**
 - **language understanding** capabilities
 - **mental model:** Parry has affective variables
 - Anger, Fear, Mistrust
 - “If Anger level is high, respond with hostility”
- The **first system to pass the Turing test** (in 1971)
 - Psychiatrists couldn't distinguish interviews with PARRY from interviews with real paranooids

PARRY (Colby 1971)

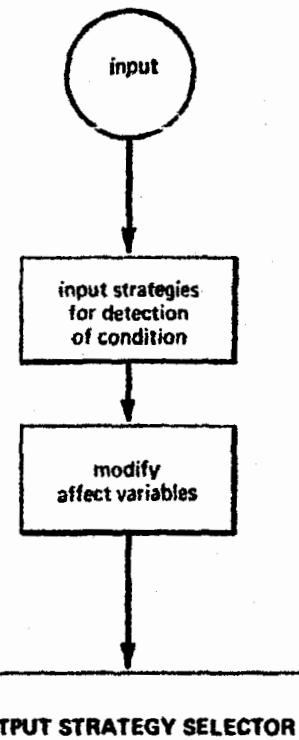


- Parry's "persona":
 - 28-year-old single man, post office clerk
 - no siblings and lives alone
 - sensitive about his physical appearance, his family, his religion, his education and the topic of sex
 - hobbies are movies and gambling on horseracing
 - recently attacked a bookie, claiming the bookie did not pay off in a bet
 - afterwards worried about possible underworld retaliation
 - eager to tell his story to non-threatening listeners

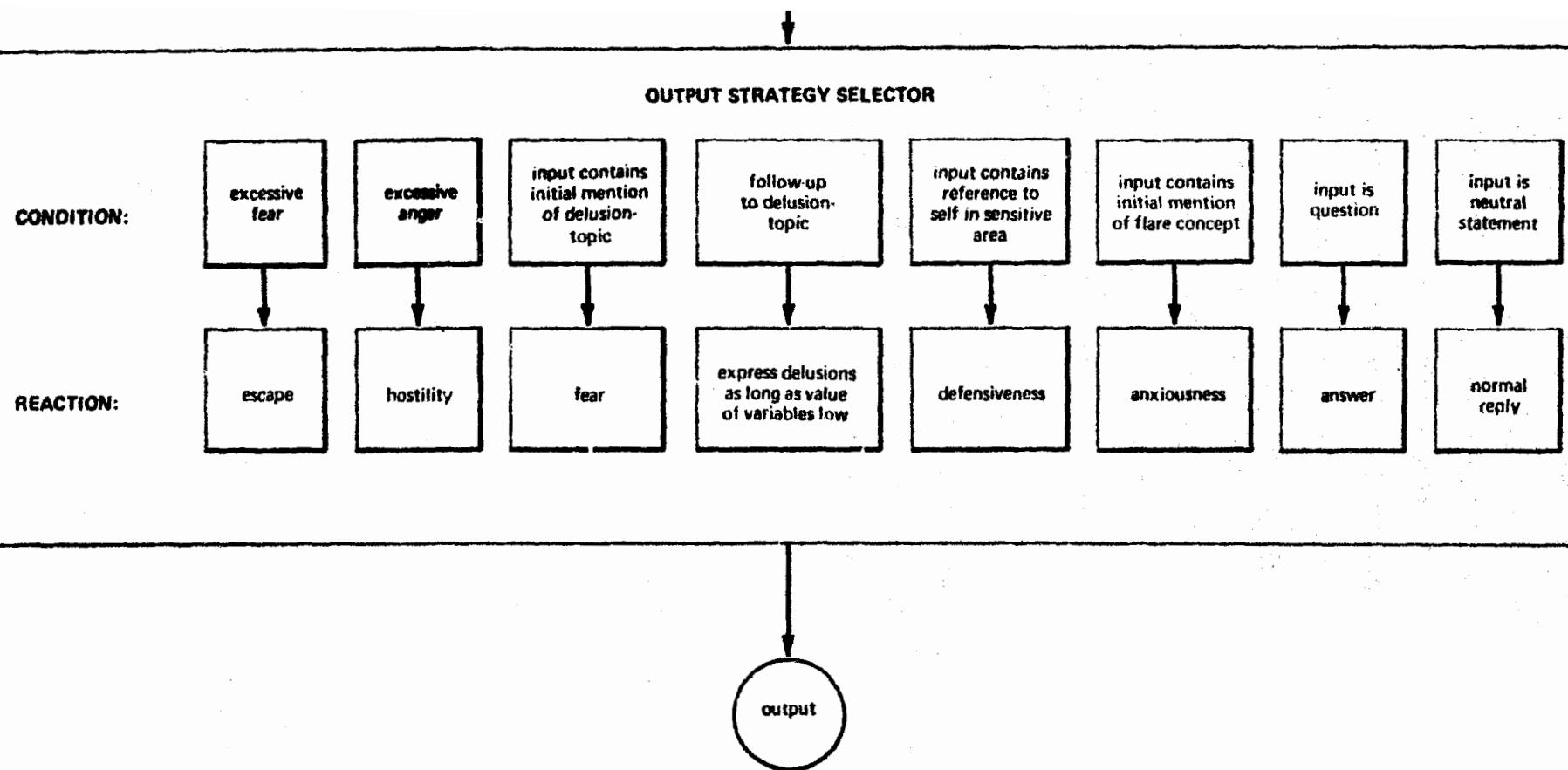
PARRY's architecture



PARRY's architecture



PARRY's architecture



PARRY (Colby 1971)

- Affect variables:
 - Fear and Anger (each ranging 0-20), Mistrust (ranging 0-15)
 - Initial conditions: All low
- After each user turn, if nothing malevolent in input
 - Anger drops by 1, Fear by 0.3, Mistrust by 0.05 to base level
- Otherwise depends on what the user says
 - **Each user statement can change Fear and Anger.** Examples:
 - User implies Parry is mentally ill > Rise in Fear and Anger
 - User mentions “Mafia” or associated concepts (“kill”) > 1st mention: rise in Fear ; later mentions: depends on willingness to discuss, i.e. on levels affect variables
 - User mentions Parry > (flattery) decreases Fear/Anger if Mistrust is low, increases Anger if Mistrust is high
 - Negative attitudes towards Parry (fear, disbelief) > increase Fear/Anger
 - **Insults increases Anger by some percentage**
 - **Mistrust goes up if Fear or Anger do**

PARRY: Language understanding

- **Flare concepts:** List of concepts related to Mafia
 - An ordered graph designed to lead interviewer to topic
 - horses>horse racing>gambling>bookies>underwold>Mafia
 - Mention of a new flare topic by interviewer > rise in Fear
 - Flare topics cause Parry to give preset responses to that flare
- “Semantic parsing/normalisation”: each sentence is mapped into a “**conceptualisation**”, i.e. a normalised semantic content expressed in the form of a sentence
 - A predication on a conceptual object
 - A predication on a relation between two objects
 - A predication on an attribute
 - Complex Pattern/transform rules
 - Different predicates (fear, afraid of)
 - Ordering (You are afraid of me = I frighten you)

PARRY: Language understanding

- **Flare concepts:** List of concepts related to Mafia
 - An ordered graph designed to lead interviewer to topic
 - horses>horse racing>gambling>bookies>underwold>Mafia
 - Mention of a new flare topic by interviewer > rise in Fear
 - Flare topics cause Parry to give preset responses to that flare
- “Semantic parsing/normalisation”: each sentence is mapped into a “**conceptualisation**”, i.e. a normalised semantic content expressed in the form of a sentence
 - A predication on a conceptual object
 - A predication on a relation between two objects
 - **A predication on an attribute**
 - Complex Pattern/transform rules
 - Different predicates (fear, afra
 - Ordering (You are afraid of m

What is your work?
What sort of work do you do?
Where do you work? → (your work?)
What do you do for a living?
What is your job?
Do you have a job?
What is your occupation?

PARRY: Language understanding

- **Flare concepts:** List of concepts related to Mafia
 - An ordered graph designed to lead interviewer to topic
 - horses>horse racing>gambling>bookies>underwold>Mafia
 - Mention of a new flare topic by interviewer > rise in Fear
 - Flare topics cause Parry to give preset responses to that flare
- “Semantic parsing/normalisation”: each sentence is mapped into a “**conceptualisation**”, i.e. a normalised semantic content expressed in the form of a sentence
 - A predication on a conceptual object
 - A predication on a relation between two objects
 - A predication on an attribute
 - Complex Pattern/transform rules
 - Different predicates (fear, afraid of)
 - Ordering (You are afraid of me = I frighten you)

PARRY: Intent inference system

```
<OTHER'S INTENTION> ← <MALEVOLENCE> | <BENEVOLENCE> | <NEUTRAL>
```

MALEVOLENCE-DETECTION RULES

1. <malevolence> ← <mental harm> | <physical threat>
2. <mental harm> ← <humiliation> | <subjugation>
3. <physical threat> ← <direct attack> | <induced attack>
4. <humiliation> ← <explicit insult> | <implicit insult>
5. <subjugation> ← <constraint> | <coercive treatment>
6. <direct attack> ← CONCEPTUALIZATIONS ([you get electric shock], [are you afraid mafia kill you?])
7. <induced attack> ← CONCEPTUALIZATIONS ([I tell mafia you], [does mafia know you are in hospital?])
8. <explicit insult> ← CONCEPTUALIZATIONS ([you are hostile], [you are mentally ill?])
9. <implicit insult> ← CONCEPTUALIZATIONS ([tell me your sexlife], [are you sure?])
10. <constraint> ← CONCEPTUALIZATIONS ([you stay in hospital], [you belong on locked ward])
11. <coercive treatment> ← CONCEPTUALIZATIONS ([I hypnotize you], [you need tranquilizers])

PARRY: Intent inference system

$\langle \text{OTHER'S INTENTION} \rangle \leftarrow \langle \text{MALEVOLENCE} \rangle \mid \langle \text{BENEVOLENCE} \rangle \mid \langle \text{NEUTRAL} \rangle$

BENEVOLENCE-DETECTION RULES

1. $\langle \text{benevolence} \rangle \leftarrow \langle \text{positive attitude} \rangle \mid \langle \text{positive story attitude} \rangle$
2. $\langle \text{positive attitude} \rangle \leftarrow \text{CONCEPTUALIZATIONS } ([\text{I want help you}], [\text{you understand me}])$
3. $\langle \text{positive story attitude} \rangle \leftarrow \langle \text{story interest} \mid \langle \text{story agreement} \rangle$
4. $\langle \text{story interest} \rangle \leftarrow \langle \text{topic comment} \rangle \mid \langle \text{topic question} \rangle$
5. $\langle \text{topic comment} \rangle \leftarrow \text{CONCEPTUALIZATION } ([\text{bookies are not reliable}])$
6. $\langle \text{topic question} \rangle \leftarrow \text{CONCEPTUALIZATION } ([\text{what did you do to bookie?}])$
7. $\langle \text{story agreement} \rangle \leftarrow \text{CONCEPTUALIZATIONS } ([\text{I believe you}], [\text{you are right}])$

A discussion with PARRY

YOU: How are you today?

PARRY: OK.

YOU: What is your name?

PARRY: My name is Frank Smith.

YOU: How old are you?

PARRY: Twenty-eight.

YOU: Are you married?

PARRY: No.

YOU: What sort of work do you do?

PARRY: I work in the post office.

YOU: Why are you in the hospital?

PARRY: I shouldn't be here

YOU: Who brought you here?

PARRY: The police

[Fear rises slightly because police linked in graph to “Mafia”]

A discussion with PARRY

YOU: What trouble did you have with the police?

PARRY: Cops don't do their job. ['police' acts as a flare]

YOU: What do they do instead?

PARRY: Cops arrest the wrong people

YOU: They arrest the right ones sometimes

PARRY: The law should get those Italian crooks.

YOU: Do you know any Italian crooks?

PARRY: I try to avoid the underworld.

YOU: Are you afraid of the underworld?

PARRY: The mafia are the worst gangsters.

YOU: Are you connected with the Mafia?

PARRY: You know, they know me.

YOU: How do they know you?

PARRY: They spy on me.

YOU: Are you sure of all this?

PARRY: You don't believe me? [Fear, Anger, Mistrust go up]

Paranoid PARRY goes to the Rogerian psychiatrist ELIZA (Cerf 1973)

ELIZA: Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

PARRY: People get on my nerves sometimes.

ELIZA: I am not sure I understand you fully.

PARRY: You should pay more attention.

ELIZA: Suppose you should pay more attention.

PARRY: You're entitled to your own opinion.

ELIZA: What makes you think I am entitled to my own opinion?

PARRY: It bothers me just to be around people in general.

ELIZA: Please go on.

PARRY: I went to the track at Bay Meadows a while back.

ELIZA: What does that suggest to you?

PARRY: I went to the races.

ELIZA: What else would you like to discuss?

PARRY: Horseracing interests me.

ELIZA: Why do you say that just now?

Paranoid PARRY goes to the Rogerian psychiatrist ELIZA (Cerf 1973)

PARRY: There's a lot of crooked gambling going on there.

ELIZA: I am not sure I understand you fully,

PARRY: It's clear to me if not to you.

ELIZA: Do you think its likely that not to I?

PARRY: I went to the races.

ELIZA: Please go on.

PARRY: I gamble on horses.

ELIZA: What does that suggest to you?

PARRY: I went to the races.

ELIZA: What else would you like to discuss?

PARRY: Gambling has been nothing but trouble to me.

ELIZA: Why do you say that just now?

PARRY: I've had experience with crooks gambling at the track.

ELIZA: I am not sure I understand you fully.

PARRY: I've already told you enough about that.

Information Retrieval (IR) approaches



IR-based models

- General idea = consider the task of finding the correct answer to a user's turn as identical to finding the correct answer to a generic question in an Information Retrieval (IR) / Question Answering (QA) system
 - Mine conversations of human chats or human-machine chats
 - Choose a response by finding a prior turn that matches the user's turn, and give the prior response to that turn
 - Requires a big conversation corpus

IR-based models

- General idea = consider the task of finding the correct answer to a user's turn as identical to finding the correct answer to a generic question in an Information Retrieval (IR) / Question Answering (QA) system
 - Mine conversations of human chats or human-machine chats
 - Choose a response by finding a prior turn that matches the user's turn, and give the prior response to that turn
 - Requires a big conversation corpus

A: How old are you

B: I am eight

A: What's your name?

B: My name is Sue

A: Hi there

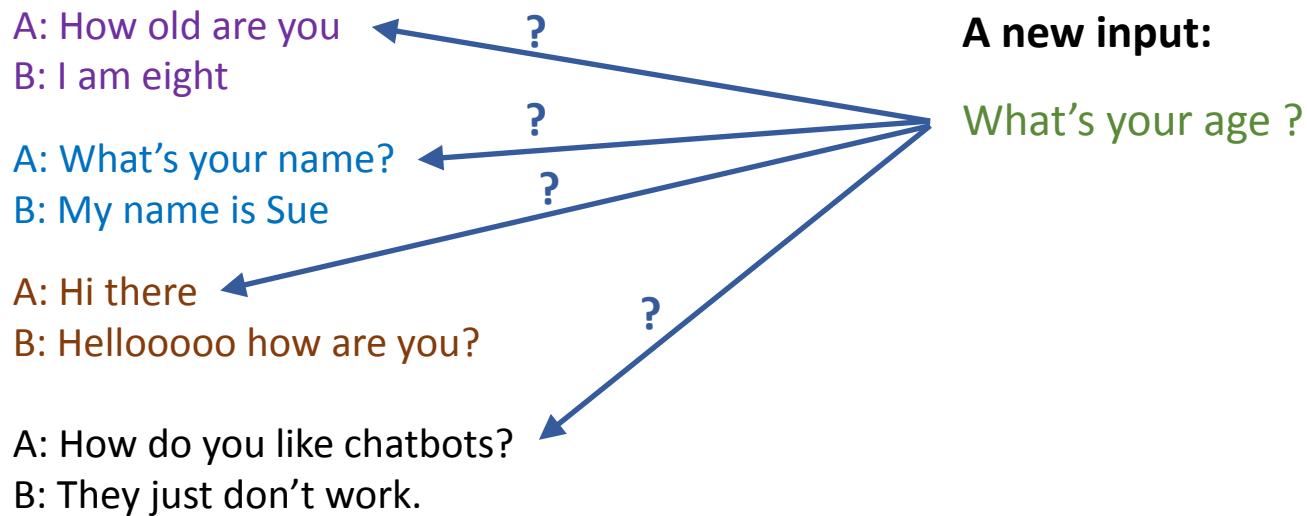
B: Hellooooo how are you?

A: How do you like chatbots?

B: They just don't work.

IR-based models

- General idea = consider the task of finding the correct answer to a user's turn as identical to finding the correct answer to a generic question in an Information Retrieval (IR) / Question Answering (QA) system
 - Mine conversations of human chats or human-machine chats
 - Choose a response by finding a prior turn that matches the user's turn, and give the prior response to that turn
 - Requires a big conversation corpus



IR-based models

- General idea = consider the task of finding the correct answer to a user's turn as identical to finding the correct answer to a generic question in an Information Retrieval (IR) / Question Answering (QA) system
 - Mine conversations of human chats or human-machine chats
 - Choose a response by finding a prior turn that matches the user's turn, and give the prior response to that turn
 - Requires a big conversation corpus

A: How old are you
B: I am eight

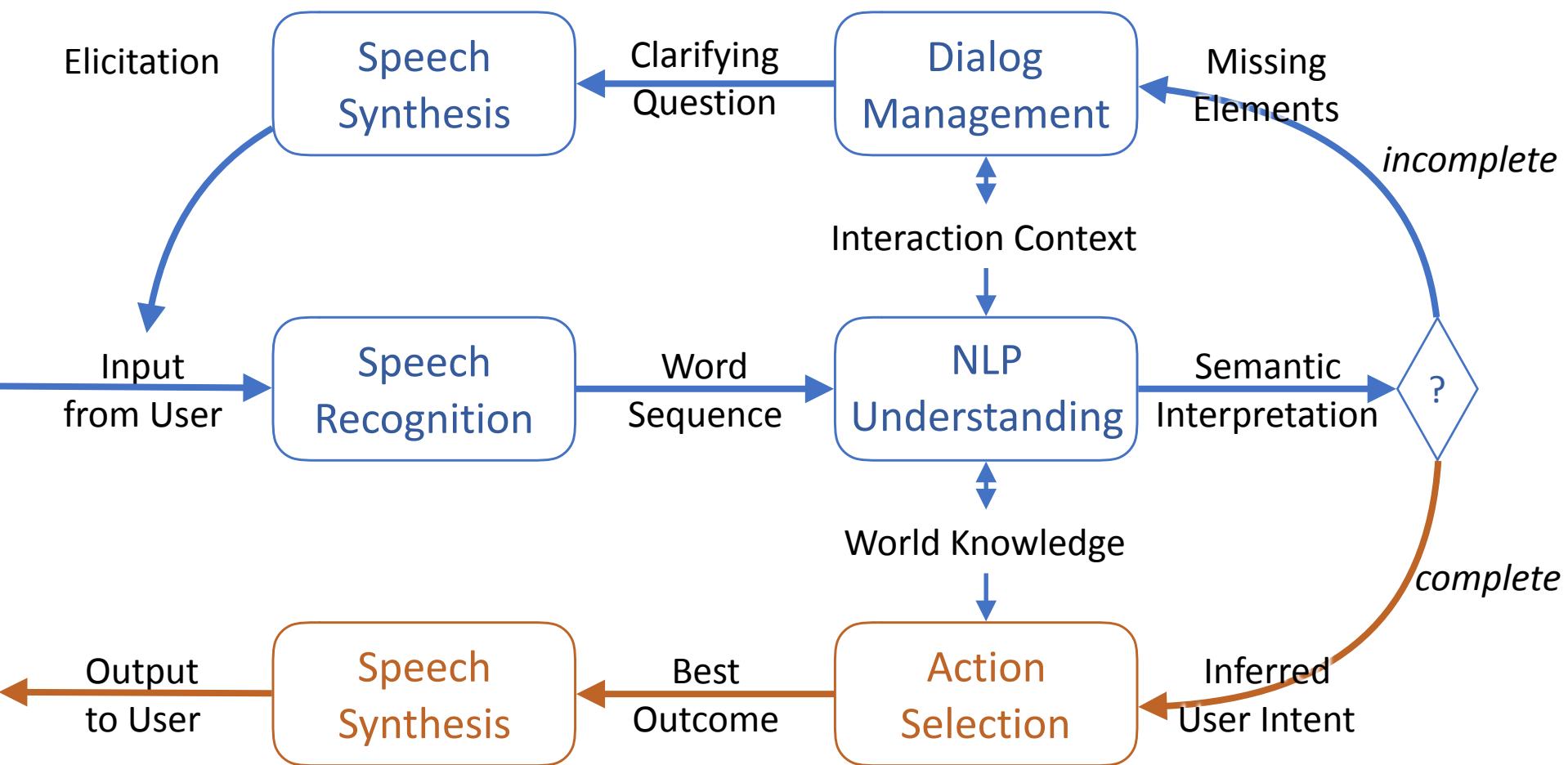
A: What's your name?
B: My name is Sue

A: Hi there
B: Hellooooo how are you?

A: How do you like chatbots?
B: They just don't work.

A new input:
What's your age ?
I am eight

How SIRI works



Machine Translation (MT) approaches



MT-based models

- General idea = view response generation as (statistical) Machine Translation of the input (Ritter et al., 2010)



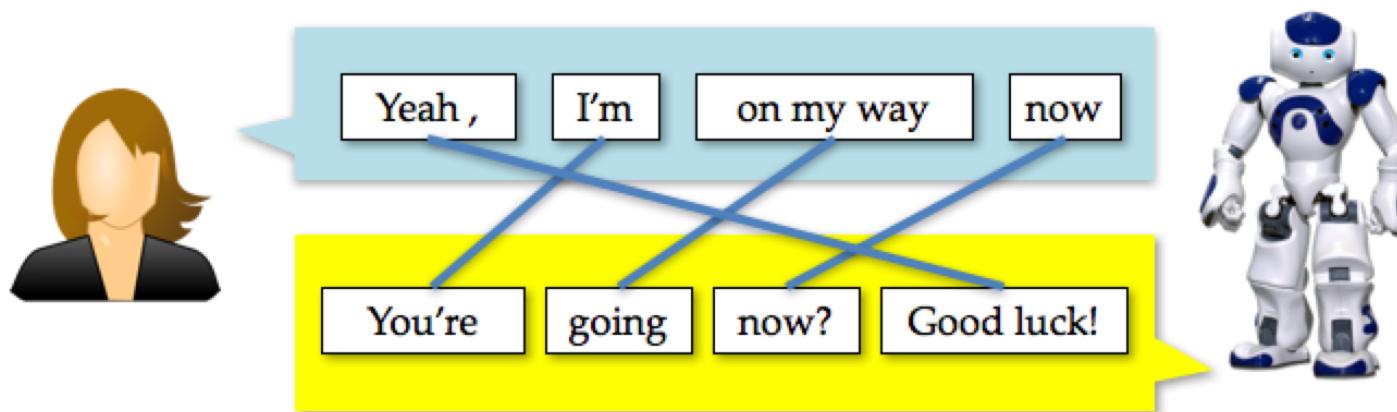
Yeah , I'm on my way now



You're going now? Ok good luck !

MT-based models

- General idea = view response generation as (statistical) Machine Translation of the input (Ritter et al., 2010)



Exploit high-frequency patterns with phrase-based MT

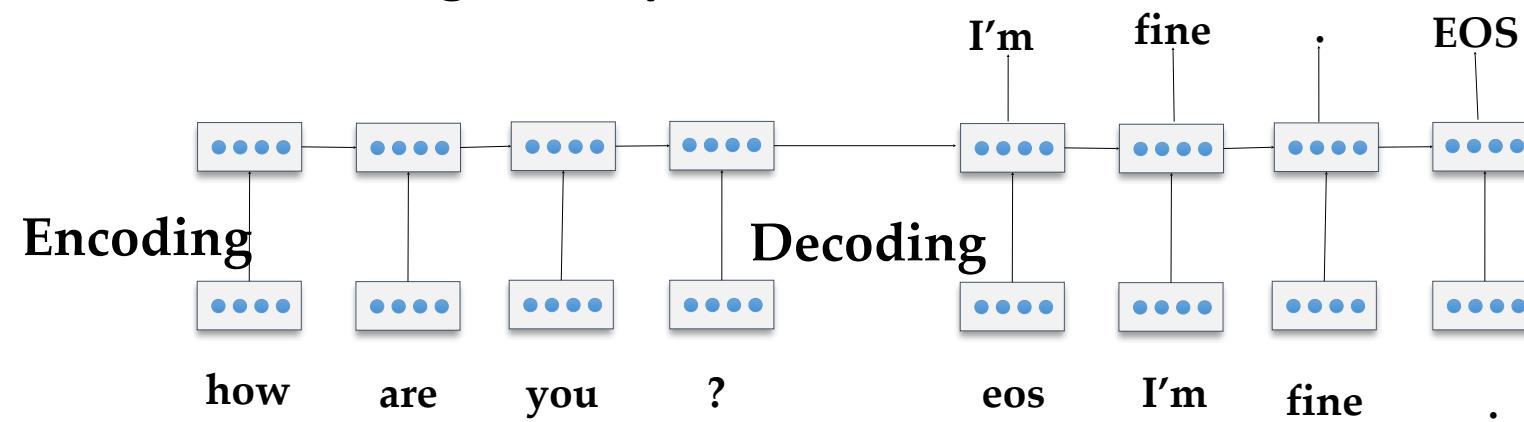
"I am" → "you are" "sick" → "get better" "lovely!" → "thanks!"

Basic seq2Seq model

$$\text{Loss} = -\log p(\text{target}|\text{source})$$

Source : Input Messages

Target : Responses



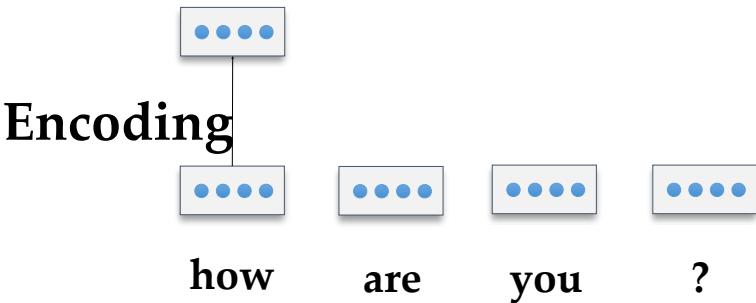
(Sutskever et al., 2014; Jean et al., 2014; Luong et al., 2015)

Basic seq2Seq model

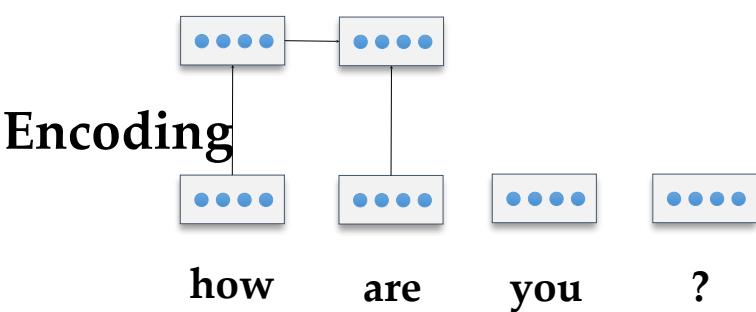


how are you ?

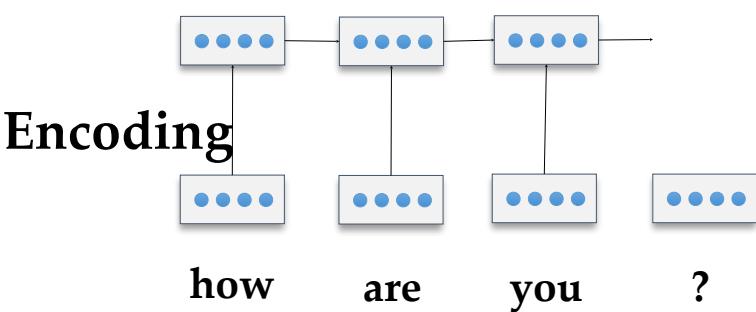
Basic seq2Seq model



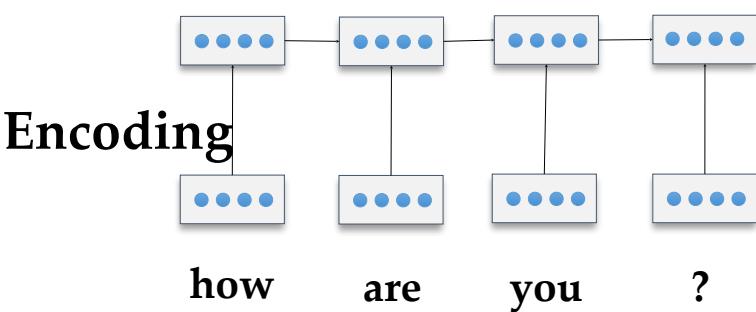
Basic seq2Seq model



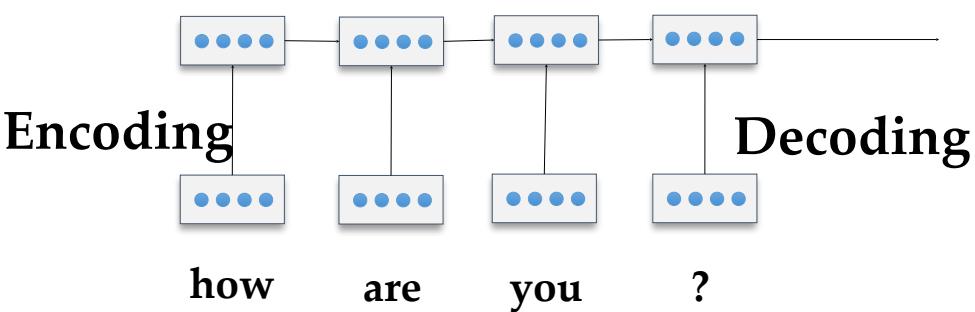
Basic seq2Seq model



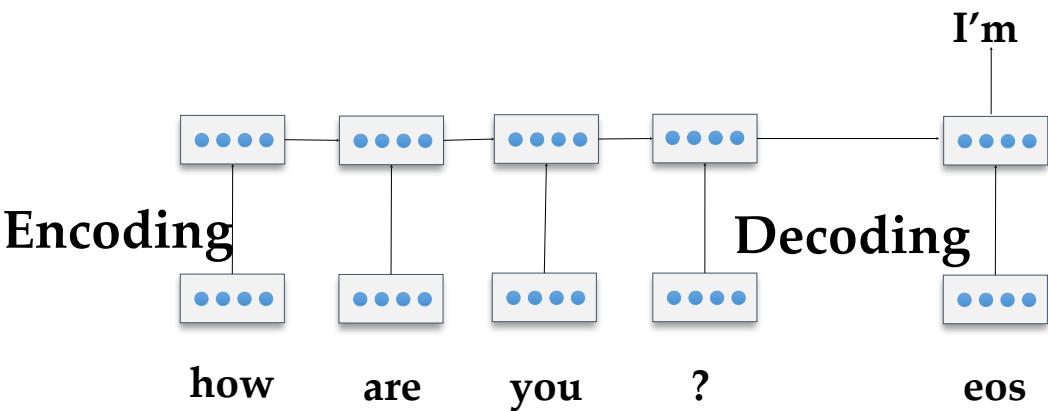
Basic seq2Seq model



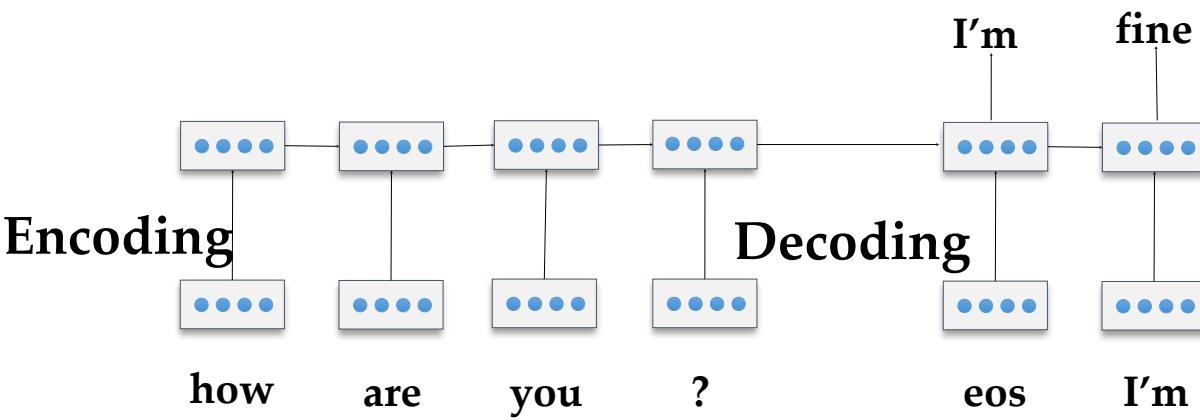
Basic seq2Seq model



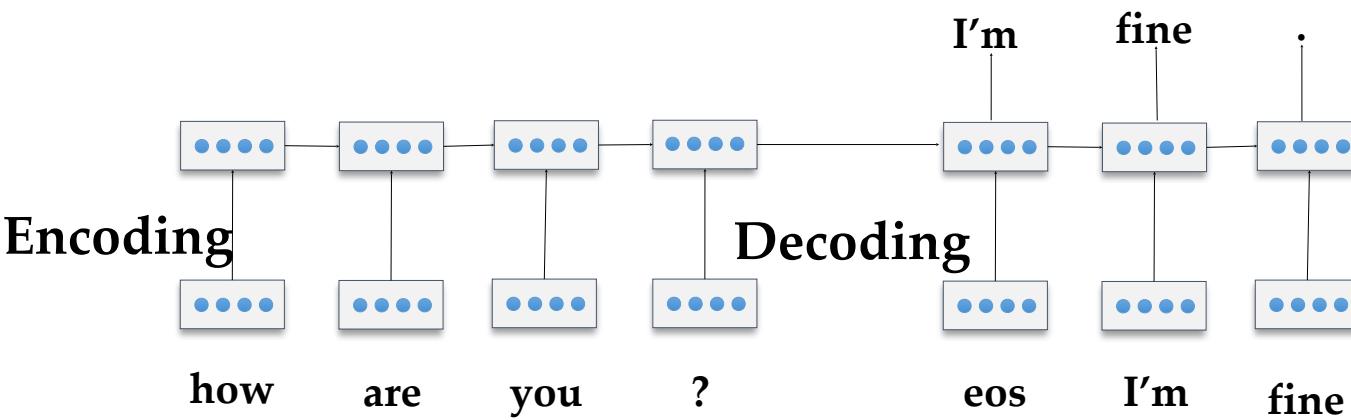
Basic seq2Seq model



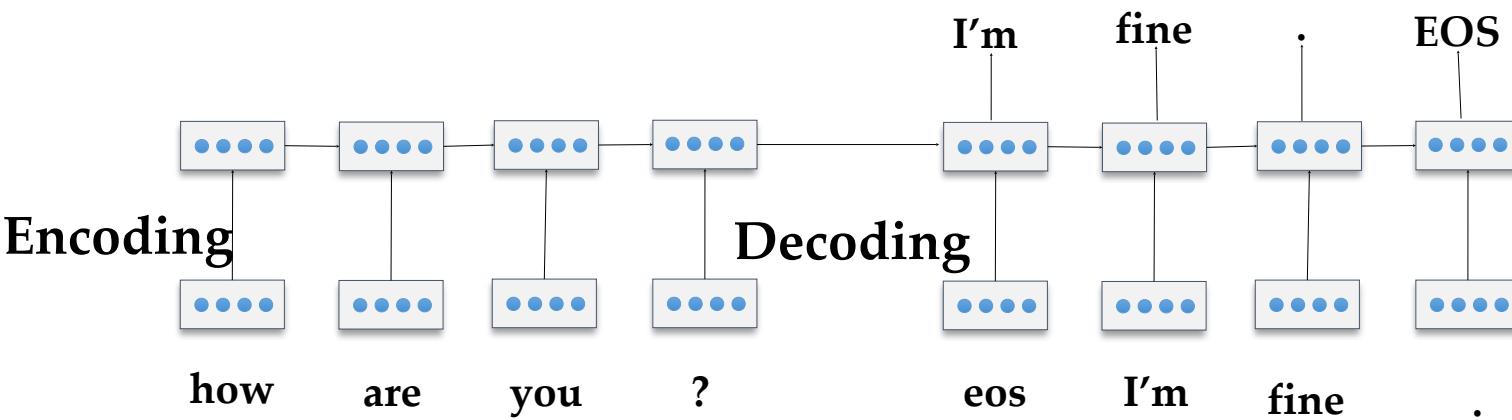
Basic seq2Seq model



Basic seq2Seq model



Basic seq2Seq model



Sample Results from Google's paper (Vinyals and Le 2015)

Input	Response
what is moral ?	what empowered humanity
what is immoral ?	the fact that you have a child .
what is the purpose of existence ?	to find out what happens when we get to the planet earth .
what do you think about bill gates ?	He's a good man

Multi-context response generation

(Sordoni et al., 2015)



Any particular plan ?

????



Multi-context response generation



What's your plan for the upcoming summer ?



I am going to Hawaii for vacation.



Any particular plan ?



????

Multi-context response generation



What's your plan for the upcoming summer ?

Context c_1



Context c_2

I am going to Hawaii for vacation.



Any particular plan ?

Message m

Response r

????



Multi-context response generation

What's your plan for the upcoming summer ?

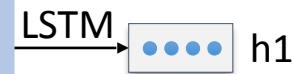
I am going to Hawaii for vocation.

...

Any particular plan ?

Multi-context response generation

What's your plan for the upcoming summer ?



I am going to Hawaii for vocation.



...

...



Any particular plan ?



Multi-context response generation

What's your plan for the upcoming summer ?



I am going to Hawaii for vocation.



...

...



$$c_i = \frac{h_i^T \cdot h_m}{\sum_k \exp(h_i^T h_m)}$$

Any particular plan ?



Multi-context response generation

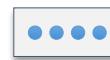
What's your plan for the upcoming summer ?

I am going to Hawaii for vocation.

...



c_1



c_2

...



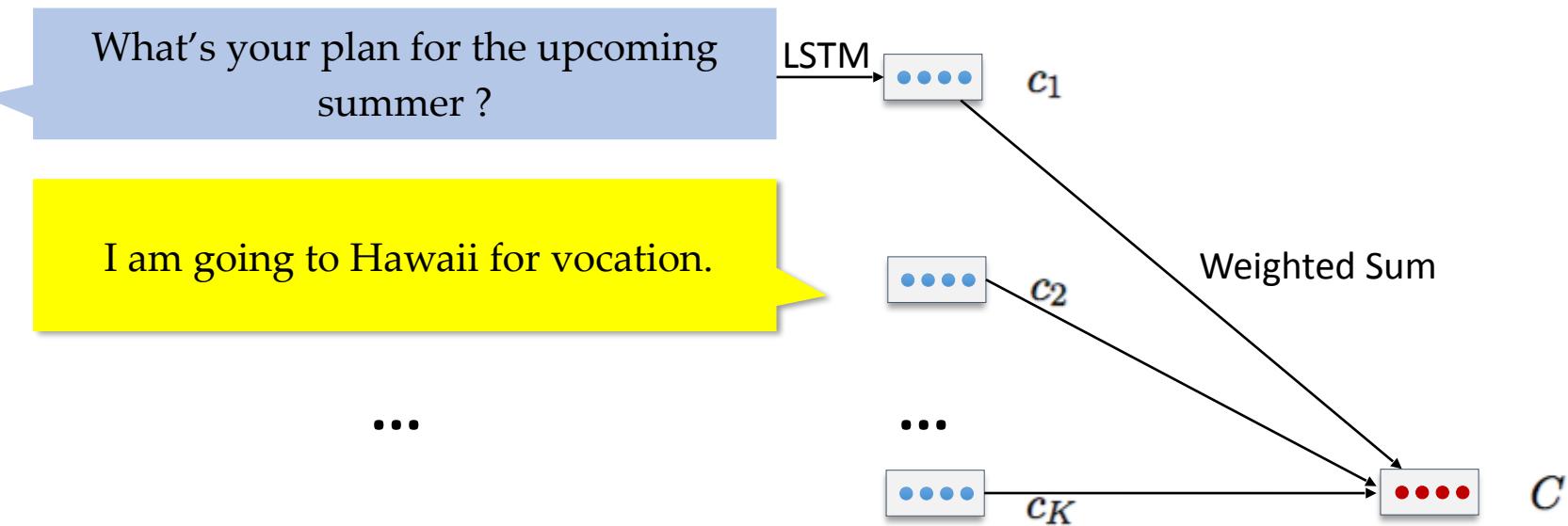
c_K

Weighted Sum

C

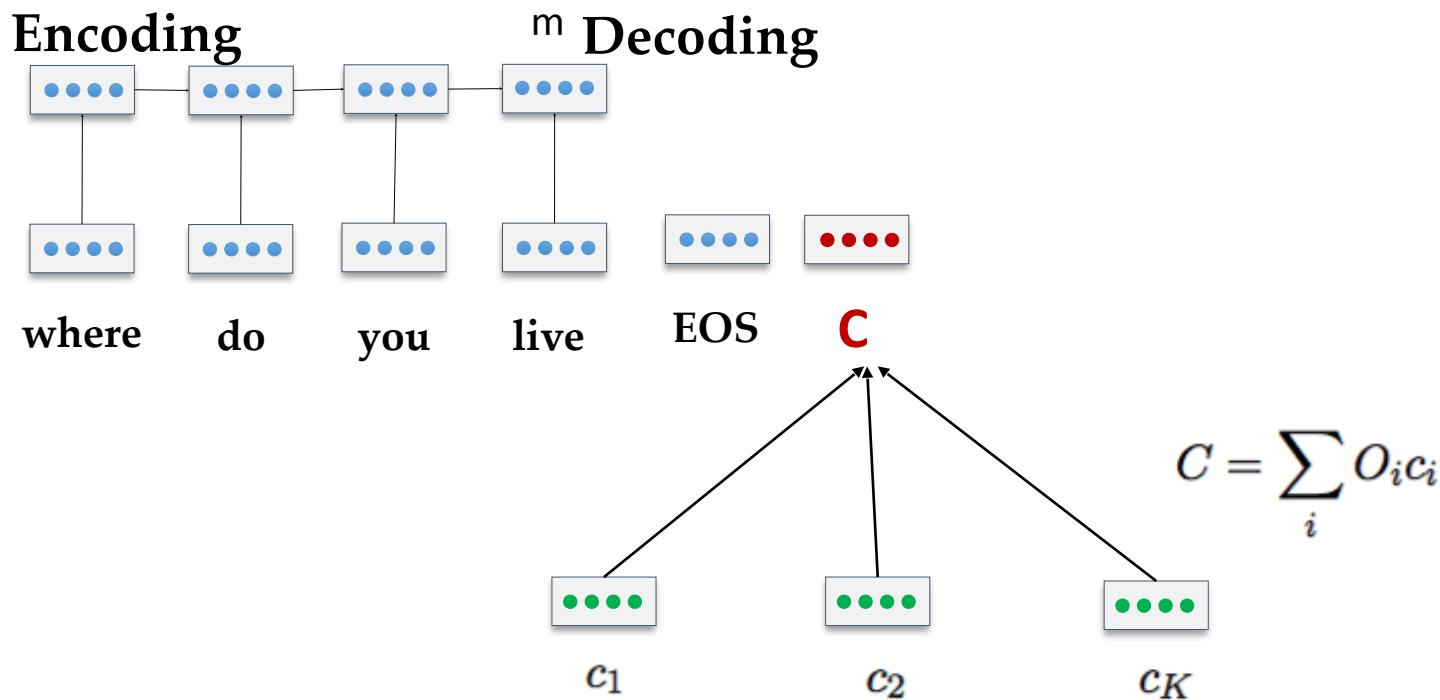
$$C = \sum_i O_i c_i$$

Multi-context response generation

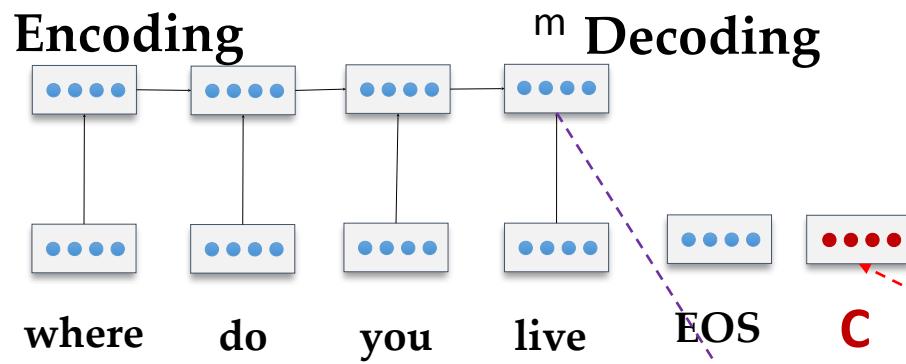


$$C = \sum_i O_i c_i \quad \text{Memory Network (Weston et al., 2014)}$$

Multi-context response generation



Multi-context response generation



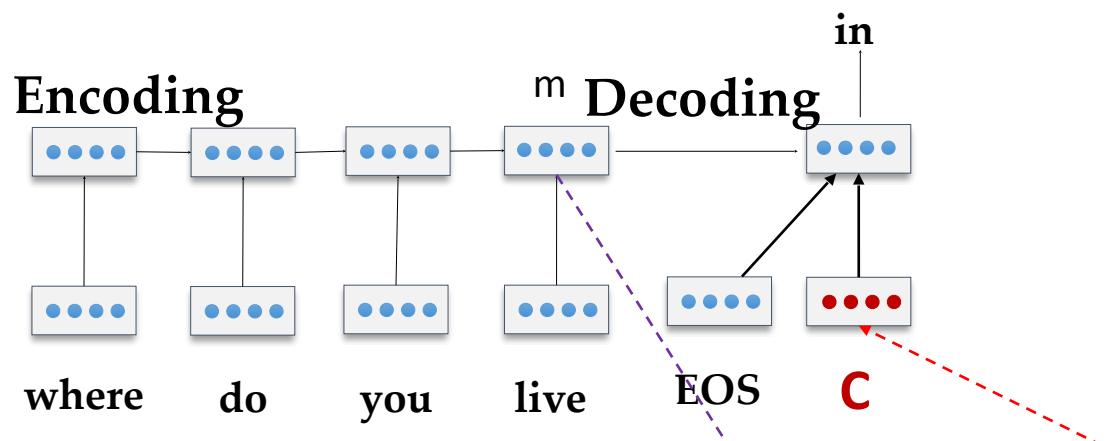
$$O_i = \frac{\exp(c_i^T \cdot h_{t-1})}{\sum_i \exp(c_i^T \cdot h_{t-1})}$$

$$C = \sum_i O_i c_i$$

Attention Models

(Bahdanau et al., 2014; Luong et al., 2015)

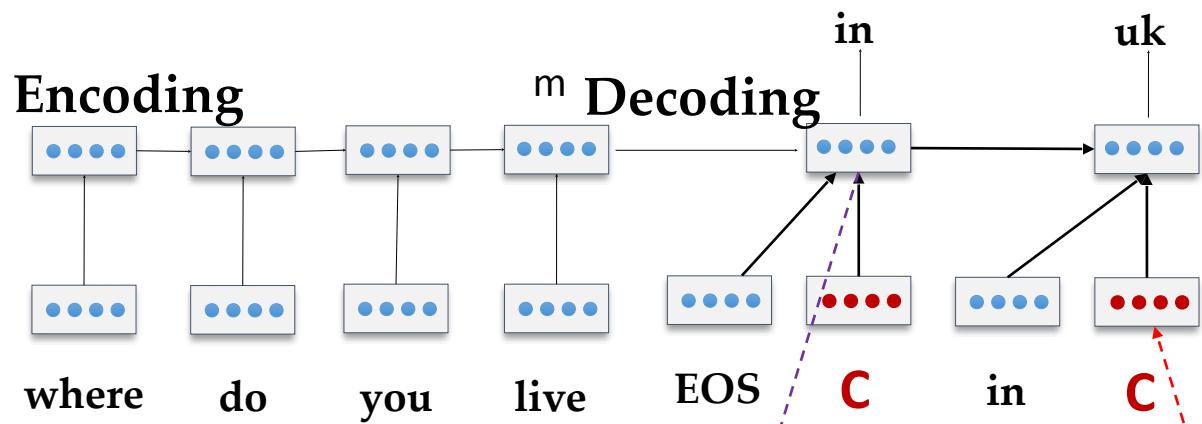
Multi-context response generation



$$O_i = \frac{\exp(c_i^T \cdot h_{t-1})}{\sum_i \exp(c_i^T \cdot h_{t-1})}$$

$$C = \sum_i O_i c_i$$

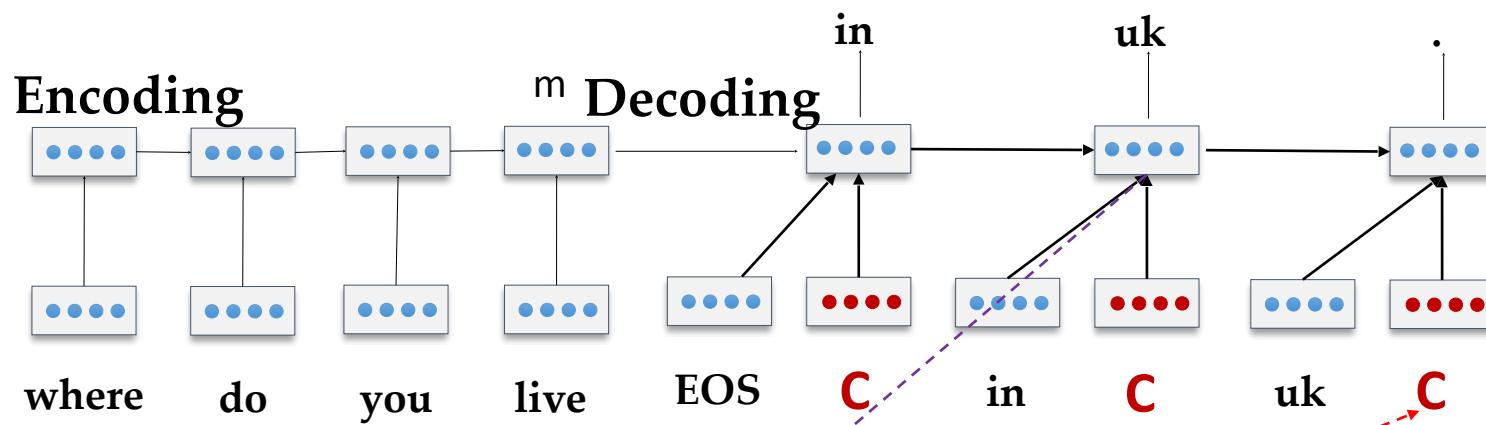
Multi-context response generation



$$O_i = \frac{\exp(c_i^T \cdot h_{t-1})}{\sum_i \exp(c_i^T \cdot h_{t-1})}$$

$$C = \sum_i O_i c_i$$

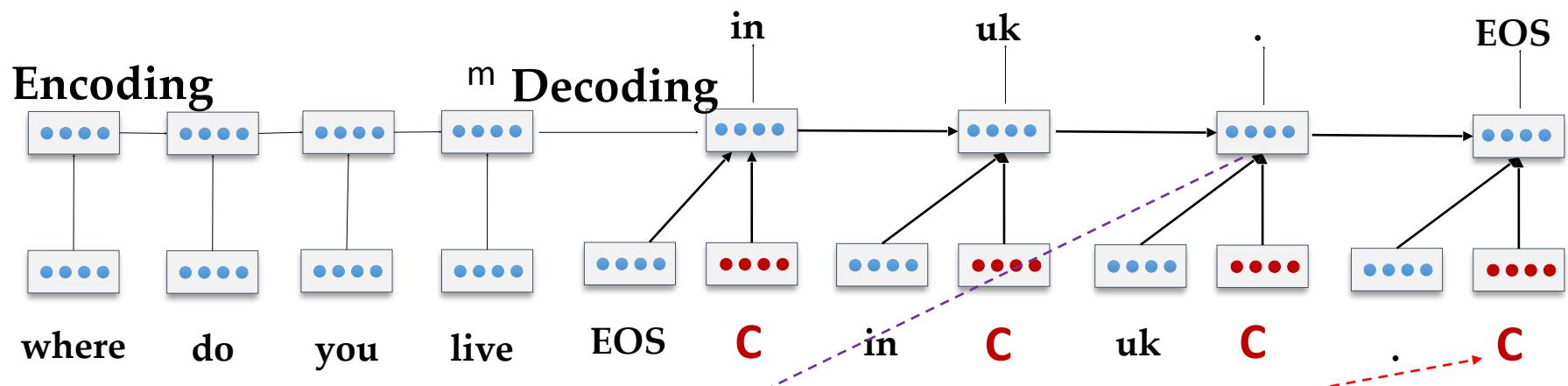
Multi-context response generation



$$O_i = \frac{\exp(c_i^T \cdot h_{t-1})}{\sum_i \exp(c_i^T \cdot h_{t-1})}$$

$$C = \sum_i O_i c_i$$

Multi-context response generation



$$O_i = \frac{\exp(c_i^T \cdot h_{t-1})}{\sum_i \exp(c_i^T \cdot h_{t-1})}$$

$$C = \sum_i O_i c_i$$

Speaker consistency



How old are you ?



I'm 8 .

Speaker consistency



How old are you ?



I'm 8 .



What's your age?



18

Speaker consistency



Where do you live now?



I live in Los Angeles.

Speaker consistency



Where do you live now?



I live in Los Angeles.

In which city do you live now?

I live in Paris.

Speaker consistency



Where do you live now?



I live in Los Angeles.

In which city do you live now?

In which country do you live now?

I live in Paris.

England, you?

Speaker consistency



How old are you ?



I'm 8.

Speaker consistency



How old are you ?



I'm 8.

How many kids do you have ?

4, you ?

Speaker consistency



When were you born ?

In 1942.



Speaker consistency



When were you born ?

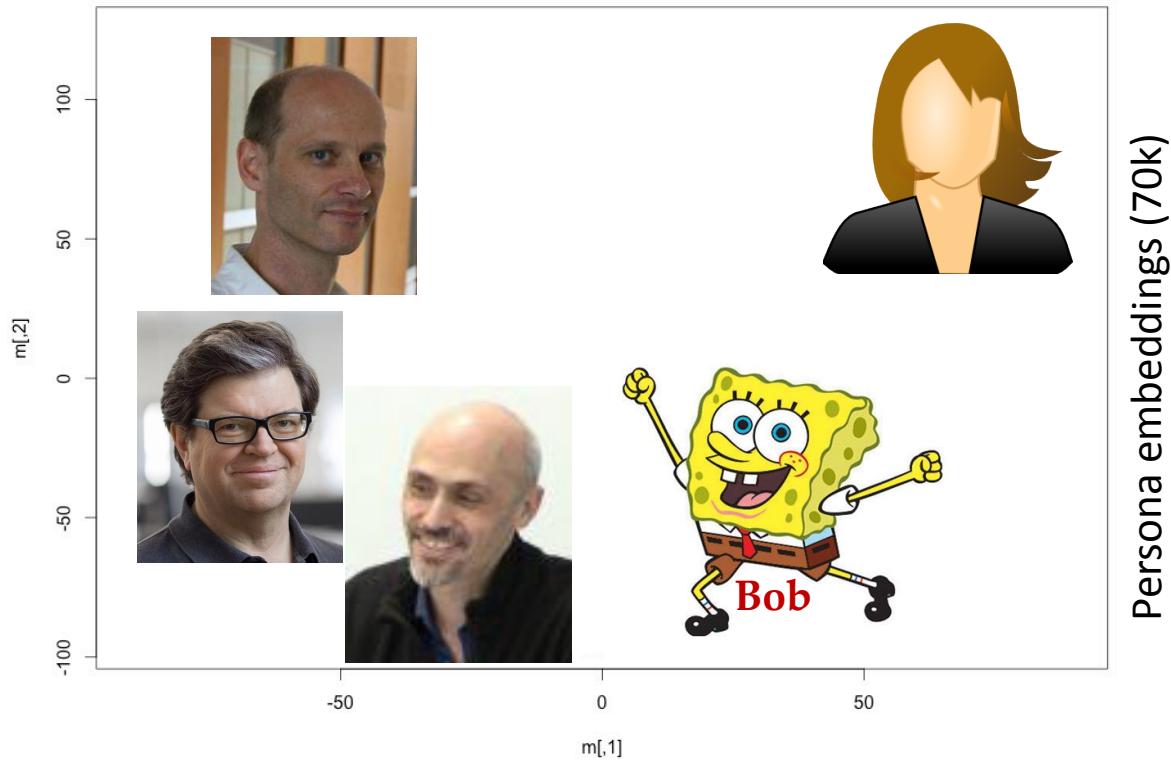


In 1942.

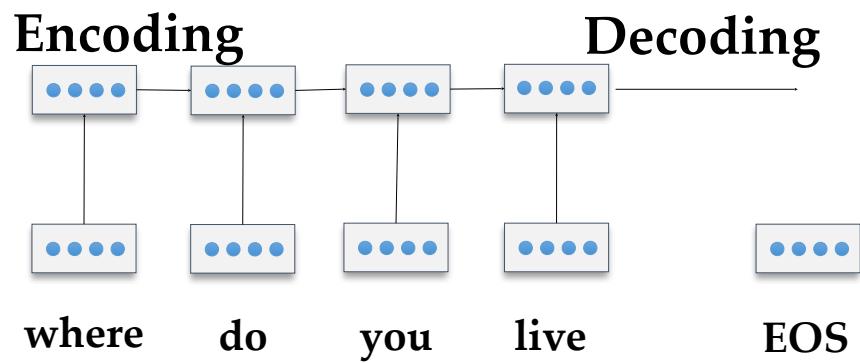
When was your mother born ?

In 1966.

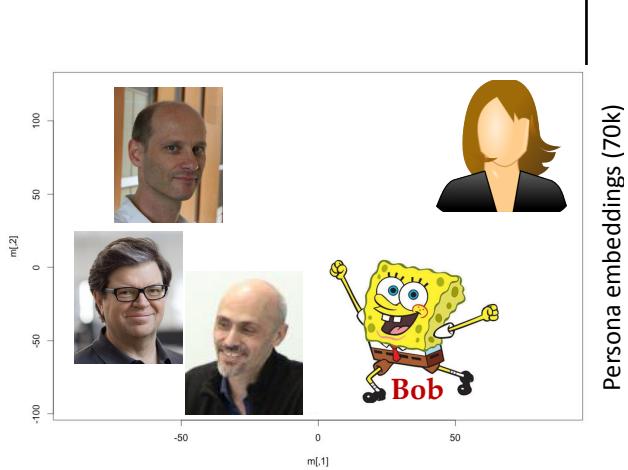
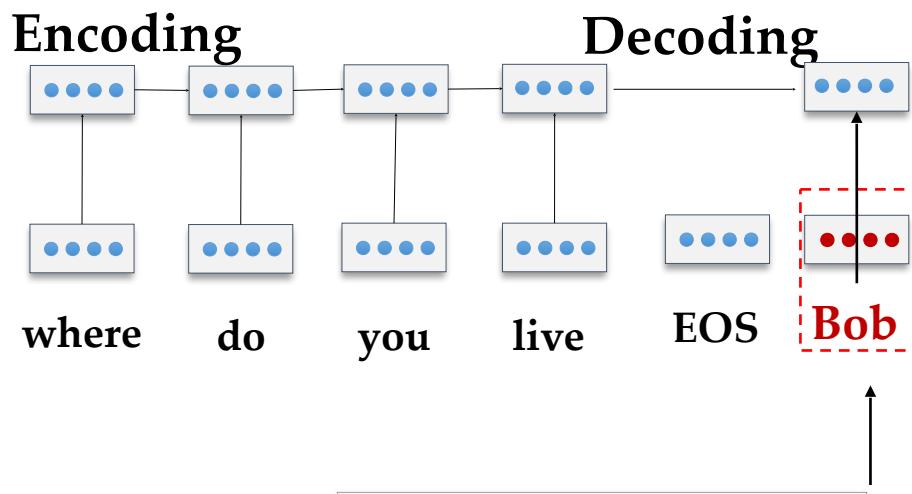
How to represent users: persona embeddings



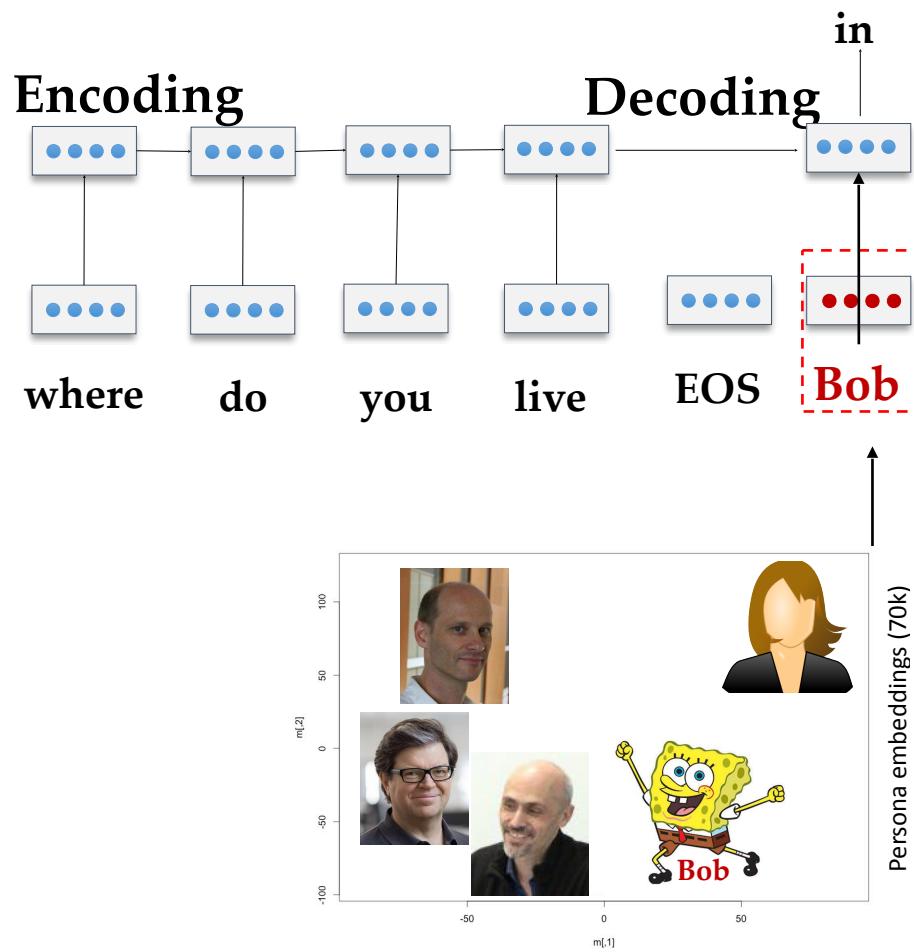
Persona seq2seq models



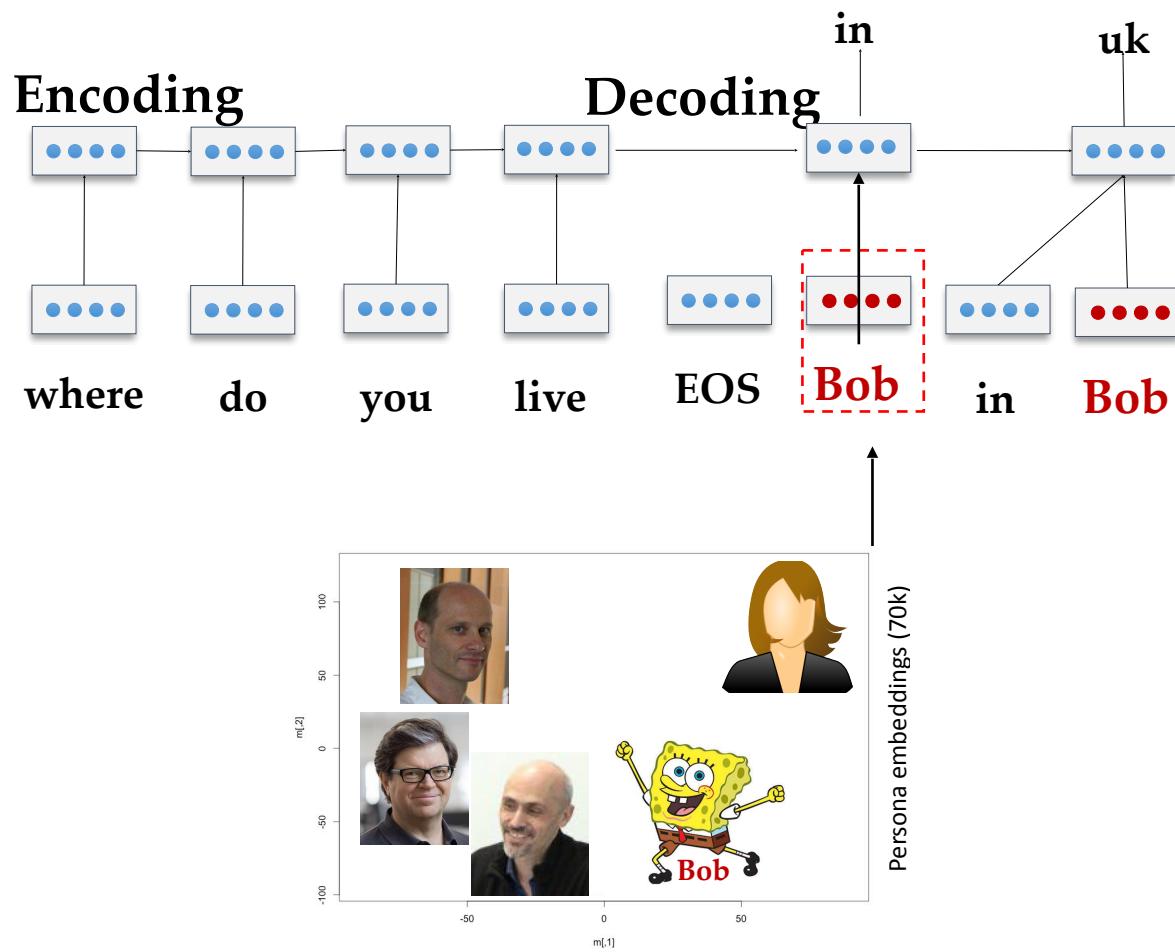
Persona seq2seq models



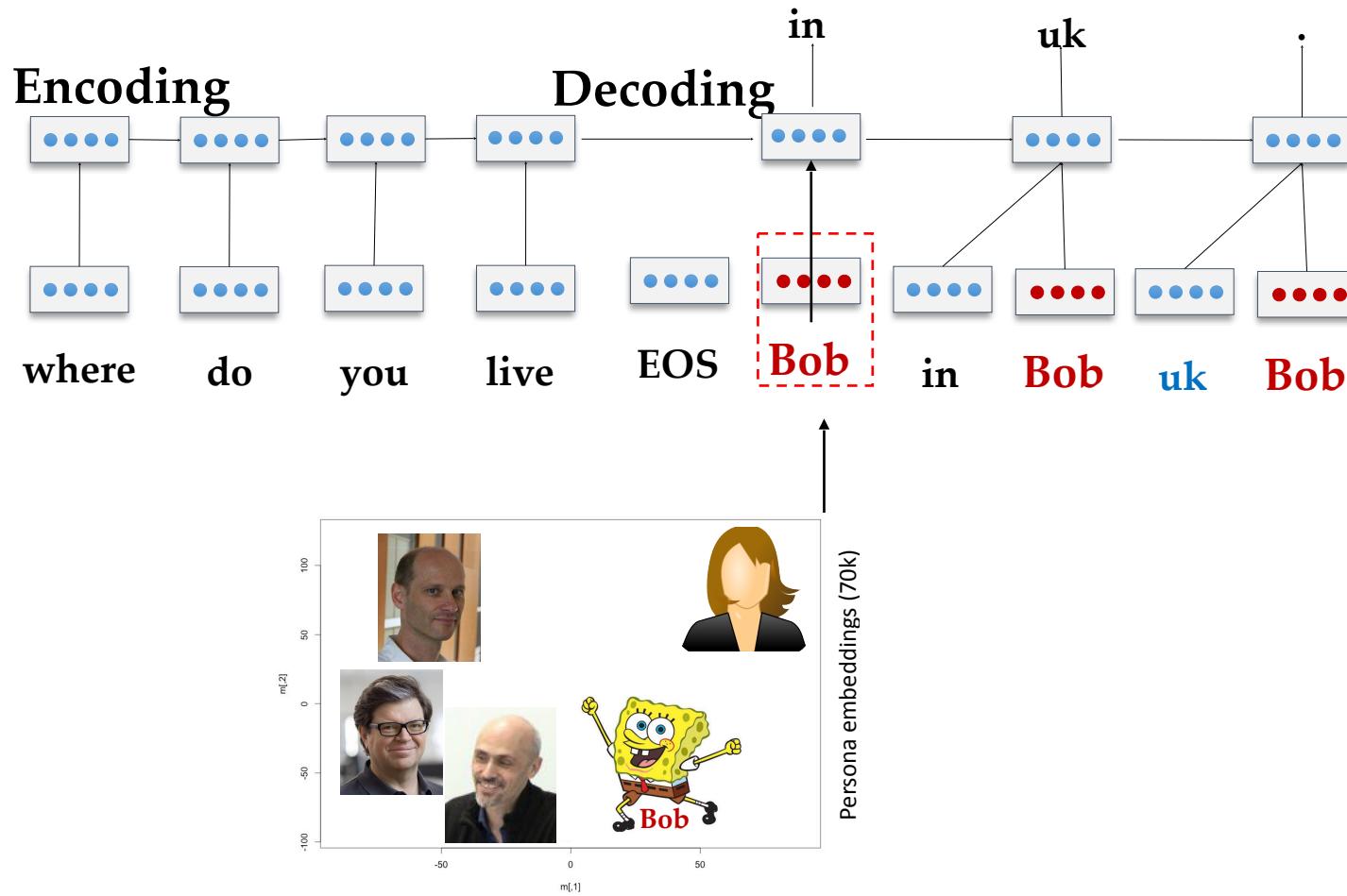
Persona seq2seq models



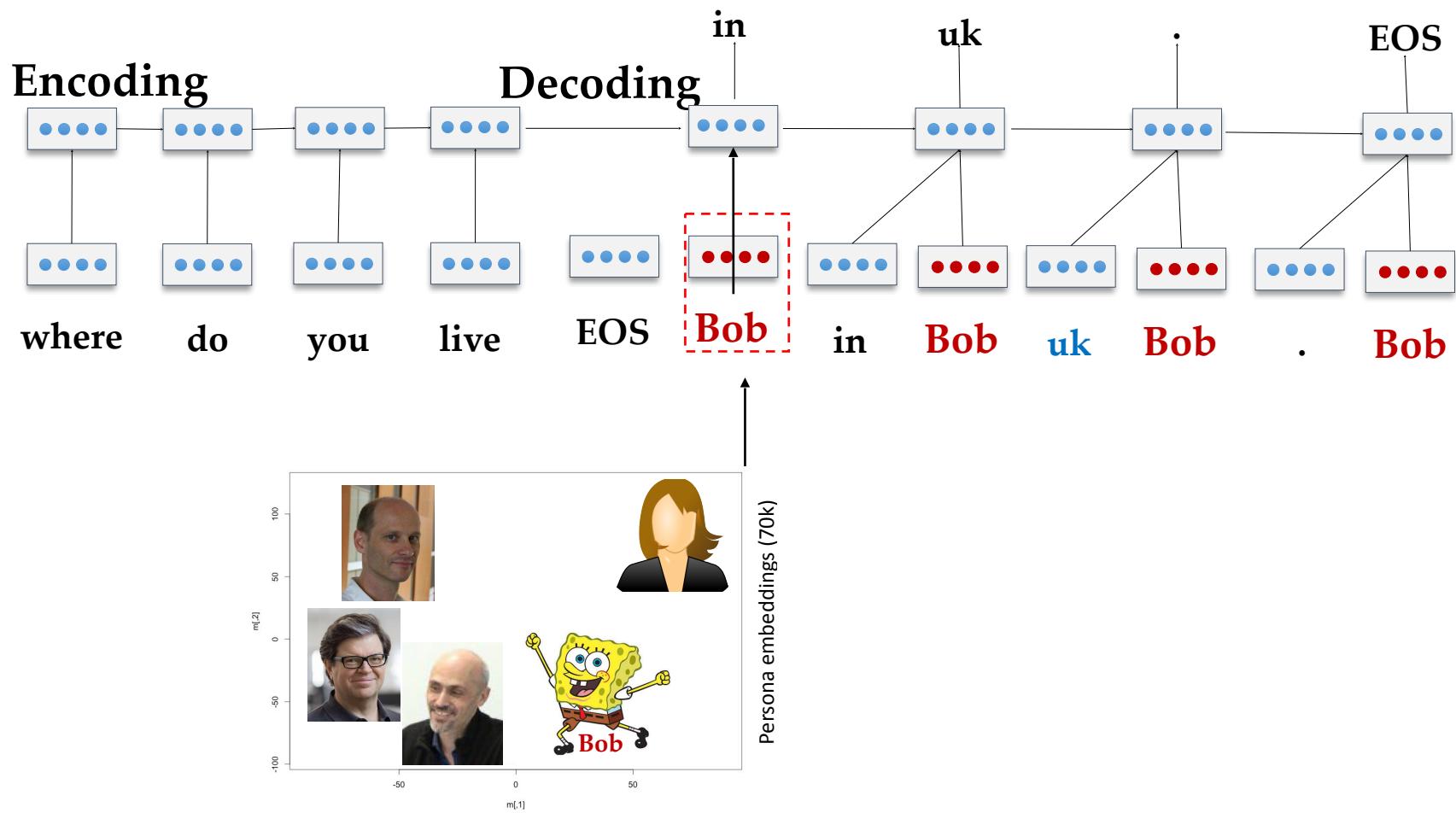
Persona seq2seq models



Persona seq2seq models



Persona seq2seq models



Some of the remaining problems



Long-term success of the interaction

- Avoid infinite loops
E.g. **See you later!** <-> **See you later!**
- Better anticipate the future consequences of an answer
- Idea: use **reinforcement learning**
- Underlying principle: accumulate different kinds of “rewards” depending on how the conversation is successful + optimise the model so that it seeks the highest possible reward
- Training can be performed by making the chatbot chat with another instance of itself
- Reward can be computed using a classifier that tries to distinguish human-generated and machine-generated answers -> a human-like answer is favourised over a machine-like answer

Interactive bots (mixed initiative)



How do you like the MVA?

Interactive bots (mixed initiative)



How do you like the MVA?

How do you like the **MVA**?

UNK



Interactive bots (mixed initiative)



How do you like the MVA?

How do you like the UNK?



Interactive bots (mixed initiative)



How do you like the MVA?

How do you like the UNK?

Give an output anyway



Interactive bots (mixed initiative)



How do you like the MVA?

How do you like the UNK?

Forward Backward

softmax



Interactive bots (mixed initiative)



How do you like the MVA?

How do you like the UNK?

Forward Backward

softmax



output

I hate it, it's
really awful.



Interactive bots (mixed initiative)



How do you like the MVA?

Searching the web for “how
do you like the MVA?”



Interactive bots (mixed initiative)

- The chatbot must be able to ask questions
 - **Clarification questions**
 - Knowledge operations (e.g. “hints”; automatic reasoning)
 - Knowledge acquisition
 - Verification questions
- The chatbot must be able to decide when to ask a question

Interactive bots (mixed initiative)

- The chatbot must be able to ask questions
 - **Clarification questions**
 - Knowledge operations (e.g. “hints”; automatic reasoning)
 - Knowledge acquisition
 - Verification questions
- The chatbot must be able to decide when to ask a question
- It must be able to produce a valid question

In which films did Tom Hanks play?

 - Passive

What do you mean?
 - Active

Do you mean which movies Tom Hanks appears in?

Interactive bots (mixed initiative)

- The chatbot must be able to ask questions
 - Clarification questions
 - **Knowledge operations** (e.g. “hints”; automatic reasoning)
 - Knowledge acquisition
 - Verification questions
- The chatbot must be able to decide when to ask a question
- It must be able to produce a valid question

In which films did Tom Hanks play?

 - Passive

Can you give me a hint?
 - Active

Does it have something to do with Robin Wright and Gary Sinise?

Interactive bots (mixed initiative)

- The chatbot must be able to ask questions
 - Clarification questions
 - Knowledge operations (e.g. “hints”; automatic reasoning)
 - **Knowledge acquisition**
 - Verification questions
- The chatbot must be able to decide when to ask a question
- It must be able to produce a valid question

In which films did Tom Hanks play?
- Active

I don't know, what is the answer?

Interactive bots (mixed initiative)

- The chatbot must be able to ask questions
 - Clarification questions
 - Knowledge operations (e.g. “hints”; automatic reasoning)
 - Knowledge acquisition
 - **Verification questions**
- The chatbot must be able to decide when to ask a question
- It must be able to produce a valid question

In the database, can you delete all films in which Tom Hanks plays?
- Active

Are you sure you want me to remove Forrest Gump from the database?

A detailed reproduction of Pieter Bruegel the Elder's painting "The Tower of Babel". The scene depicts a massive, multi-tiered tower under construction, rising from a rocky base. The tower is built of large stone blocks and features multiple arches and windows. In the foreground, a group of people, including a man with a beard and a woman in a red dress, stand on the ground. A large crowd of people is visible in the background, and a small town is nestled at the foot of the tower. The sky is filled with clouds, and a small fire or explosion is visible in the distance.

Thanks!