

HOEL LE CAPITAINE, POLYTECH NANTES  
INFO5  
2021/2022

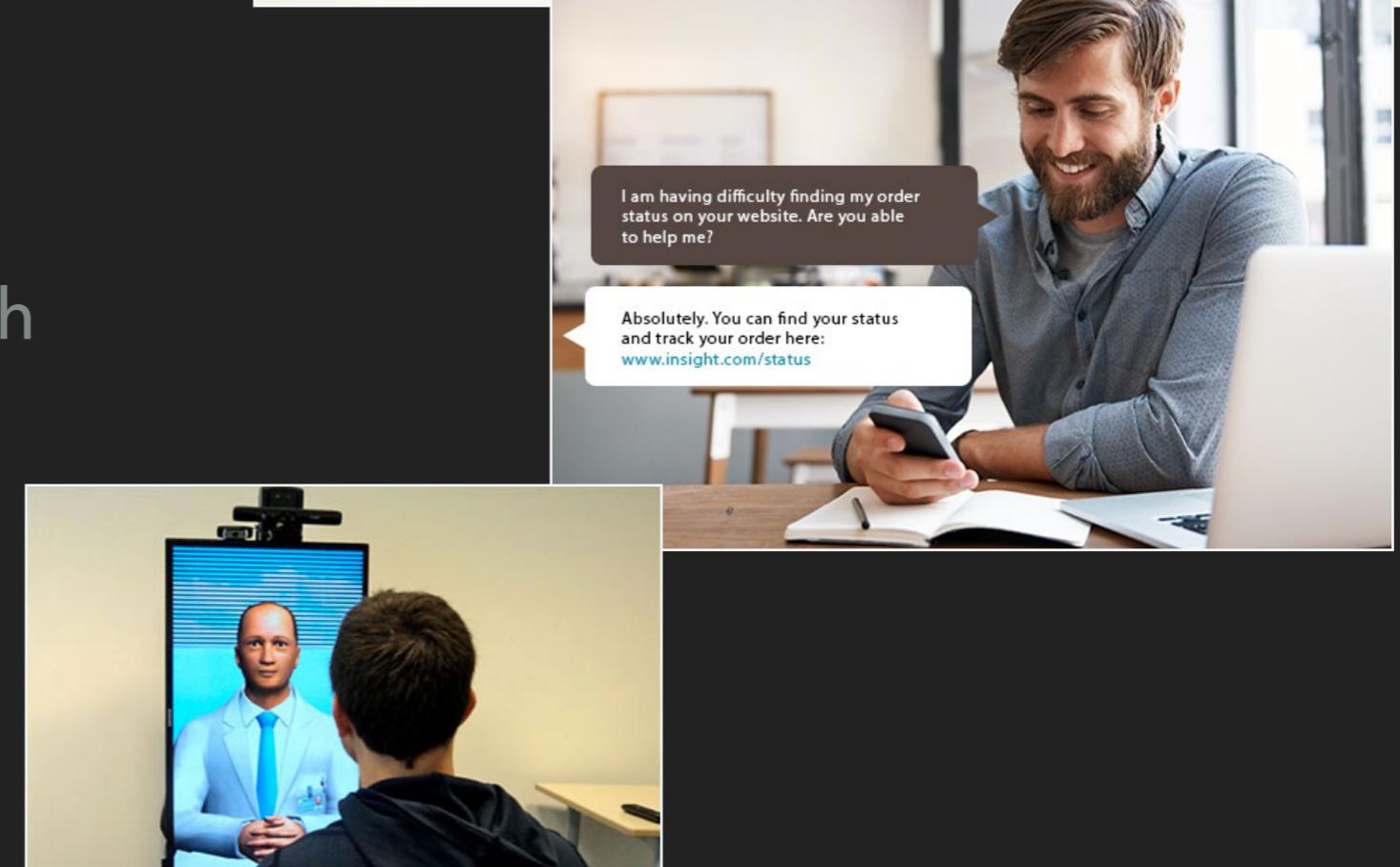
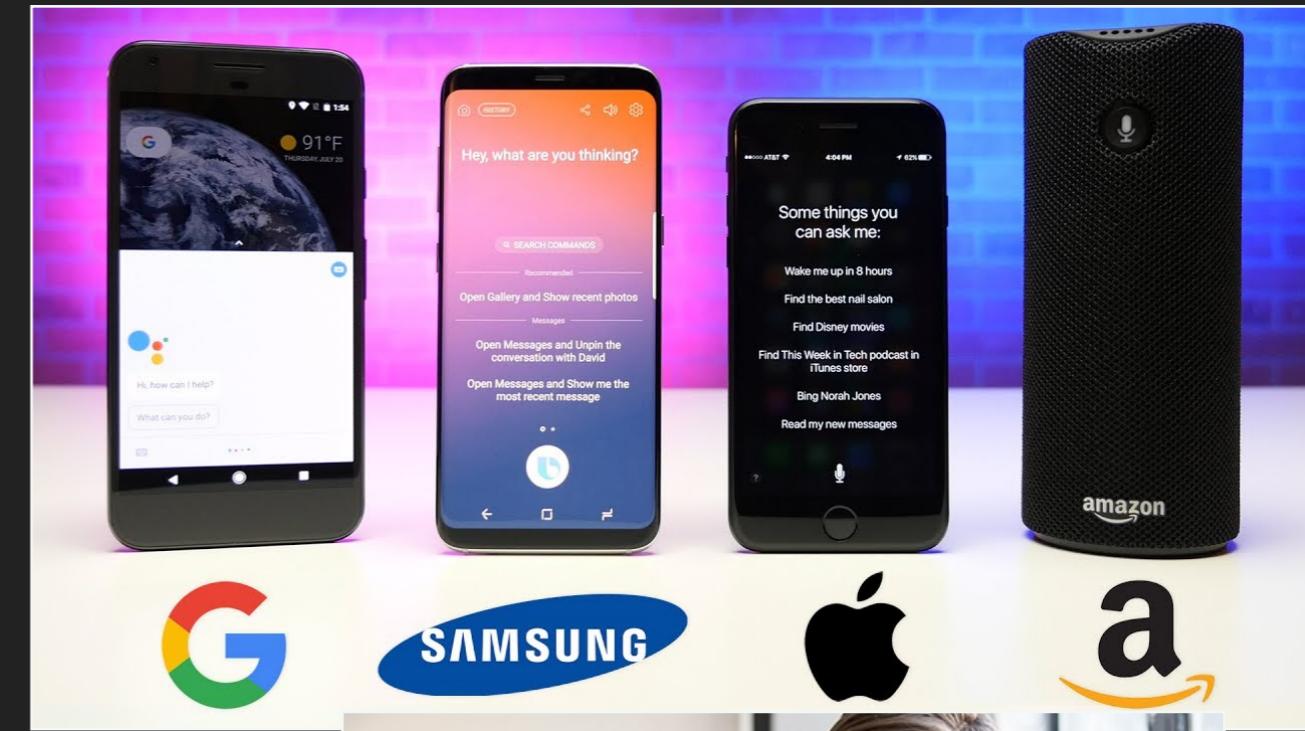
---

# CONVERSATIONAL AGENTS

courtesy of D. Jurafsky, Stanford

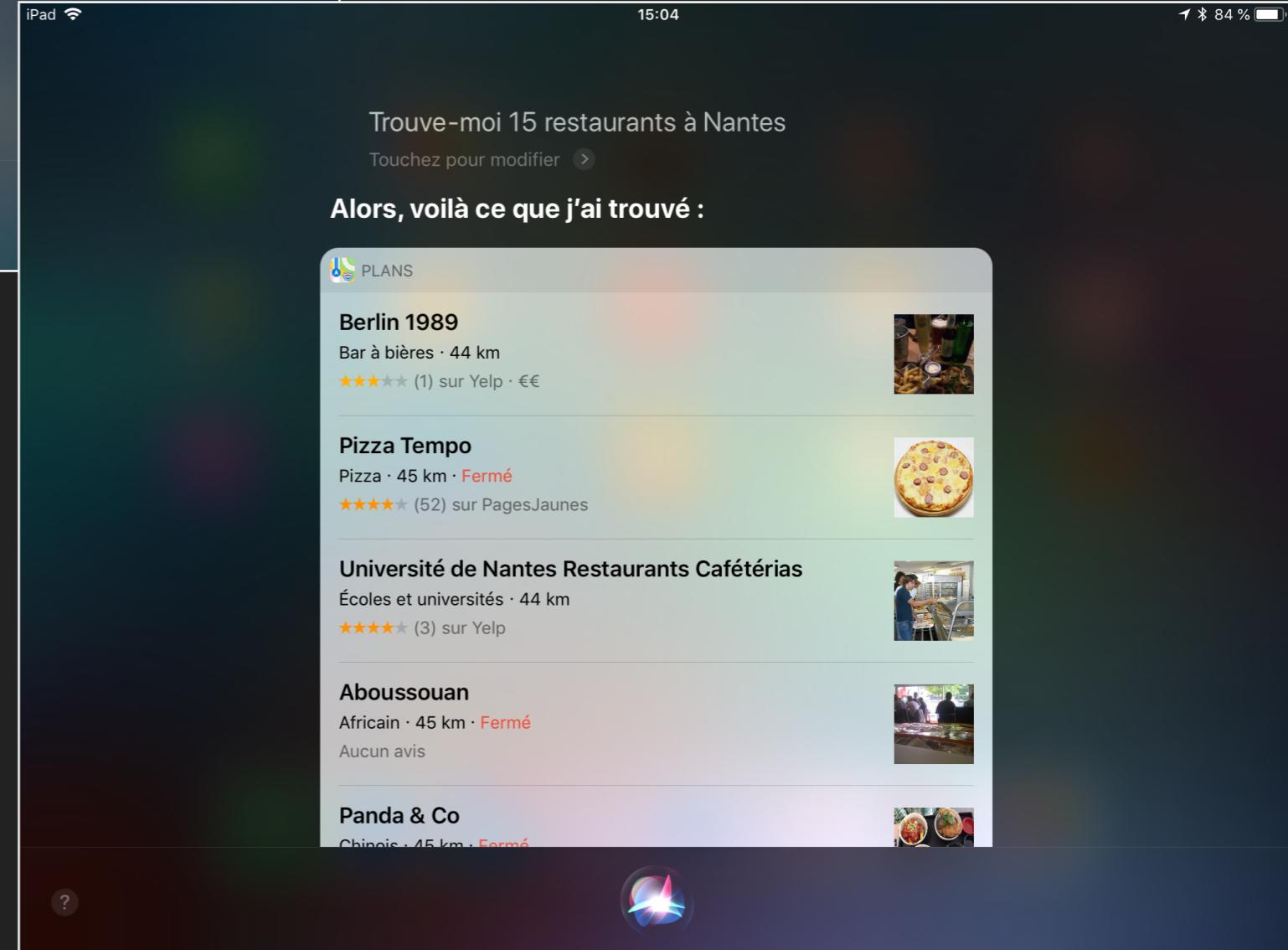
# DIALOG AGENTS

- ▶ Phone-based personal assistants:  
Siri, Cortana, Google Now, Alexa,  
...
- ▶ talking to your car
- ▶ communication with robots
- ▶ clinical uses for mental health
- ▶ chatting for fun



## TWO CLASSES OF SYSTEMS

1. Goal-based dialog systems, e.g. SIRI, interfaces to machines such as cars, robots, computers
2. Chatbots



iPad 15:07 83 %

Ce restaurant est-il ouvert  
Touchez pour modifier >

J'ai trouvé neuf restaurants correspondant à "-il".

Touchez celui que vous recherchez :

PLANS

**Il bambino**  
1 Rue Charles Lindbergh, Bouguenais · 47 km  
★★★★★ (128) sur TripAdvisor

**Il Ristorante**  
5 Avenue des Lions, Saint-Herblain · 49 km · Fermé  
★★★★★ (146) sur LaFourchette

**Il Toscano**  
Les Carreaux, La Mézière · Fermé  
★★★★★ (526) sur TripAdvisor

**Il Basilico**  
25 Avenue Jean Janvier, Rennes · Fermé

15:04 \* 84 %

me 15 restaurants à Nantes  
pour modifier >

ilà ce que j'ai trouvé :

**Pizzeria Rennes**  
Rennes · 44 km  
★★★★★ (1) sur Yelp · €€

**Pizza Tempo**  
Pizza · 45 km · Fermé  
★★★★★ (52) sur PagesJaunes

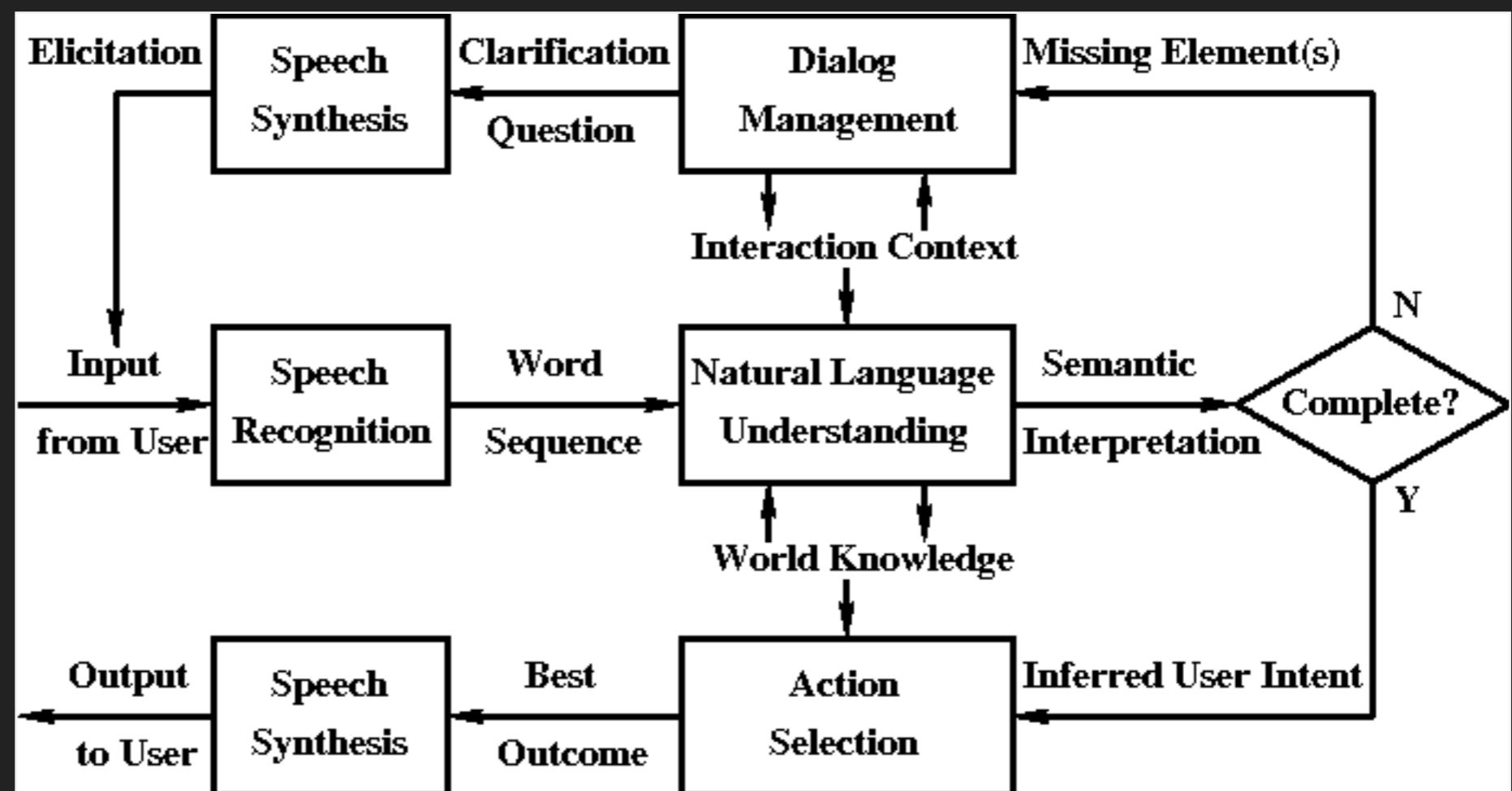
**Université de Nantes Restaurants Cafétérias**  
Écoles et universités · 44 km  
★★★★★ (3) sur Yelp

**Aboussouan**  
Africain · 45 km · Fermé  
Aucun avis

**Panda & Co**  
Chinois · 45 km · Fermé

# ARCHITECTURES FOR PRACTICAL DIALOG SYSTEMS

- ▶ Finite-state (passwords, credit cards, ...)
- ▶ Active ontology/frame based (SIRI, etc ...)

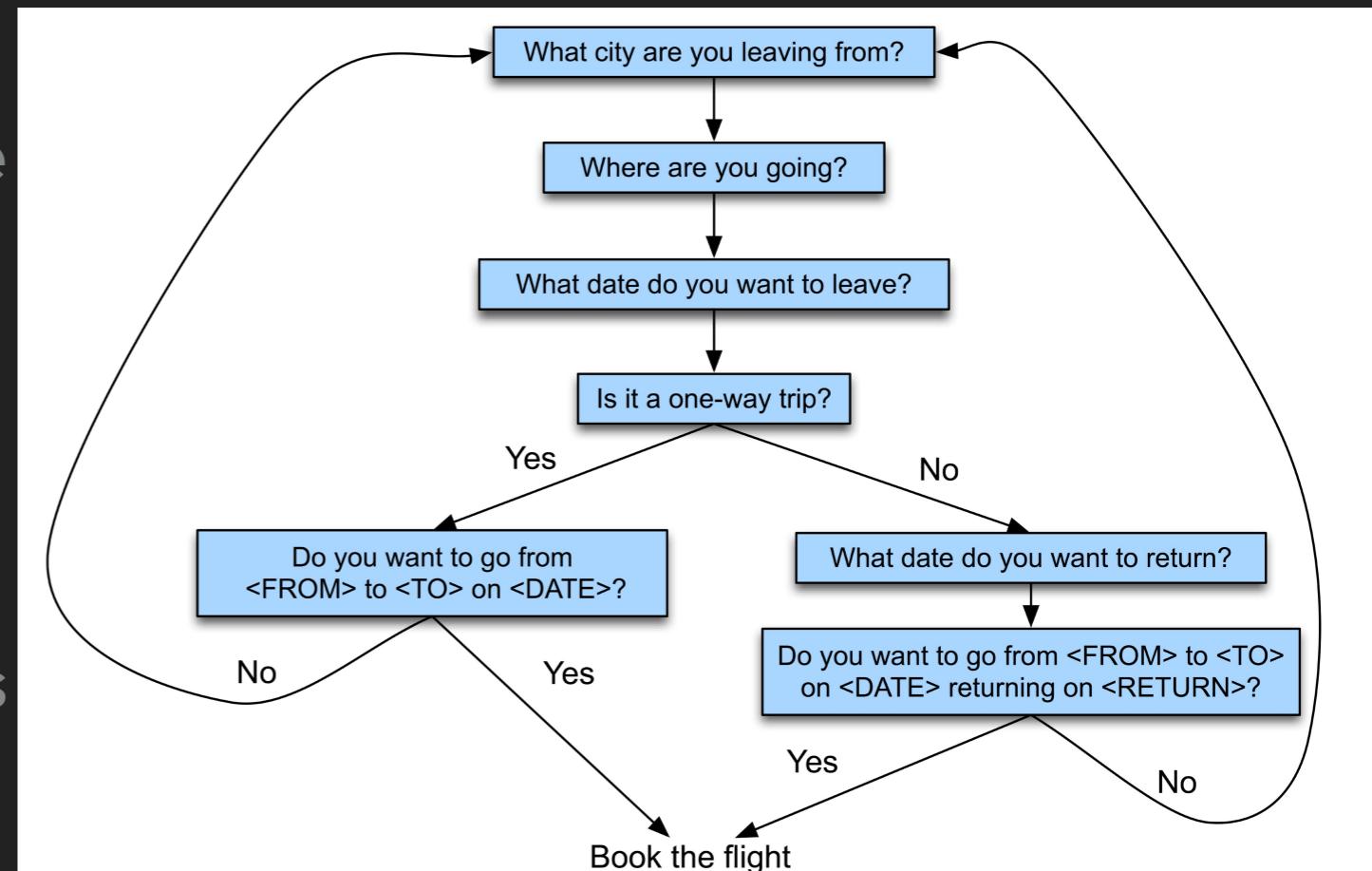


## FINITE-STATE DIALOG MANAGEMENT

- ▶ 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

# FINITE-STATE DIALOG MANAGERS

- ▶ System completely controls the conversation with the user
- ▶ It asks the user a series of questions
- ▶ Ignoring anything the user says that is not a direct answer to system's expectations
- ▶ Systems that control conversation like this are called **system initiative**, or **single initiative**.



**INITIATIVE: WHO HAS CONTROL OF CONVERSATION**

**IN NORMAL HUMAN-HUMAN DIALOG, INITIATIVE SHIFTS BACK AND FORTH BETWEEN PARTICIPANTS**

# SYSTEM INITIATIVE

- ▶ 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 (Automatic Speech Recognition)
  - ▶ known topic: better performance from NLU(Natural Language Understanding)
- ▶ OK for **very** simple tasks (entering a credit card, login name and password)
- ▶ Too limited, no flexibility

## PROBLEMS WITH SYSTEM INITIATIVE

- ▶ Real dialog involves give and take!
- ▶ In travel planning, users might want to say something that is not the direct answer to the question
- ▶ Take for example more than one question in a sentence

EYH, I'D LIKE TO FLY FROM PARIS TUESDAY MORNING;  
I WANT A TRAIN FROM NANTES TO PARIS ONE WAY  
LEAVING AFTER 6PM ON MONDAY

## SINGLE INITIATIVE + UNIVERSALS

- ▶ One can give users a little more flexibility by adding **universals**: commands you can say anywhere
- ▶ Same principle as augmenting every state of FSA with
  - ▶ Help
  - ▶ Start over
  - ▶ Correct
- ▶ This describes many implemented systems
- ▶ But still does not 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 dialog:

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?

# CHATBOT ARCHITECTURES

- ▶ Rule-based
- ▶ pattern-action rules (Eliza, 1966)
- ▶ + a mental model:  
affective variables (Parry, 1968)

(0 you 0 me)

[pattern]

(What makes you think I 3 you) [transform]

You hate me

What makes you think I hate you

Welcome to

EEEEEE	LL	IIII	ZZZZZZZ	AAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLL	IIII	ZZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?  
YOU: Men are all alike.  
ELIZA: What is the connection, do you suppose ?  
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: Is it important to you that your boyfriend made you come here ?  
YOU: He says I'm depressed much of the time.  
ELIZA: I am sorry to hear that you are depressed.  
YOU: It's true. I am unhappy.  
ELIZA: Can you explain what made you unhappy ?  
YOU: █

## IMPLICATIONS OF FIRST DIALOGS

- ▶ People became deeply emotionally involved with the program
- ▶ Weizenbaum tells the story of his secretary who would ask him to leave the room when she talked with ELIZA
- ▶ Suggestions that he might store all the conversations for later analysis already pointed out privacy implications
- ▶ Anthropomorphicism and the Heider-Simmel illusion (1944),  
<https://www.youtube.com/watch?v=76p64j3H1Ng&feature=related>

## INFORMATION-STATE AND DIALOG ACTS

- ▶ If we want a dialog system to be more than just form-filling, must be able to:
  - ▶ decide when the user has asked a question, made a proposal, rejected a suggestion
  - ▶ ground a user's utterance, ask clarification questions, suggestion plans
- ▶ This suggests that C.A. need sophisticated models of interpretation and generation
  - ▶ in terms of speech acts and grounding
  - ▶ needs more complex representation of dialog context (rather than just a list of slots)

# INFORMATION RETRIEVAL-BASED MODELS

- ▶ Cleverbot (1988)
- ▶ Xiaolce (2014)
- ▶ 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

The screenshot shows the Cleverbot homepage. At the top is the Cleverbot logo with a brain icon. Below it is a statistic: "56420 people talking". A conversation transcript follows:

Where are you from?  
United States of America.  
Which state?  
Are you asking where I was born or raised?  
Born.  
I wasn't born in a state, and I wasn't raised in any particular area. [Go share!](#)

Below the transcript is a blue search bar containing "say to cleverbot...". Underneath the search bar are three buttons: "think about it", "think for me", and "thoughts so far".

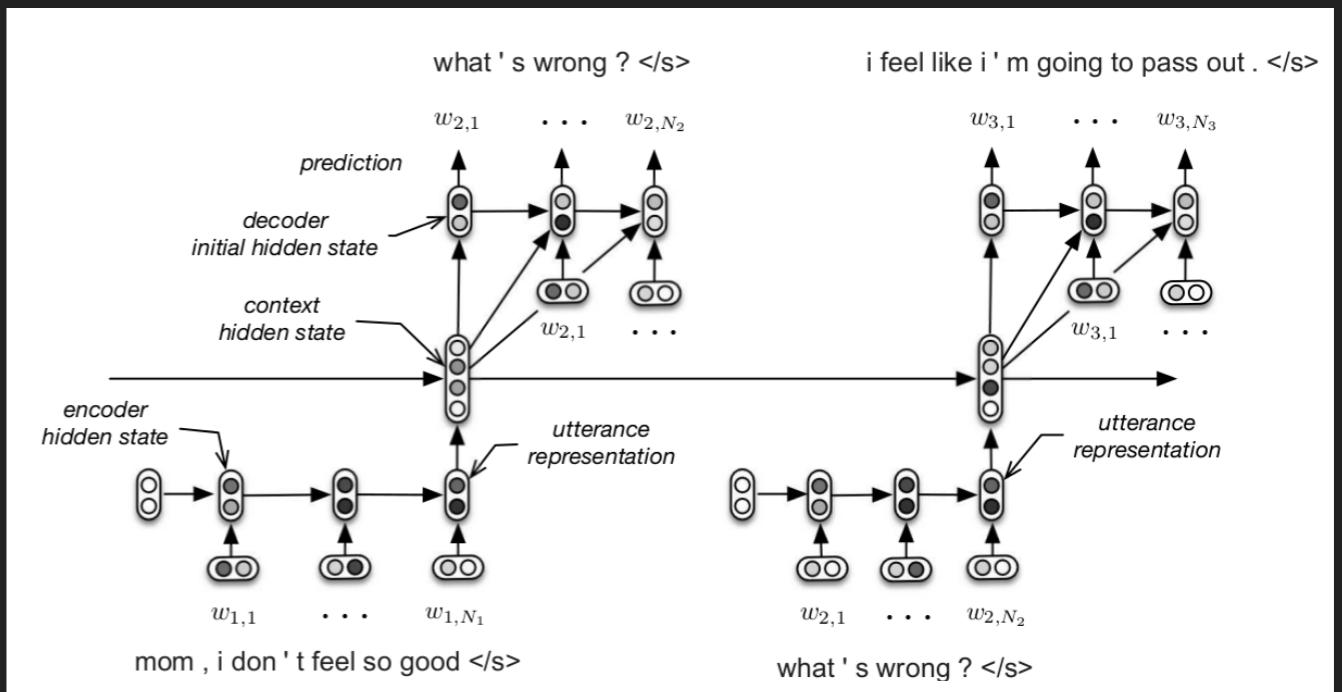
# DEEP LEARNING CHATBOTS

- ▶ Train on sequential models on:
  - ▶ movie dialog databases
  - ▶ twitter conversations
  - ▶ forum discussions

trained on 127M Twitter

context-message-response triples

Seq2seq, 2015



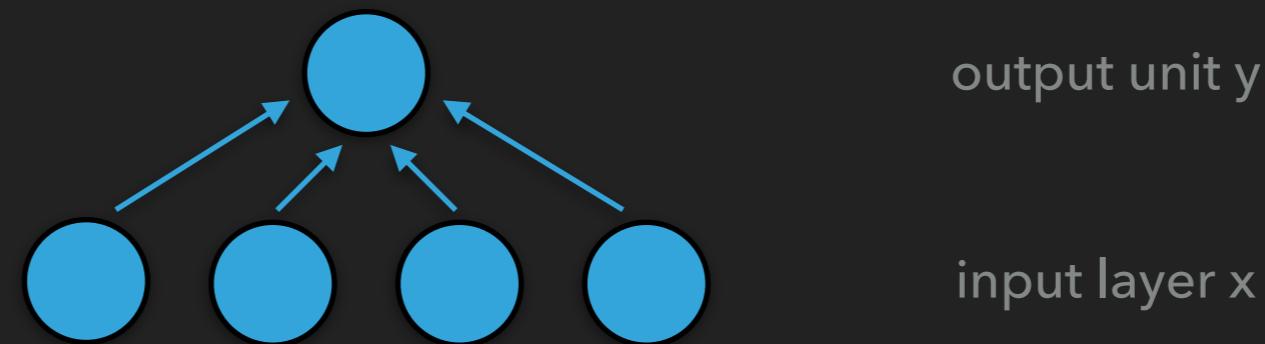
# Context $c$	Message $m$	Generated Response $r$
1 noone can help this time its just i need a support system	well im here to support you. whenever you need me.	thank you, i appreciate it thank you
2 nothin much, and how's the book?!	its good but i'm only like halfway through cuz i don't feel like reading. i'm so bored ...	that's good! i have the book but i'm bored too.
3 i know! what have you been up to?	nothing really. just enjoying summer and working. what about you?	that's good! nothing much just listening to music and you?
4 thanks, your way better though	i'm just good at expressing it that's but trust me you are freakin amazing and waay better	i know, i don't trust you, i love you too! thank you
5 thanks. yeah i am good thanks! good, but angry!	sorry you're angry, been there, i hope it dies down.	i hope so too, i'm not angry!!
6 figures. bass. never ate bass.	bass is extremely mild with very few bones. my favorite after catfish	that's one of the bones, chuck bass is mine too.
7 today i will walk for miles. clear sky clear mind.	love walking. it's too hot to walk this month. i haven't moved. frozen. heat freeze.	it's too hot for you to get up to honey if you have frozen yogurt to you.
8 arresting people over "nasty tweets"? please. i'm not condoning what the lad did but arresting him is a tad silly.	it appears he may have sent death threats to others while all this was kicking off too.	i did! he was armed with nuclear threats? that's what's happening to you.

## DEEP LEARNING?

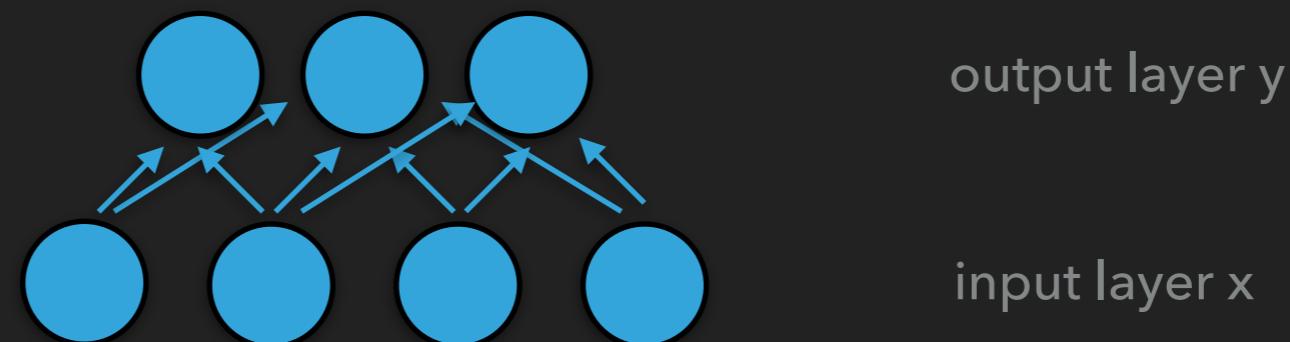
- ▶ Neural networks, typically several hidden layers (depth corresponds to the number of hidden layers)
  - ▶ single-layer neural nets are linear classifiers
  - ▶ multi-layer neural nets are more expressive
- ▶ Impressive performances in both **computer vision** and **speech recognition** over the last several years
- ▶ Neural nets date back from the 50's, why they suddenly made a comeback?
  - ▶ fast computers (GPUs), and very large datasets have made it possible to train very complex models

## NNETS: RECALLS

- ▶ Simplest variant: single layer feedforward net
  - ▶ For **binary classification**, single output unit



- ▶ For multi class classification, **K** output units

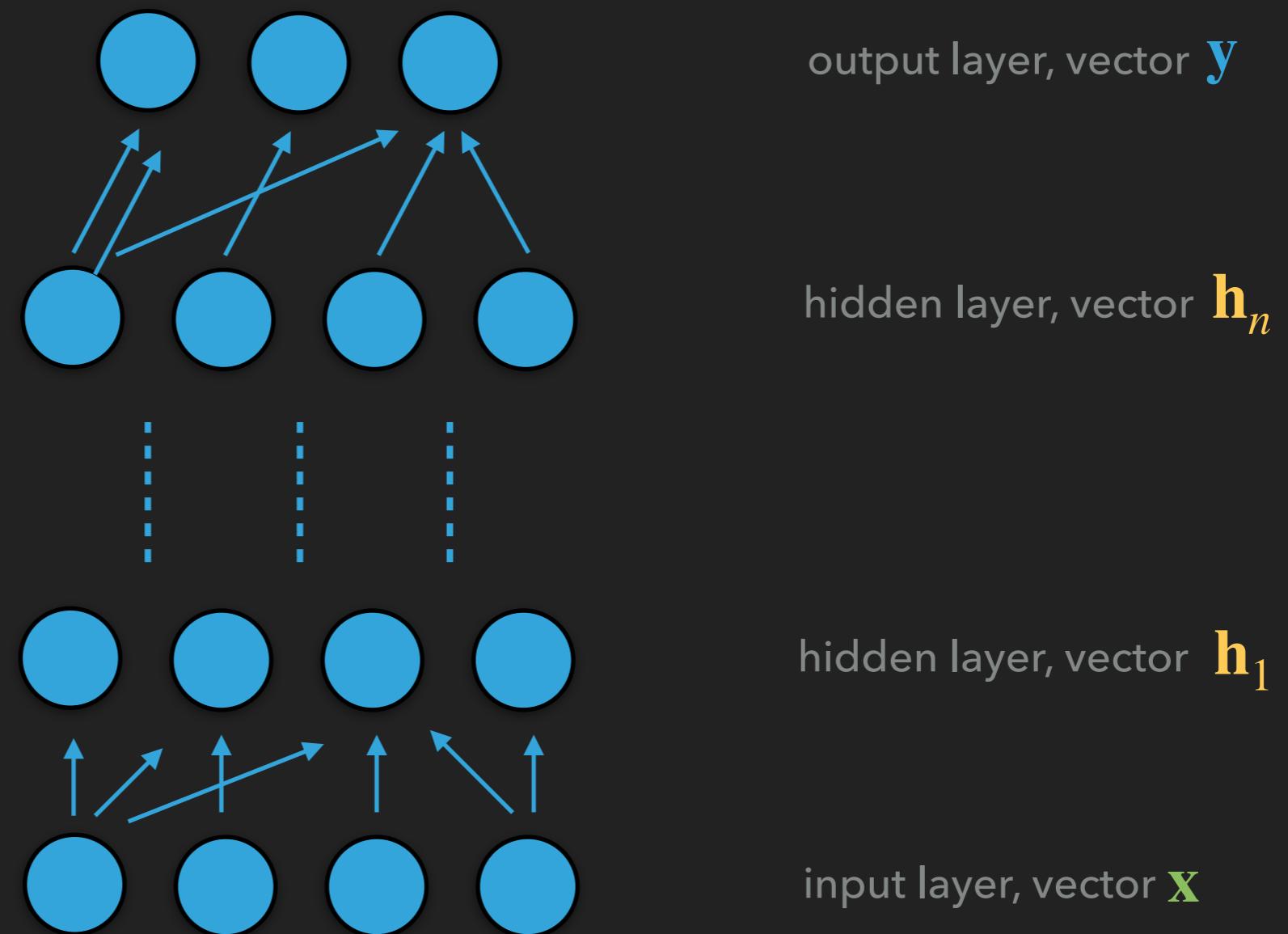


## MULTICLASS MODELS

- ▶ Predict one of the K classes, returns the class i with the highest score:  $\text{argmax}_i(y_i)$
- ▶ With nnets, done with softmax function, that maps a N-dimensional vector into a distribution over the N outputs
- ▶ For a vector  $\mathbf{z}$ ,  $P(i) = \text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{k=1}^N e^{z_k}}$

# MULTI-LAYER FEEDFORWARD NETWORKS

- ▶ Generalization to multi-layer networks



## CHALLENGES IN USING NNETS FOR NLP

- ▶ inputs are discrete: words, labels, structures
- ▶ Nnets work best with continuous vectors
  - ▶ we typically want to learn a mapping (embedding) from discrete words (input) to dense vectors
- ▶ The input to a nnet is a fixed-length vector. How to represent a variable-length sequence as a vector?
  - ▶ use recurrent neural nets: read in one word at a time to predict a vector, use that vector and the next word to predict a new vector

# NLP APPLICATIONS OF NEURAL NETS

- ▶ Word embeddings (word2vec, Glove, ...)
  - ▶ train a nn to predict a word from its context, gives a dense vector representation of each word
- ▶ Neural language models
  - ▶ use recurrent neural nets (RNN) to predict word sequences (LSTMs)
- ▶ Sequence-to-sequence (seq2seq) models:
  - ▶ from machine translation: use one RNN to encode source string, another RNN to decode it into a target string. Also used for automatic image captioning.
- ▶ *Recursive neural networks:*
  - ▶ used for parsing text

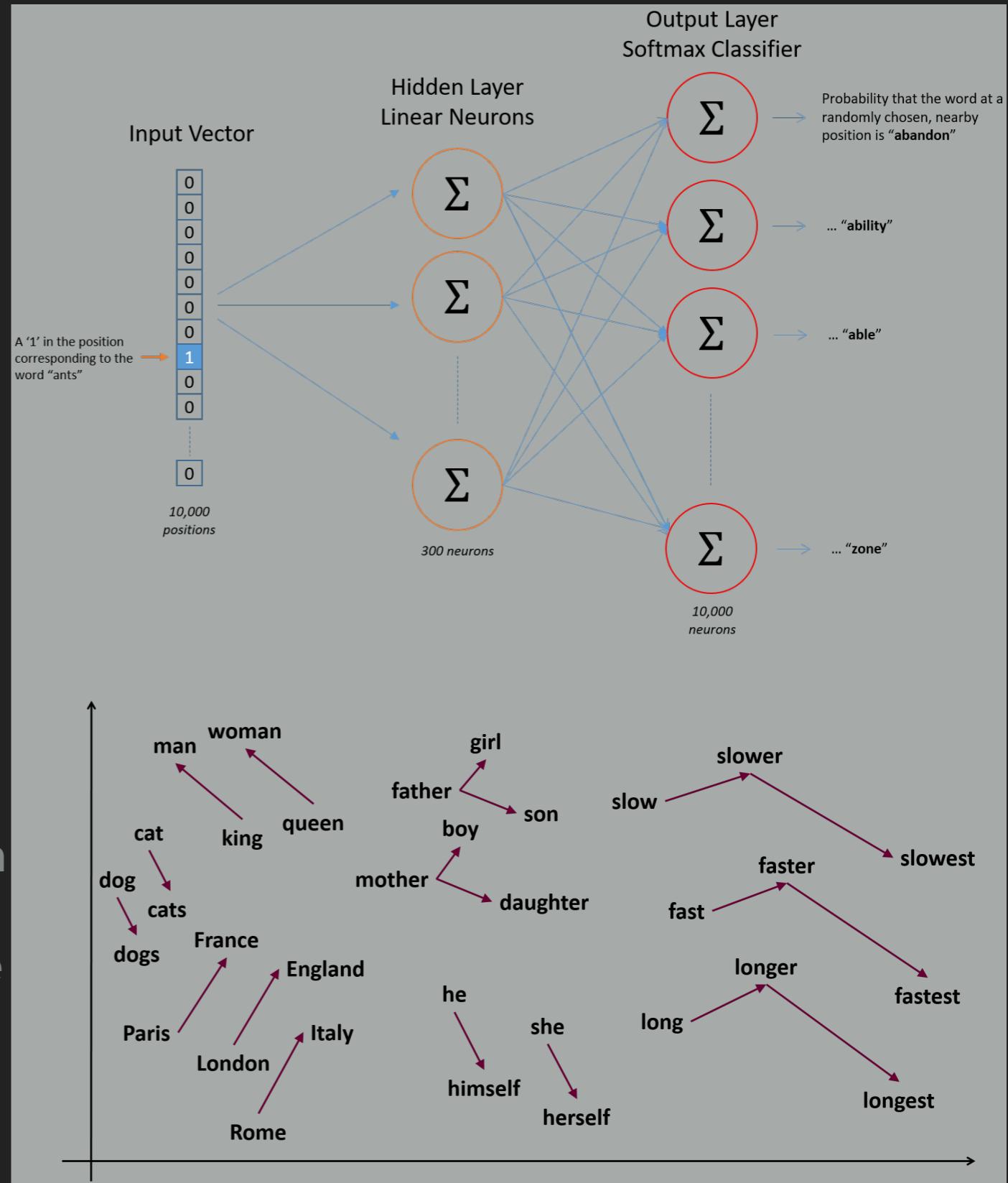
# WORD EMBEDDINGS

## ► Principle

If you use a feedforward net to predict probability of words that appear in a context, the hidden layer of the network provides a dense vector representation of the word

Words appearing in similar contexts will have similar vector representation

These models can be trained on large amounts of raw text (pre-trained embeddings can be downloaded)



# NEURAL LANGUAGE MODELS

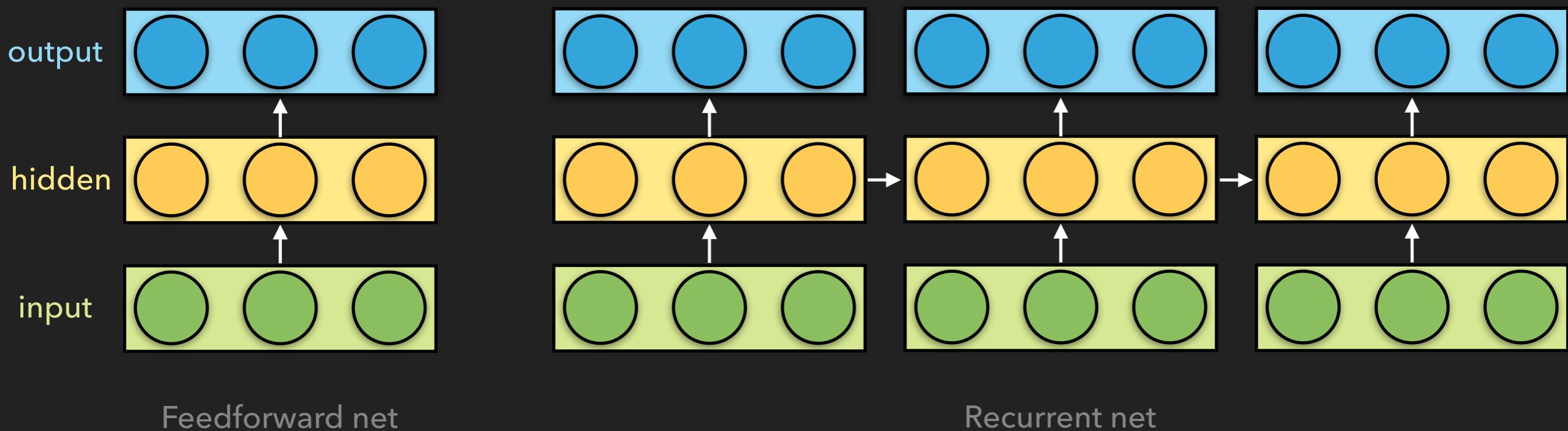
- ▶ LMs define a distribution over strings:  $P(w_1, \dots, w_k)$
- ▶ It factors  $P(w_1, \dots, w_k)$  into the probability of each word:

$$P(w_1, \dots, w_k) = P(w_1)P(w_2 | w_1)P(w_3 | w_2, w_1)\dots, P(w_k | w_{k-1}, \dots, w_1)$$

- ▶ A neural LM needs to define a distribution over the  $V$  words in the vocabulary, conditioned on the preceding words.
- ▶ **output layer**:  $V$  units (one per word in the vocabulary) with softmax to get a distribution
- ▶ **input**: represent each preceding word by its  $d$ -dimensional embedding
  - ▶ fixed-length history (n-gram): use preceding  $n-1$  words
  - ▶ variable-length history: use a recurrent neural net

## RECURRENT NEURAL NETWORKS

- ▶ Basic RNN: modify the feedforward architecture (which predicts a string  $w_0, \dots, w_n$  one word at a time) such that the output of the current step ( $w_i$ ) is given as additional input to the next time step (when predicting the output for  $w_{i+1}$ )

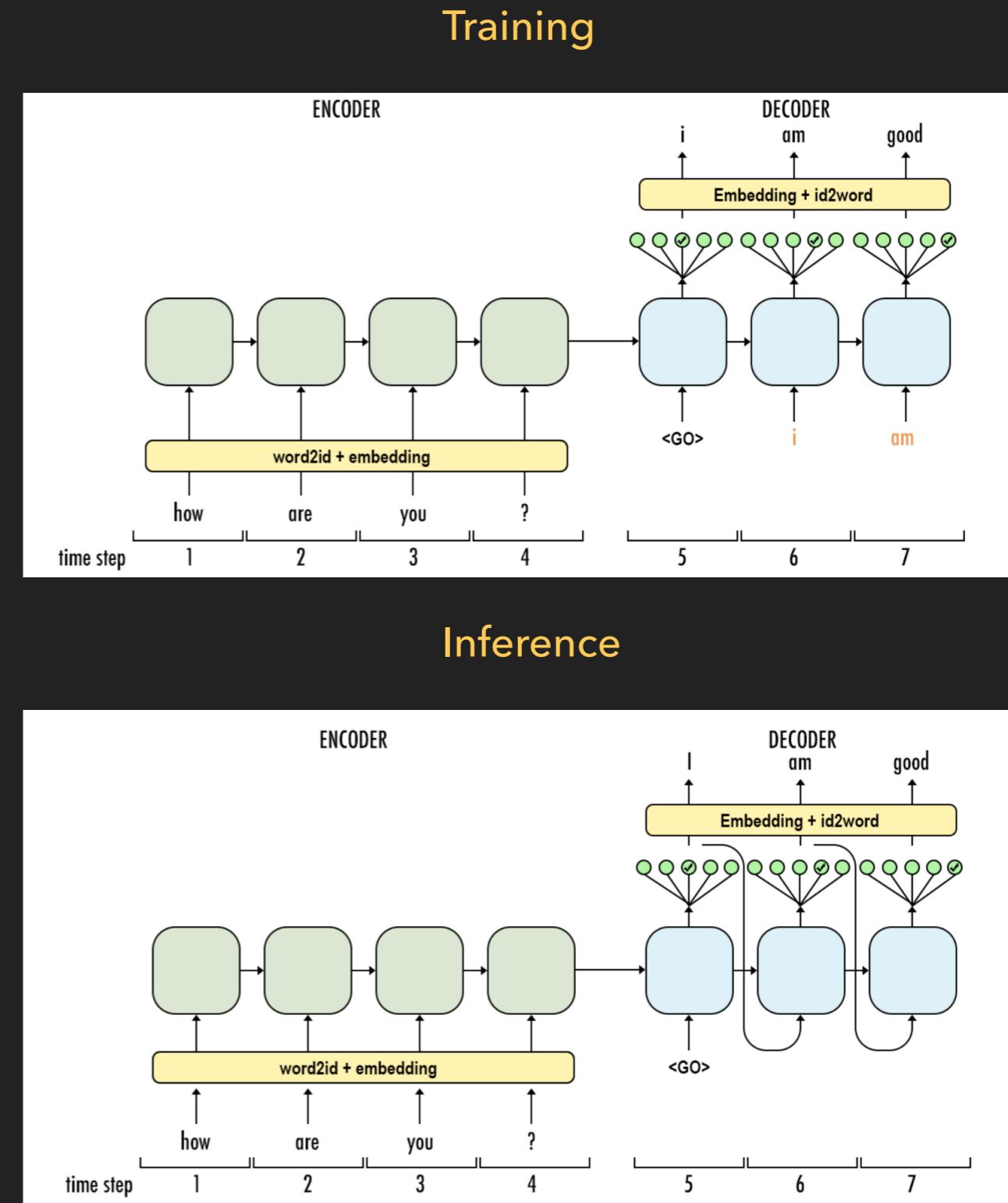


# SEQ2SEQ MODELS

- ▶ Task (e.g. machine translation):
  - ▶ given one variable length sequence as input, return another variable length sequence as output

## Principle

- ▶ Use one RNN to encode the input sequence (encoder)
- ▶ Feed the last hidden state as input to a second RNN (decoder) that then generates the output sequence



## FURTHER READINGS AND LINKS

- ▶ deep learning for NLP: <http://cs224d.stanford.edu>
- ▶ Yoav Goldberg's Primer on Nnets for NLP:  
<http://arxiv.org/pdf/1510.00726.pdf>
- ▶ seq2seq chatbot code: <https://github.com/topics/seq2seq-chatbot>
- ▶ Demo: <https://projector.tensorflow.org>
- ▶ Exercise: <https://radimrehurek.com/gensim/scripts/word2vec2tensor.html>
- ▶