

# CS 481

## *Artificial Intelligence Language Understanding*

April 6, 2023

# Announcements / Reminders

- Please follow the Week 12 To Do List instructions
- Programming Assignment #02 is due on ~~Sunday 04/02/23~~ Friday 04/07/23 11:59 PM CST
- Final Exam date:  
**Thursday 04/27/2023 (last week of classes!)**
  - Ignore the date provided by the Registrar
  - Section 02 [Online]: contact Mr. Charles Scott ([scott@iit.edu](mailto:scott@iit.edu)) to arrange your exam

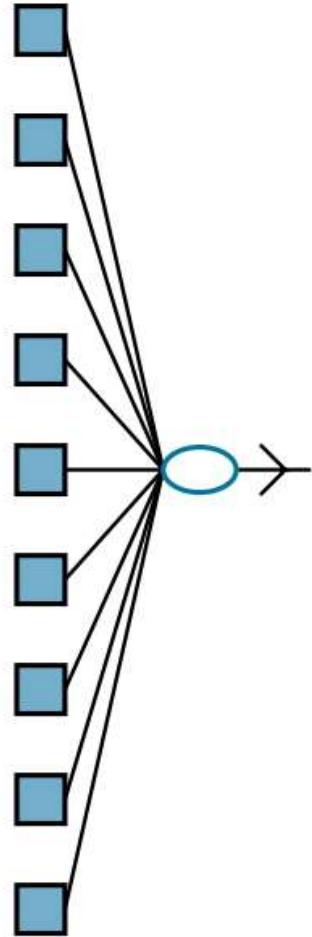
# Plan for Today

- NLP and Neural Networks
- Information Extraction
- Chatbots and Dialogue Agents (pre-ChatGPT)

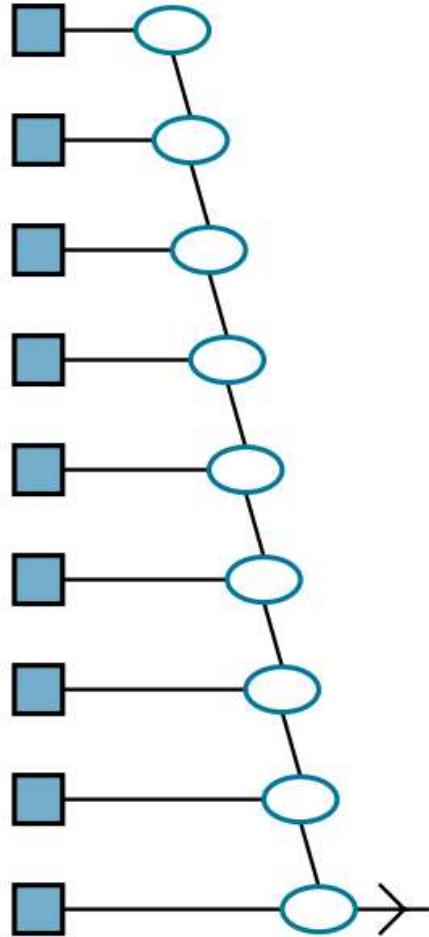
# Deep Learning

Deep learning is a broad family of techniques for machine learning (also a sub-field of ML) in which hypotheses take the form of **complex algebraic circuits with tunable connections**. The word “deep” refers to the fact that the circuits are **typically organized into many layers**, which means that **computation paths from inputs to outputs have many steps**.

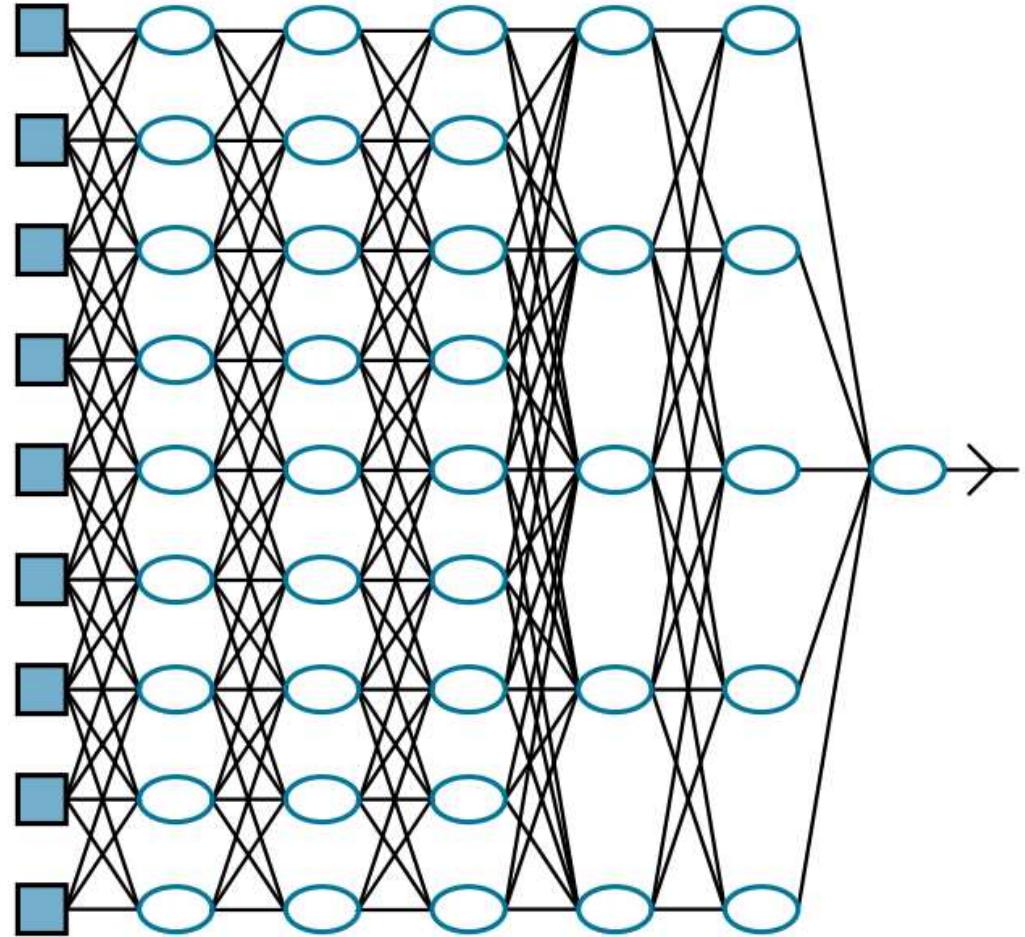
# Shallow vs. Deep Models



Shallow  
Model



Shallow  
Model

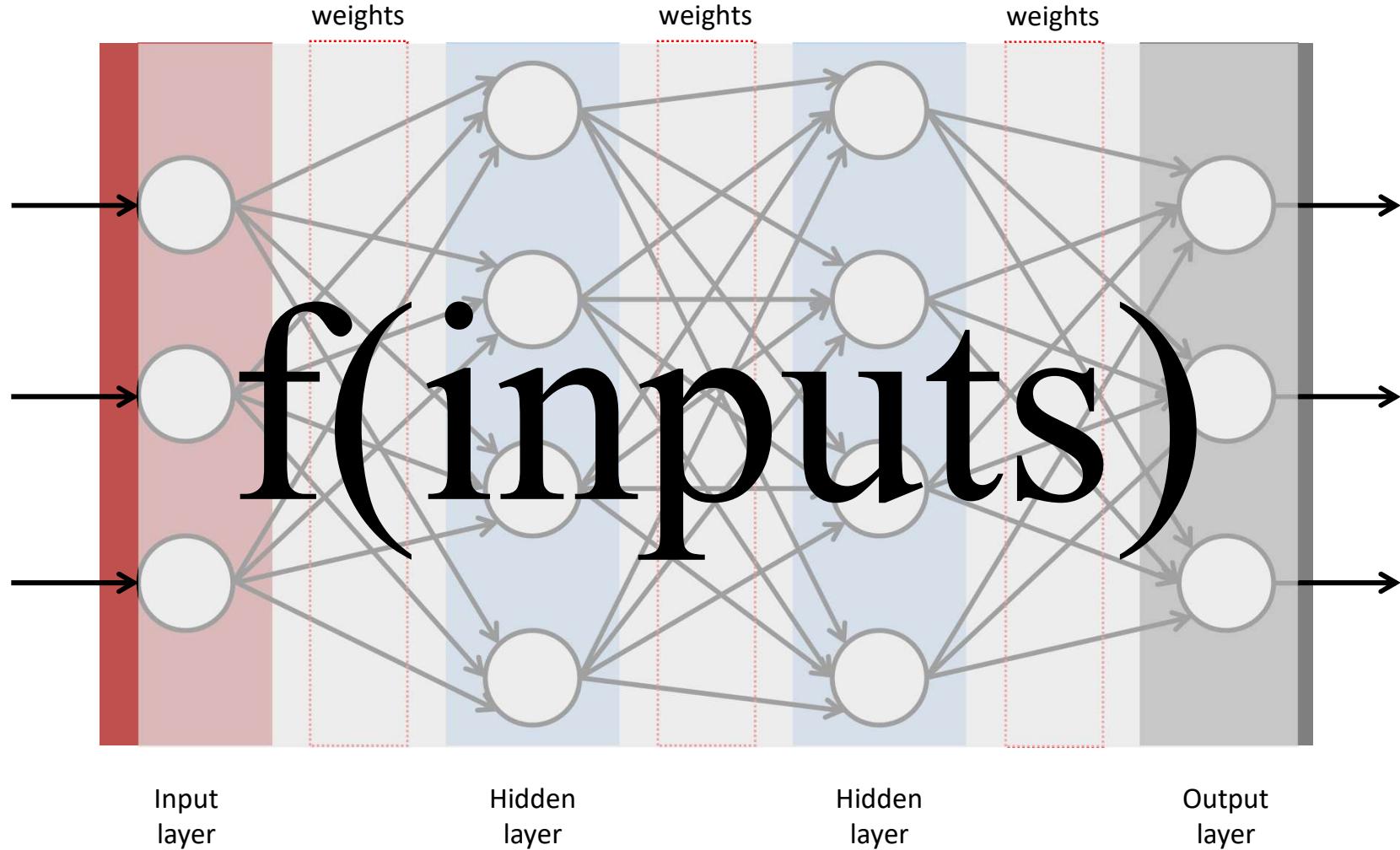


Deep  
Model

Longer computation path

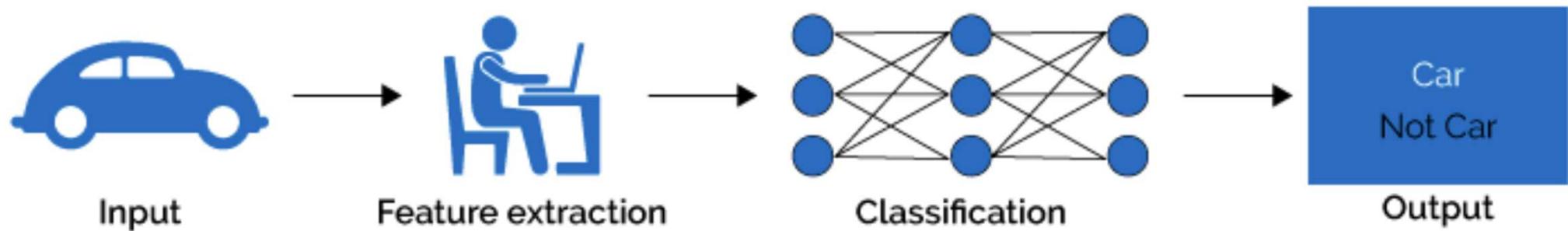
# ANN as a Complex Function

In ANNs **hypotheses take form of complex algebraic circuits** with tunable connection strengths (weights).

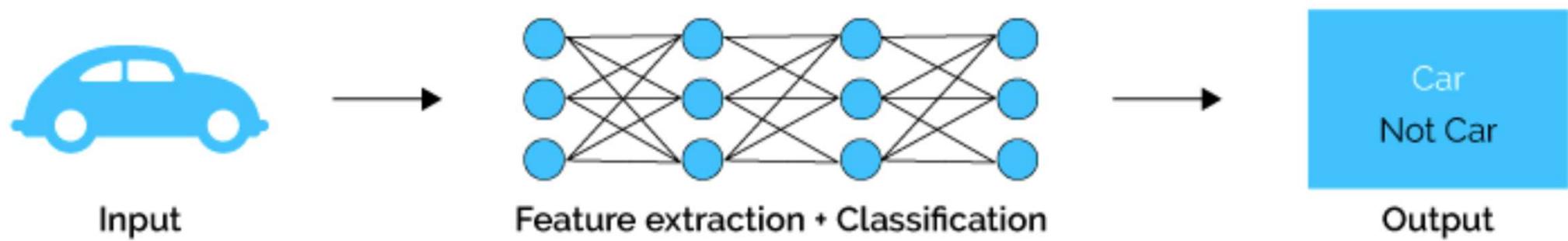


# Machine Learning vs. Deep Learning

## Machine Learning



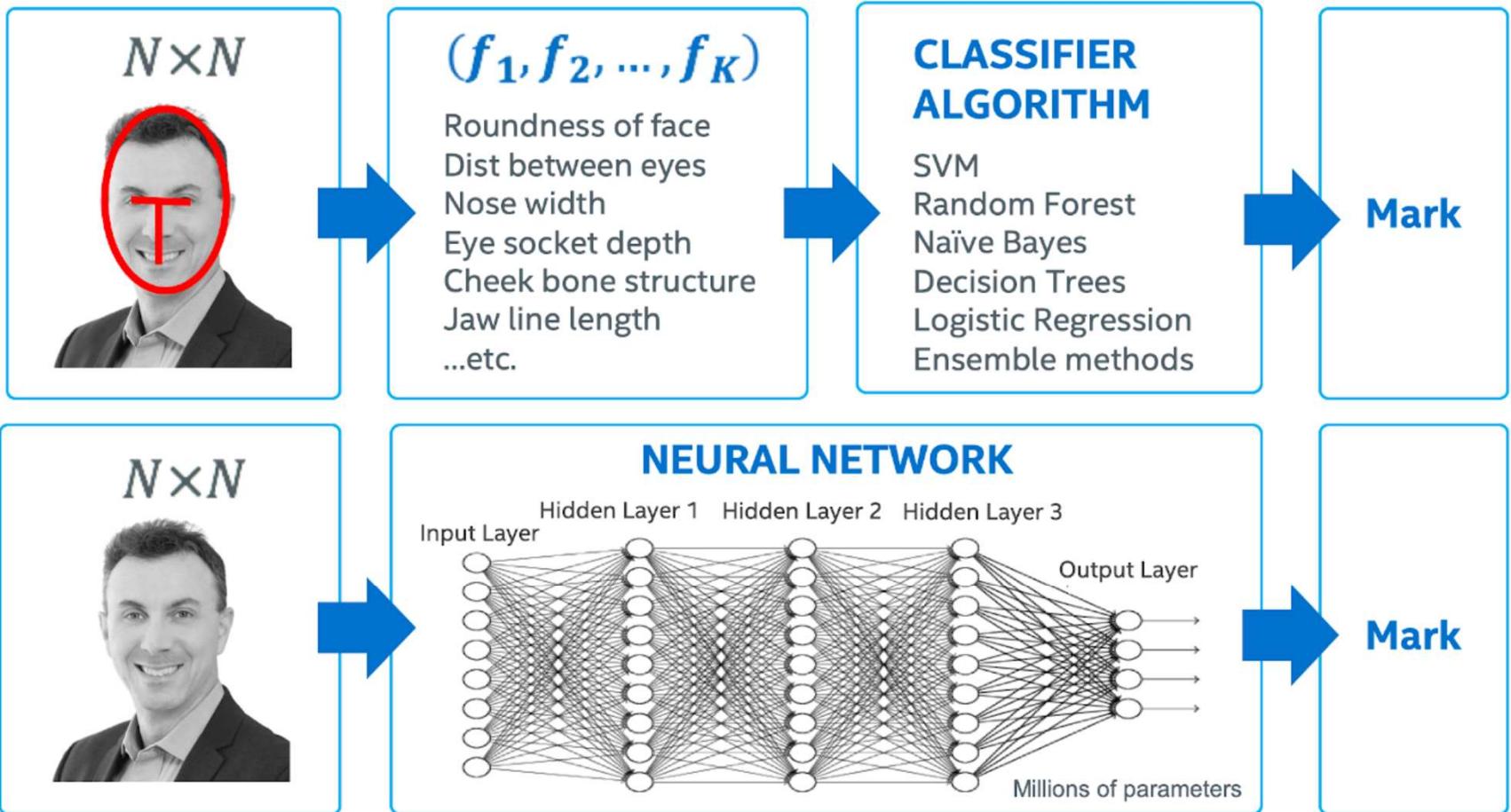
## Deep Learning



Source: <https://www.quora.com/What-is-the-difference-between-deep-learning-and-usual-machine-learning>

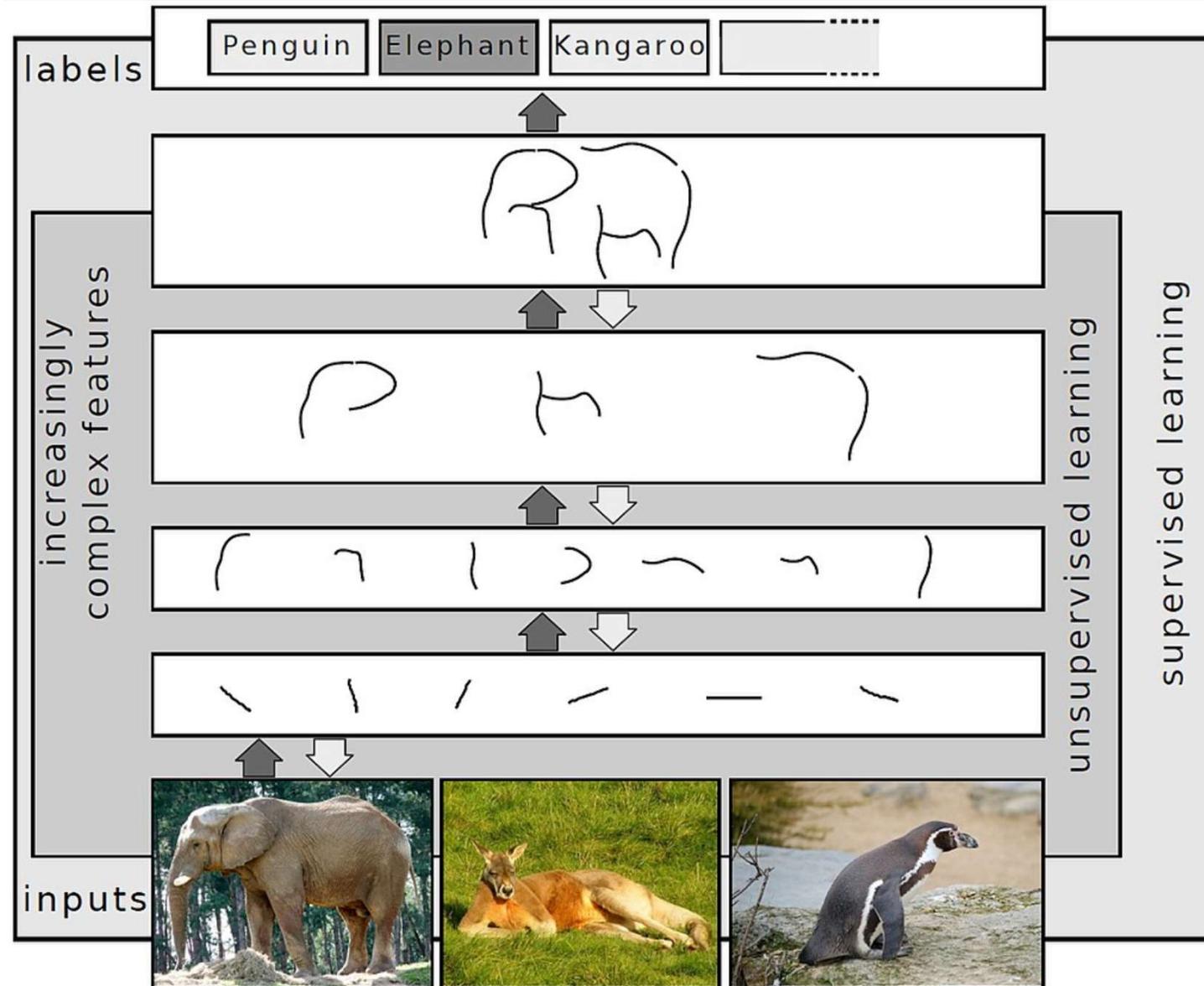
# Machine Learning vs. Deep Learning

## Classic Machine Learning



Source: <https://www.intel.com/content/www/us/en/artificial-intelligence/posts/difference-between-ai-machine-learning-deep-learning.html>

# Deep Learning: Feature Extraction



Source: [https://en.wikipedia.org/wiki/Deep\\_learning](https://en.wikipedia.org/wiki/Deep_learning)

# Texts in Different Sizes: Ideas

Some simple solutions:

1. Make the input the **length of the longest sample**
  - if shorter then pad with zero embeddings
  - truncate if you get longer reviews at test time
2. Create **a single "sentence embedding"** (the same dimensionality as a word) to represent all the words
  - take the mean of all the word embeddings
  - take the element-wise max of all the word embeddings
  - for each dimension, pick the max value from all words

# Language Models Revisited

**Language Modeling:** Calculating the probability of the next word in a sequence given some history.

- N-gram based language models
- other: neural network-based?

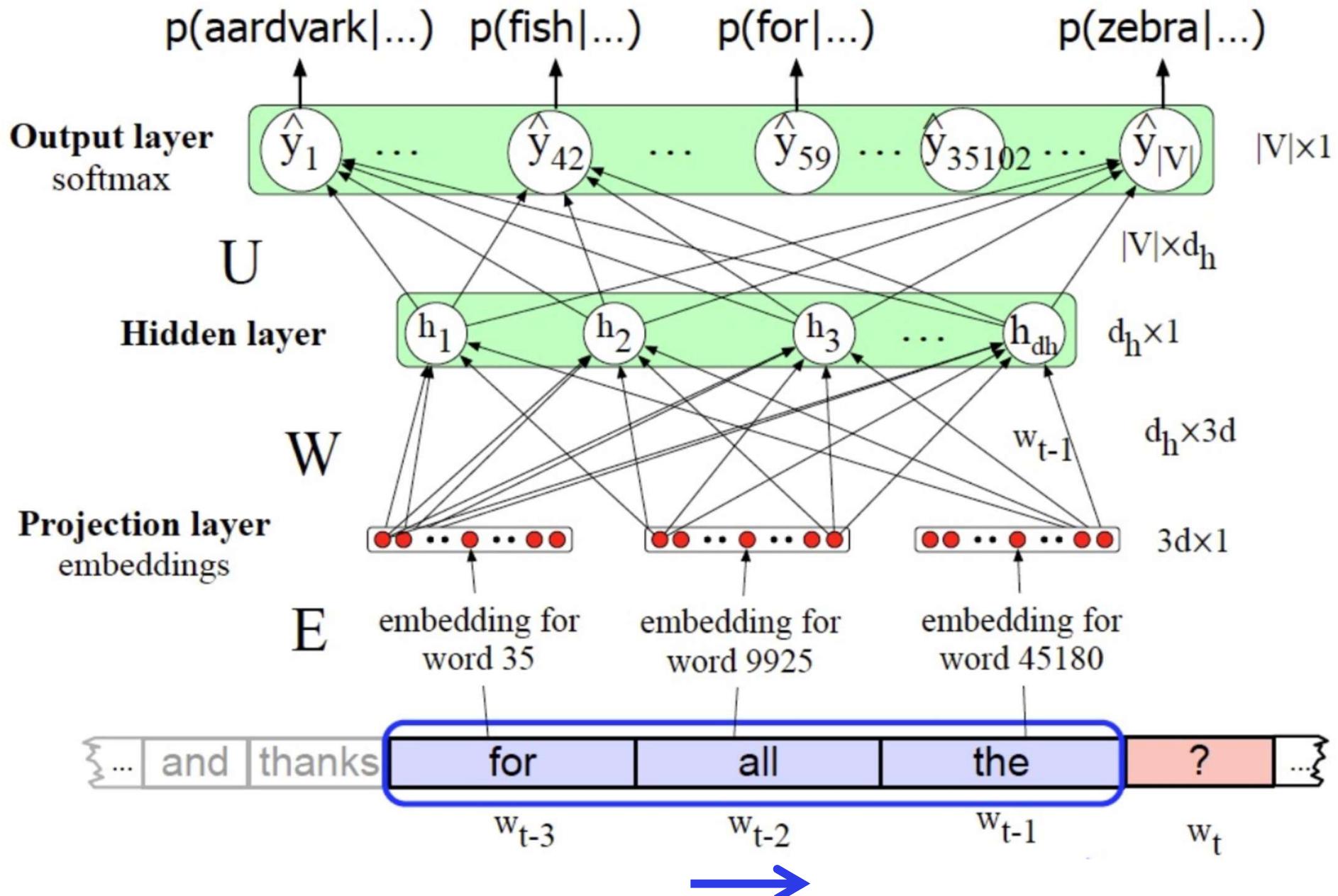
**Task:** predict next word  $w_t$   
given prior words  $w_{t-1}, w_{t-2}, w_{t-3}, \dots$

**Problem:** Now we're dealing with sequences of arbitrary length.

**Solution:** Sliding windows (of fixed length)

$$P(w_t | w_1^{t-1}) \approx P(w_t | w_{t-N+1}^{t-1})$$

# Neural Language Model



# Neural LM Better Than N-Gram LM

**Training data:**

We've seen: I have to make sure that the cat gets fed.

Never seen: dog gets fed

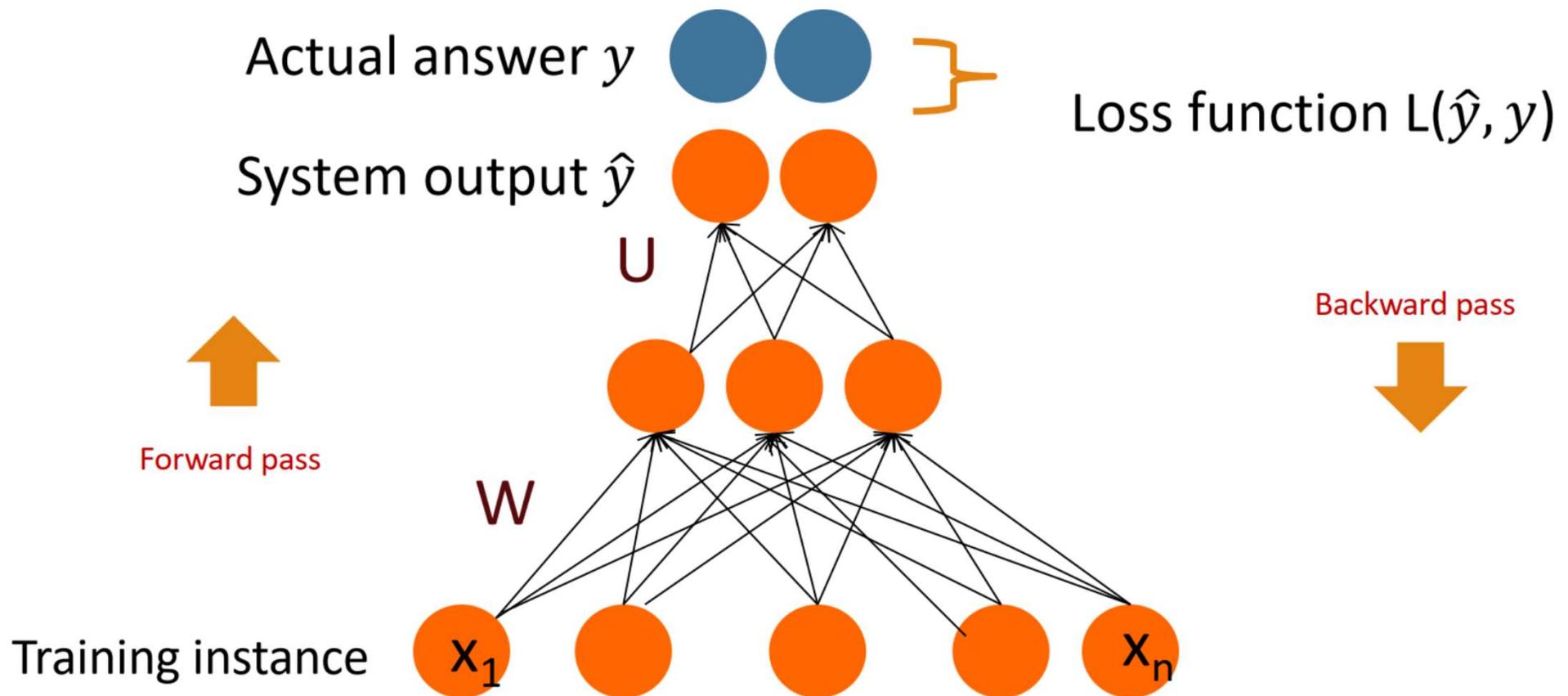
**Test data:**

I forgot to make sure that the dog gets \_\_

N-gram LM can't predict "fed"!

Neural LM can use similarity of "cat" and "dog" embeddings to generalize and predict "fed" after dog

# Training Neural Networks



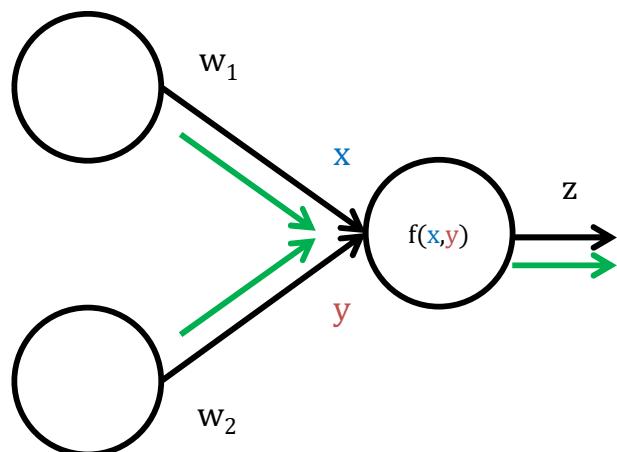
# Training Neural Networks: Intuition

For every training tuple  $(\mathbf{x}, \mathbf{y}) = (\text{feature vector}, \text{label})$

- Run forward computation to find estimate  $\hat{\mathbf{y}}$
- Run backward computation to update weights:
  - For every output node
    - Compute loss  $L$  between true  $\mathbf{y}$  and the estimated  $\hat{\mathbf{y}}$
    - For every weight  $w$  from hidden layer to the output layer
    - Update the weight
  - For every hidden node
    - Assess how much blame it deserves for the current answer
    - For every weight  $w$  from input layer to the hidden layer
      - Update the weight

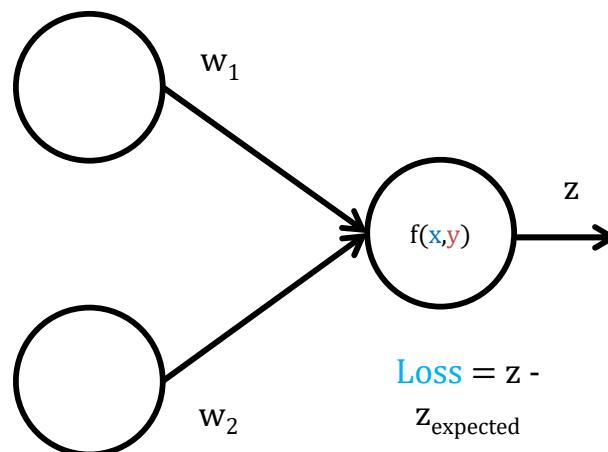
# Back-propagation

Feed forward



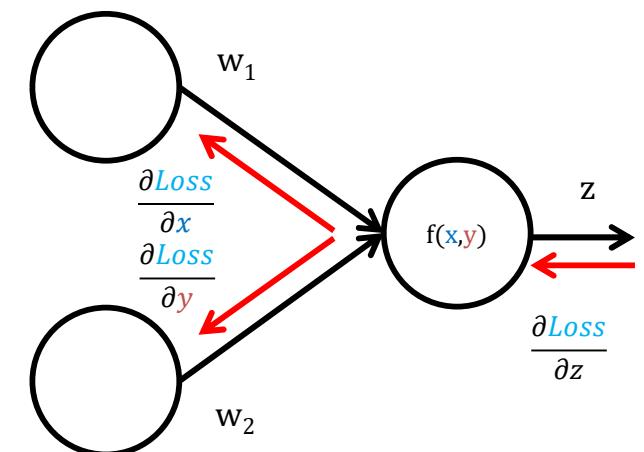
Feed a **labeled sample** through the network

Evaluate Loss



How “incorrect” is the result compare to the label?

Back-propagation



Update weights  
(use **Gradient Descent**)

# Cross-Entropy Loss: Calculation

**Goal:** **maximize** probability of the correct label  $P(y|x)$

$$\begin{aligned}\log p(y|x) &= \log [\hat{y}^y (1 - \hat{y})^{1-y}] \\ &= y \log \hat{y} + (1 - y) \log (1 - \hat{y})\end{aligned}$$

**Goal:** **minimize** error / cross-entropy loss

$$L_{\text{CE}}(\hat{y}, y) = -\log p(y|x) = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})]$$

which yields:

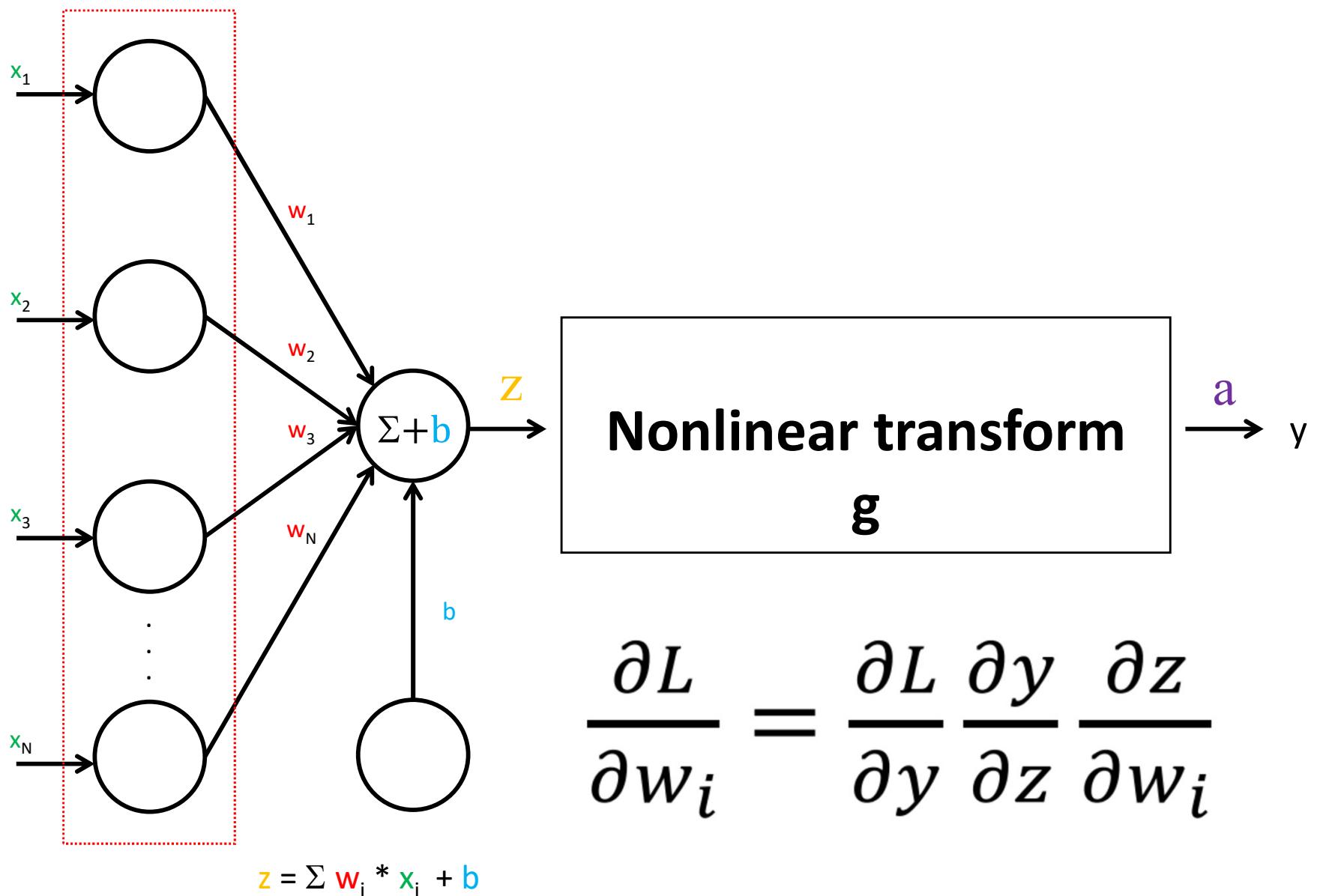
$$L_{\text{CE}}(\hat{y}, y) = -[y \log \sigma(w \cdot x + b) + (1 - y) \log (1 - \sigma(w \cdot x + b))]$$

# Gradients and Learning Rate

- The value of the gradient (slope in our example)  $\frac{d}{dw} L(f(x; w), y)$  weighted by a **learning rate**  $\eta$
- Higher learning rate means move **w** faster

$$w^{t+1} = w^t - \eta \frac{d}{dw} L(f(x; w), y)$$

# NN Node: Derivative of the Loss



# Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents and other electronically represented sources.

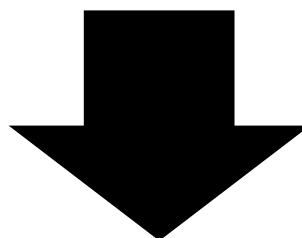
In most of the cases this activity concerns processing human language texts by means of natural language processing (NLP).

Related: processing video, etc.

# Information Extraction

## Unstructured data (document):

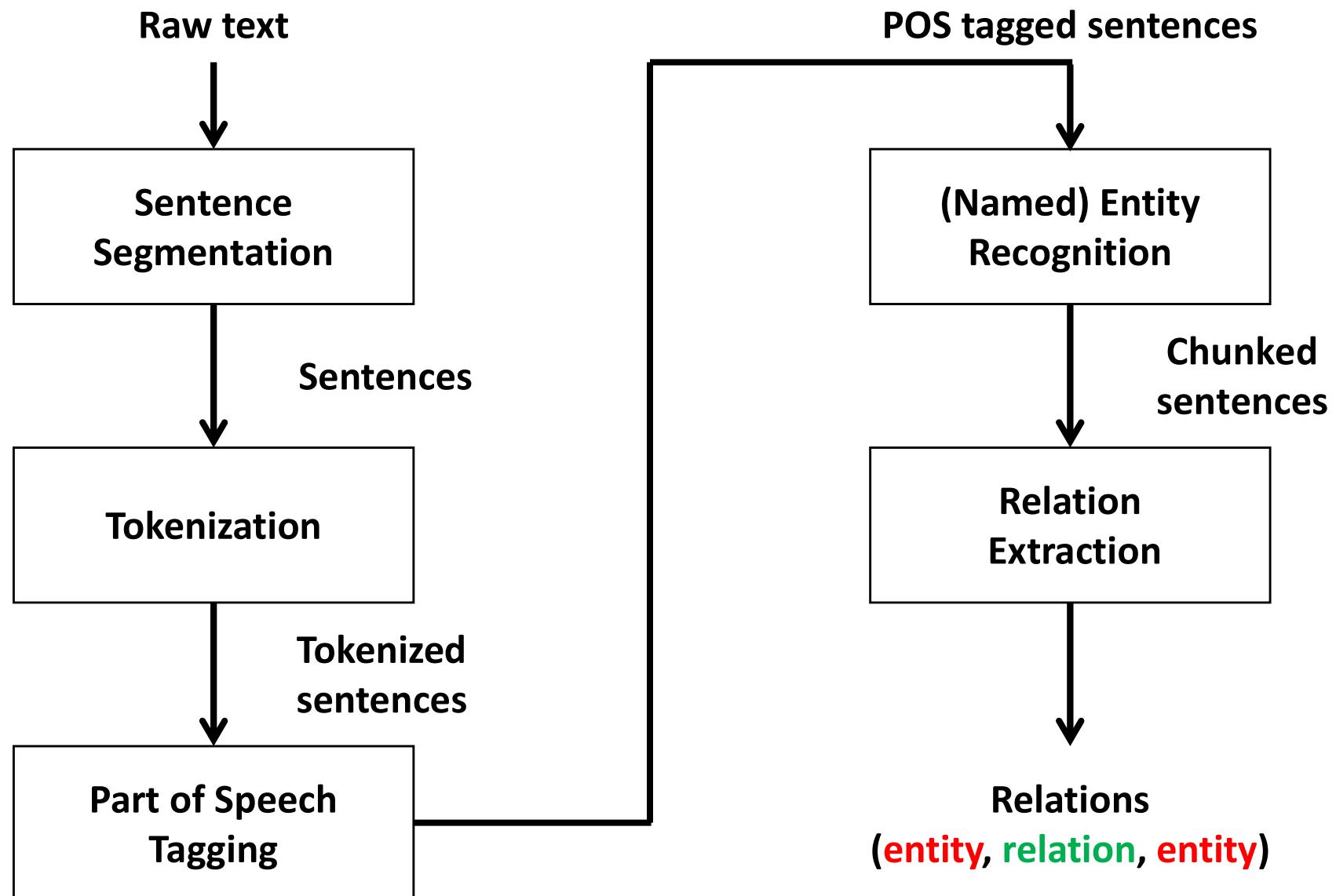
*The fourth Wells account moving to another agency is the packaged paper-products division of Georgia-Pacific Corp., which arrived at Wells only last fall. Like Hertz and the History Channel, it is also leaving for an Omnicom-owned agency, the BBDO South unit of BBDO Worldwide. BBDO South in Atlanta, which handles corporate advertising for Georgia-Pacific, will assume additional duties for brands like Angel Soft toilet tissue and Sparkle paper towels, said Ken Haldin, a spokesman for Georgia-Pacific in Atlanta.*



## Structured data:

Organization	Location
BBDO South	Atlanta
Georgia-Pacific	Atlanta

# Information Extraction Architecture



# Sample POS-Tagged Sentence

There/PRO were/VERB 70/NUM children/NOUN  
there/ADV ./PUNC

Preliminary/ADJ findings/NOUN were/AUX  
reported/VERB in/ADP today/NOUN 's/PART  
New/PROPN England/PROPN Journal/PROPN  
of/ADP Medicine/PROPN

# **Entity Recognition: Motivation**

**Entity Recognition is a supporting task for a variety of other problems:**

- **Sentiment analysis: consumer's sentiment toward a particular company or person?**
- **Question Answering: answer questions about an entity?**
- **Information Extraction: Extracting facts about entities from text.**

# Named Entity Recognition

Named-entity recognition (NER):

- also known as: entity identification, entity chunking, and entity extraction
- a subtask of NLP that **seeks to locate and classify named entities** mentioned in unstructured text into pre-defined categories such as:
  - person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

# Named Entities

Named entity, in its core usage, means **anything that can be referred to with a proper name.**

Four most common entity tags:

- PER (Person): “*Marie Curie*”
- LOC (Location): “*New York City*”
- ORG (Organization): “*Stanford University*”
- GPE (Geo-Political Entity): “*Boulder, Colorado*”

Often multi-word phrases. The term is also extended to things that are not entities:dates, times, prices

# Sample NER-Tagged Text

Citing high fuel prices, [United Airlines]<sub>ORG</sub> said [Friday]<sub>TIME</sub> it has increased fares by [\$6]<sub>MONEY</sub> per round trip on flights to some cities also served by lower-cost carriers. [American Airlines]<sub>ORG</sub>, a unit of [AMR Corp.]<sub>ORG</sub>, immediately matched the move, spokesman [Tim Wagner]<sub>PER</sub> said. [United]<sub>ORG</sub>, a unit of [UAL Corp.]<sub>ORG</sub>, said the increase took effect [Thursday]<sub>TIME</sub> and applies to most routes where it competes against discount carriers, such as [Chicago]<sub>LOC</sub> to [Dallas]<sub>LOC</sub> and [Denver]<sub>LOC</sub> to [San Francisco]<sub>LOC</sub>.

# Named Entity Tagging

The tasks of named entity recognition (NER):

- find **spans of text** that constitute proper names
- **tag the type** of the entity.

# Named Entity Recognition

## Unstructured data (document):

*The fourth Wells account moving to another agency is the packaged paper-products division of Georgia-Pacific Corp., which arrived at Wells only last fall. Like Hertz and the History Channel, it is also leaving for an Omnicom-owned agency, the BBDO South unit of BBDO Worldwide. BBDO South in Atlanta, which handles corporate advertising for Georgia-Pacific, will assume additional duties for brands like Angel Soft toilet tissue and Sparkle paper towels, said Ken Haldin, a spokesman for Georgia-Pacific in Atlanta.*

## Unstructured data (document) AFTER applying the Named Entity Recognition Process:

*The fourth [Wells]<sub>ORG</sub> account moving to another agency is the packaged paper-products division of [Georgia-Pacific Corp.]<sub>ORG</sub>, which arrived at [Wells]<sub>ORG</sub> only last fall. Like [Hertz]<sub>ORG</sub> and the History Channel, it is also leaving for an Omnicom-owned agency, the [BBDO South]<sub>ORG</sub> unit of [BBDO Worldwide]<sub>ORG</sub>. [BBDO South]<sub>ORG</sub> in [Atlanta]<sub>LOC</sub> which handles corporate advertising for Georgia-Pacific, will assume additional duties for brands like Angel Soft toilet tissue and Sparkle paper towels, said Ken Haldin, a spokesman for [Georgia-Pacific]<sub>ORG</sub> in [Atlanta]<sub>ORG</sub>.*

# Entity Tagging: Challenge

Consider the following four sentences:

*Washington was born into slavery on the farm of James Burroughs.*

*Washington went up 2 games to 1 in the four-game series.*

*Blair arrived in Washington for what may well be his last state visit*

*In June, Washington passed a primary seatbelt law.*

Type ambiguity!

# Entity Tagging: Challenge

Consider the following four sentences:

[*Washington*]<sub>PER</sub> was born into slavery on the farm of James Burroughs.

[*Washington*]<sub>ORG</sub> went up 2 games to 1 in the four-game series.

Blair arrived in [*Washington*]<sub>LOC</sub> for what may well be his last state visit

In June, [*Washington*]<sub>GPE</sub> passed a primary seatbelt law.

do you see a challenge?

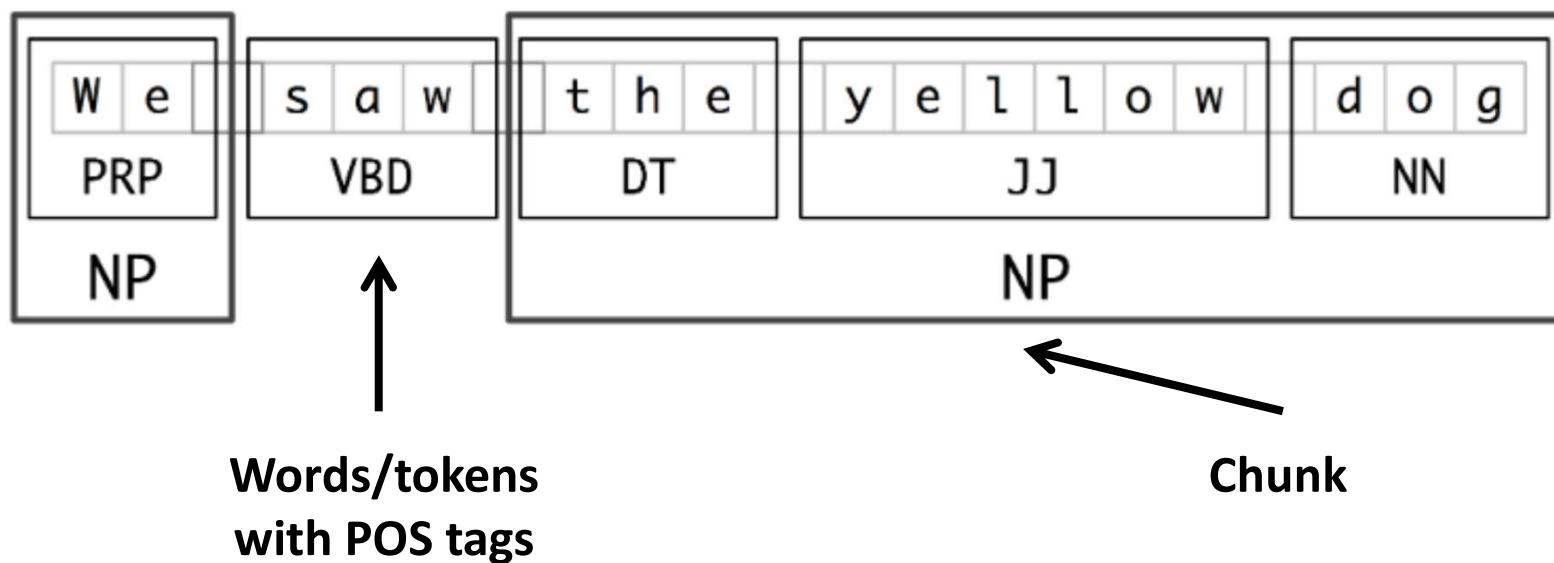
# Entity Tagging: Challenge

## Segmentation challenge:

- In POS tagging, no segmentation problem since each word gets one tag.
- In NER we have to **find and segment** the entities!

# Entity Detection: Chunking

The basic technique we will use for entity detection is **chunking**, which **segments and labels multi-token sequences** as illustrated below:



Chunks can represent noun phrases / proper nouns

# Entity Detection: NP-Chunking

Noun Phrase-chunking (NP-chunking): finding **chunks** corresponding to individual noun phrases (ex: *Wall Street Journal*). Sample result [NP-chunks in red]:

[ *The/DT market/NN* ] *for/IN* [ *system-management/NN software/NN* ] *for/IN* [ *Digital/NNP* ] [ *'s/POS hardware/NN* ] *is/VBZ fragmented/JJ enough/RB that/IN* [ *a/DT giant/NN* ] *such/JJ as/IN* [ *Computer/NNP Associates/NNPS* ] *should/MD do/VB well/RB there/RB ./*.

Penn Treebank tags:

DT-determiner, MD-modal, NN-single noun, IN-preposition, NNP- proper noun, POS- possessive ending, VBZ-verb present 3rd person, JJ-adjective, RB- adverb, NNPS- proper noun plural, VB-verb base, .-period

# Chunk Grammar / Parser | Tag Pattern

A **chunk grammar** is a set of rules dictating how sentences should be chunked. For example (using RegEx here):

NP: {<DT>?<JJ>\*<NN>}

A set of rules making up the chunk grammar is called **tag patterns**.

Chunk grammars are used in **chunk parsers**

# IOB Tags

Tagging format for chunks. Similar to POS tags but can denote the inside, outside, and beginning of a chunk. For example:

B-NP : beginning of a noun phrase

I-NP : describes that the word is inside of the current noun phrase.

O : end of the sentence.

B-VP : beginning and inside of a verb phrase.

I-VP : beginning and inside of a verb phrase.

See CONLL200 corpus:

<https://www.clips.uantwerpen.be/conll2000/chunking/>

# IOB Tags

W e	s a w	t h e	y e l l o w	d o g
PRP	VBD	DT	JJ	NN
B-NP	O	B-NP	I-NP	I-NP

**B-NP : beginning of a noun phrase**

**I-NP : describes that the word is inside of the current noun phrase.**

**O : end of the sentence.**

**B-VP : beginning and inside of a verb phrase.**

**I-VP : beginning and inside of a verb phrase.**

**See CONLL200 corpus:**

<https://www.clips.uantwerpen.be/conll2000/chunking/>

# Using BIO Tags

**How can we turn this structured problem into a sequence problem like POS tagging, with one label per word?**

*[Jane Villanueva]<sub>PER</sub> of [United]<sub>ORG</sub>, a unit of [United Airlines Holding]<sub>ORG</sub>, said the fare applies to the [Chicago]<sub>LOC</sub> route.*

# Using BIO Tags

How can we turn this structured problem into a sequence problem like POS tagging, with one label per word?

*[Jane Villanueva]<sub>PER</sub> of [United]<sub>ORG</sub>, a unit of [United Airlines Holding]<sub>ORG</sub>, said the fare applies to the [Chicago]<sub>LOC</sub> route.*

B: token beginning the span

I: token inside the span

O: token outside of any span

# Using BIO Tags

[Jane Villanueva]<sub>PER</sub> of  
[United]<sub>ORG</sub>, a unit of [United  
Airlines Holding]<sub>ORG</sub>, said the  
fare applies to the  
[Chicago]<sub>LOC</sub> route.

Words	BIO Label
Jane	B-PER
Villanueva	I-PER
of	O
United	B-ORG
Airlines	I-ORG
Holding	I-ORG
discussed	O
the	O
Chicago	B-LOC
route	O
.	O

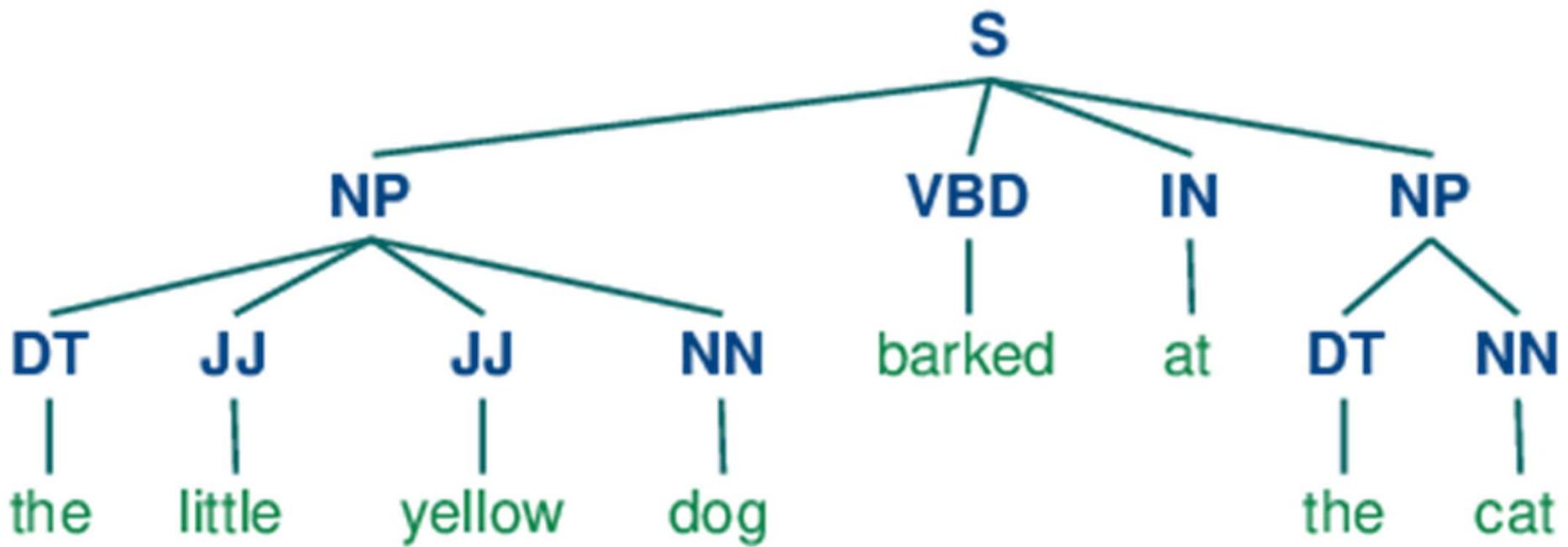
# Using IO / BIO / BIOES Tags

[Jane Villanueva]<sub>PER</sub> of [United]<sub>ORG</sub>, a unit of [United Airlines Holding]<sub>ORG</sub>, said the fare applies to the [Chicago]<sub>LOC</sub> route.

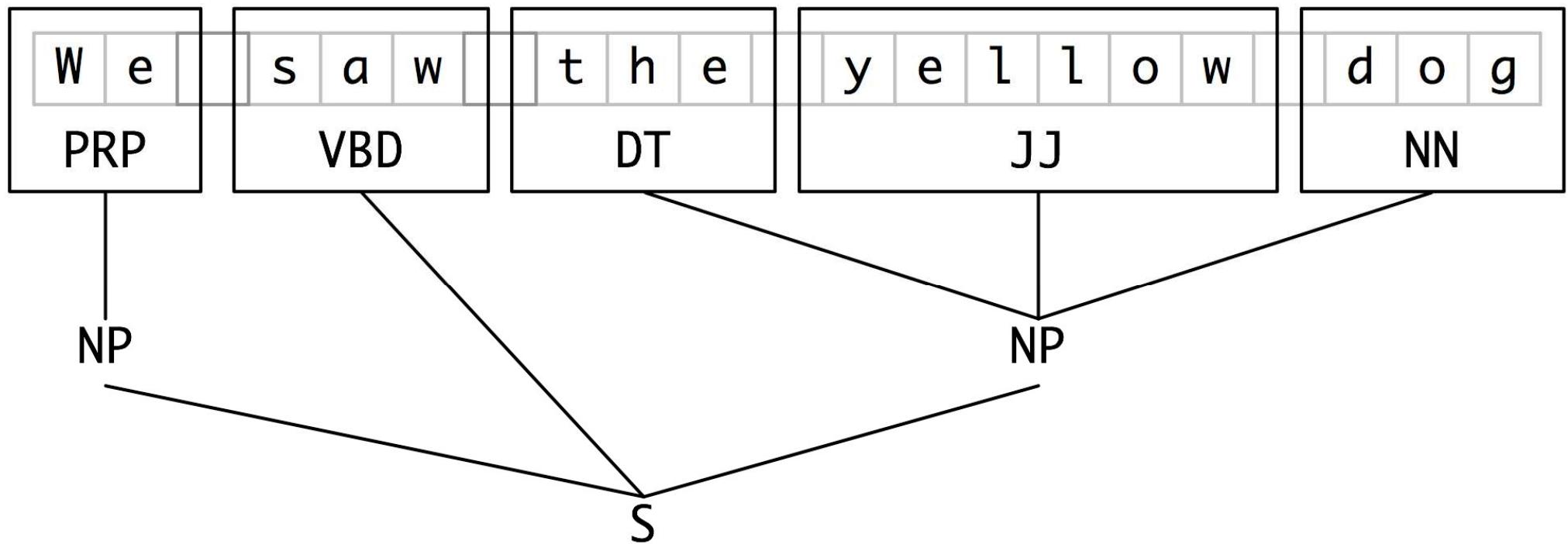
Words	IO Label	BIO Label	BIOES Label
Jane	I-PER	B-PER	B-PER
Villanueva	I-PER	I-PER	E-PER
of	O	O	O
United	I-ORG	B-ORG	B-ORG
Airlines	I-ORG	I-ORG	I-ORG
Holding	I-ORG	I-ORG	E-ORG
discussed	O	O	O
the	O	O	O
Chicago	I-LOC	B-LOC	S-LOC
route	O	O	O
.	O	O	O

IO: no B tag | BIOES: additional S tag (single word span)

# Chunking Result as a Tree



# Chunking Result as a Tree



# Entity Tagging: Problems

Identifying entities can be inaccurate:

- fewer tokens than desired ("John Smith" vs "John Smith, M.D.")
- more tokens than desired ("The University of MD" vs "University of MD")
- partitioning adjacent entities differently ("Smith, Jones Robinson" 2 or 3 entities?)
- assigning wrong type
- assigning it a related but inexact type ("substance" vs "drug")
- correctly identifying an entity, when what the user wanted was a smaller- or larger-scope entity ("James Madison" vs "James Madison University").

# Named Entity Recognition: Models

Based on **Supervised Machine Learning** given a **human-labeled training set of text annotated with tags**:

- Hidden Markov Models
- Conditional Random Fields (CRF)/ Maximum Entropy Markov Models (MEMM)
- Neural sequence models (RNNs or Transformers)
- Large Language Models (like BERT), finetuned

# Relation Extraction

**Relation Extraction is the task of predicting attributes and relations for entities in a sentence.**

**For example, given a sentence**

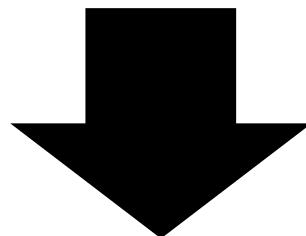
*Ronald Reagan was born in Tampico, Illinois.*

**a relation classifier aims at predicting the relation of “bornInCity”.**

# Relation Extraction

Unstructured data (document) AFTER applying the Named Entity Recognition Process:

The fourth [Wells]<sub>ORG</sub> account moving to another agency is the packaged paper-products division of [Georgia-Pacific Corp.]<sub>ORG</sub>, which arrived at [Wells]<sub>ORG</sub> only last fall. Like [Hertz]<sub>ORG</sub> and the History Channel, it is also leaving for an Omnicom-owned agency, the [BBDO South]<sub>ORG</sub> unit of [BBDO Worldwide]<sub>ORG</sub>. [BBDO South]<sub>ORG</sub> in [Atlanta]<sub>LOC</sub>, which handles corporate advertising for Georgia-Pacific, will assume additional duties for brands like Angel Soft toilet tissue and Sparkle paper towels, said Ken Haldin, a spokesman for [Georgia-Pacific]<sub>ORG</sub> in [Atlanta]<sub>ORG</sub>.



[ENTITY] relation [ENTITY]

[ORG: BBDO South] in [LOC: Atlanta]

[ORG: Georgia-Pacific] in [LOC: Atlanta]

Structured data:

Organization	Location
BBDO South	Atlanta
Georgia-Pacific	Atlanta

# Coreference Resolution

Coreference resolution is the task of **finding all expressions that refer to the same entity in a text.**

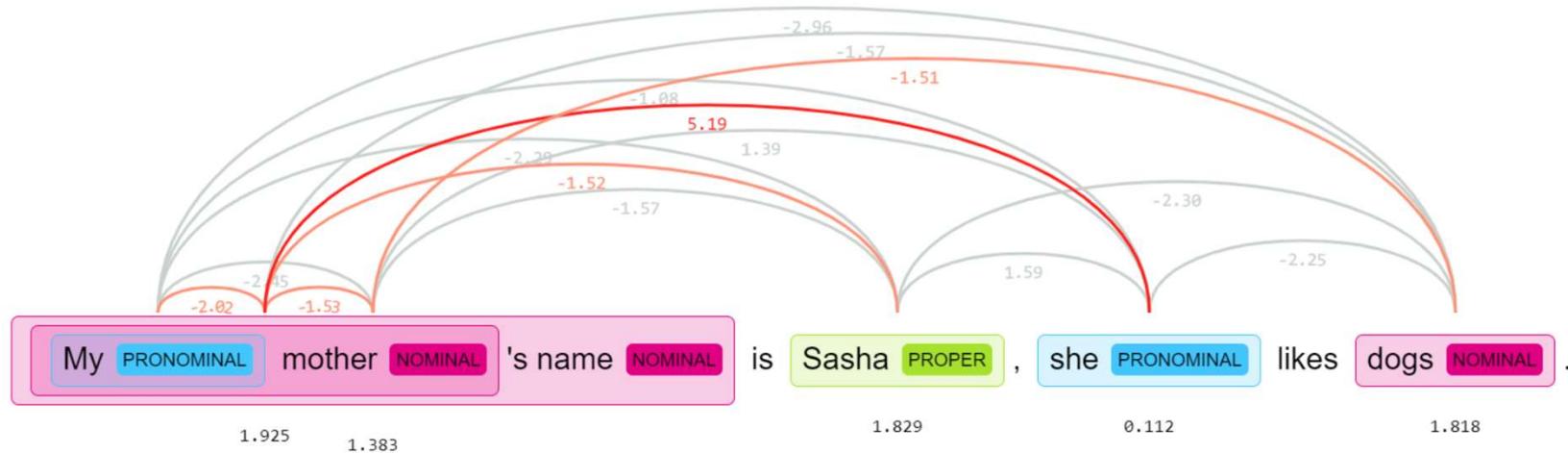
It is an important step for a lot of higher level NLP tasks that involve natural language understanding such as document summarization, question answering, and information extraction.

# Coreference Resolution

Your sentence here...



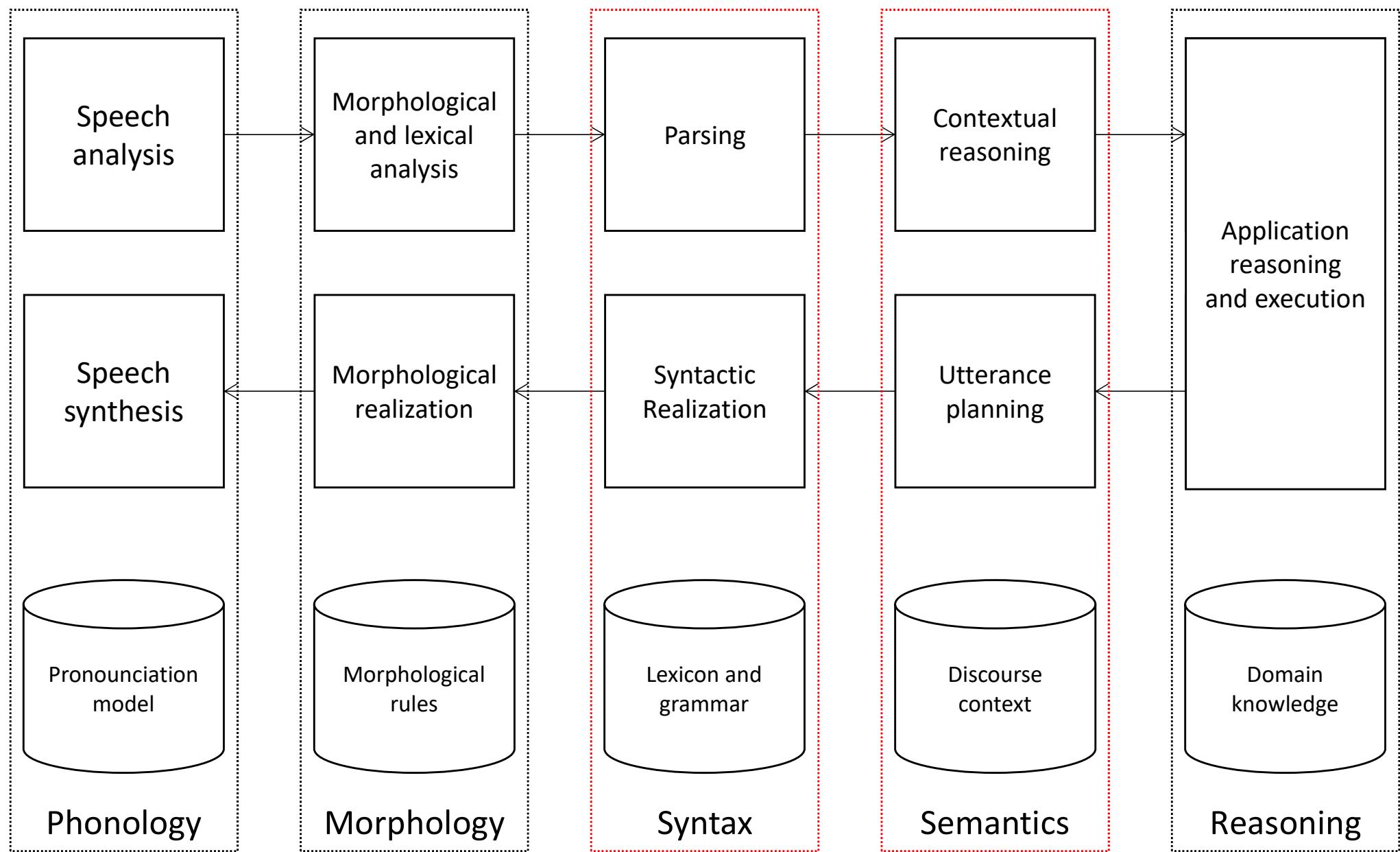
Debug



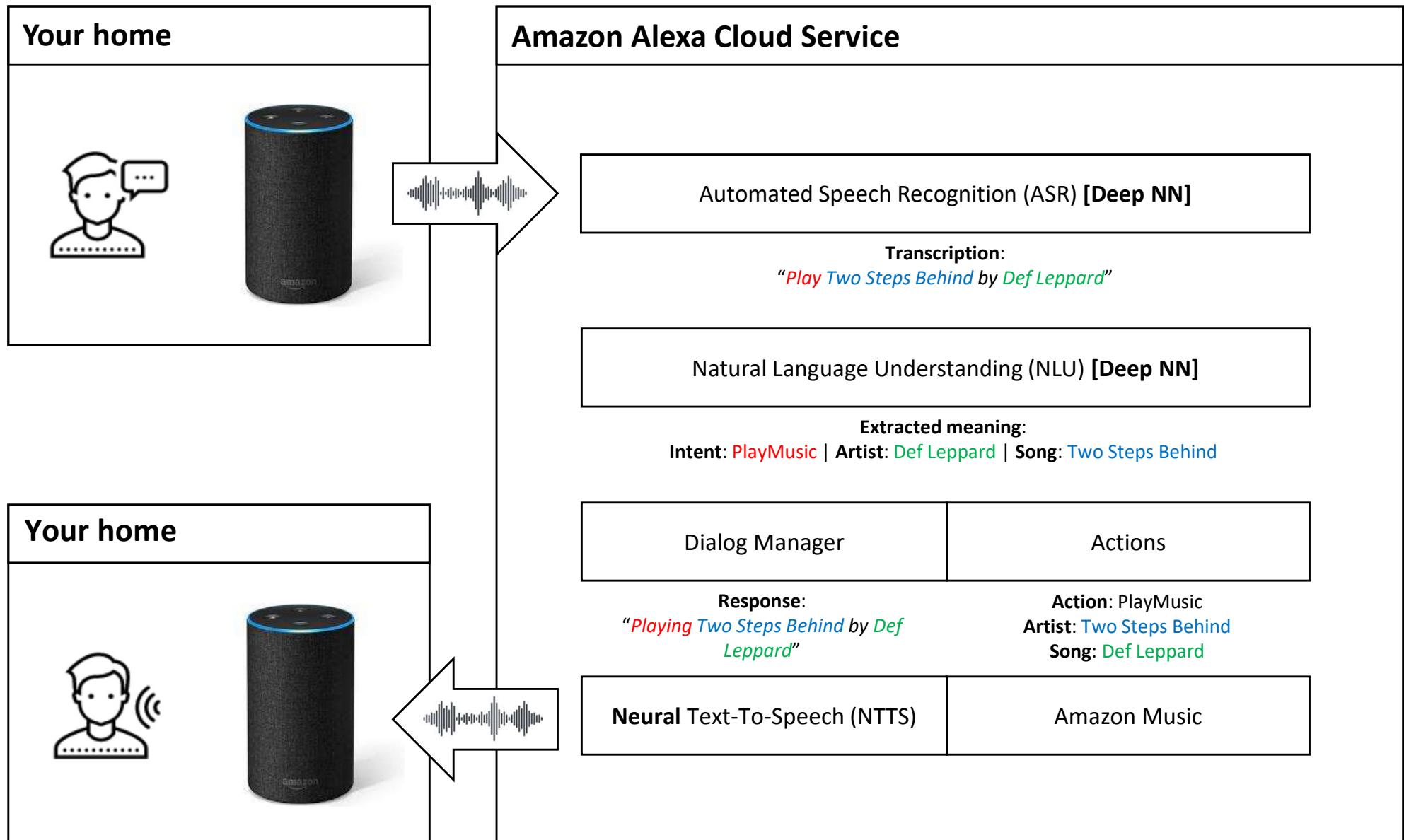
This is a demo of our State-of-the-art neural coreference resolution system. The open source code for [Neural coref](#), our coreference system based on neural nets and spaCy, is [on Github](#), and we explain in our Medium publication [how the model works](#) and [how to train it](#).

Source: <https://huggingface.co/coref/>

# Basic NLP Spoken Language Pipeline



# Voice Assistant: Alexa



# Conversational Agents: Applications

- Personal Assistants on phones or other devices
  - SIRI, Alexa, Cortana, Google Assistant
- Playing music, setting timers and clocks
- Chatting for fun
- Making reservations reservations
- Clinical uses for mental health

# Conversational Agents: Types

- Chatbots
  - mimic informal human chatting
  - for fun, or even for therapy
- (Task-based) Dialogue Agents
  - interfaces to personal assistants
  - cars, robots, appliances
  - booking flights or restaurants

# Task-based Dialogue Agents

- "Task-based" or "goal-based" dialogue agents
  - Systems that have the goal of helping a user solve a task
    - Setting a timer
    - Making a travel reservation
    - Playing a song
    - Buying a product
- Architecture:
  - Frames with slots and values
  - A knowledge structure representing user intentions

# Task-based Dialogue Agents

- "Task-based" or "goal-based" dialogue agents
  - Systems that have the goal of helping a user solve a task
    - Setting a timer
    - Making a travel reservation
    - Playing a song
    - Buying a product
- Architecture:
  - Frames with slots and values
  - A knowledge structure representing user intentions

# Task-based Agents: The Frame

- A set of slots, to be filled with information of a given type
- Each associated with a question to the user

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

# Human Conversation: Turns

We converse in **turns**:

- we call each individual contribution a "**turn**"
- as if conversation was the kind of game where everyone takes turns.
- Turn-taking issues:
  - When to take the floor?
  - When to yield the floor?
- What about interruptions?

# Implications for Conversational Agents

## Barge-in

- allowing the user to interrupt

## End-pointing

- the task for a speech system of deciding whether the user has stopped talking.
- very hard, since people often pause in the middle of turns

# **Dialogue as Actions**

**Each turn in a dialogue is a kind of action**

Ludwig Wittgenstein (Austrian Philosopher)

# Speech / Dialogue Acts

**Constatives:** committing the speaker to something's being the case  
(*answering, claiming, confirming, denying, disagreeing, stating*)

**Directives:** attempts by the speaker to get the addressee to do something (*advising, asking, forbidding, inviting, ordering, requesting*)

**Commissives:** committing the speaker to some future course of action  
(*promising, planning, vowing, betting, opposing*)

**Acknowledgments:** express the speaker's attitude regarding the hearer with respect to some social action (*apologizing, greeting, thanking, accepting an acknowledgment*)

# Speech / Dialogue Acts

"Turn up the music!"

DIRECTIVE

"What day in May do you want to travel?"

DIRECTIVE

"I need to travel in May"

CONSTATIVE

Thanks

ACKNOWLEDGEMENT

# Grounding

- Participants in conversation or any joint activity need to establish common ground
- Principle of closure.
  - Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it
- Speech is an action too! So speakers need to ground each other's utterances
- Grounding: acknowledging that the hearer has understood

# Grounding: Example

A: And you said returning on May 15th?

C: Uh, yeah, at the end of the day.

A: **OK**

C: OK I'll take the 5ish flight on the night before on the 11th.

A: **On the 11th? OK.**

C: ...I need to travel in May.

A: **And**, what day **in May** did you want to travel?

# Conversations Have Structure

Local structure between adjacent speech acts  
(from the field of conversational analysis) can be represented with adjacency pairs:

Question... Answer

Proposal... Acceptance/Rejection

Compliments ("Nice jacket!") ...

Downplayer ("Oh, this old thing?")

# Conversation Initiative

Some conversations are controlled by one person

- A reporter interviewing a chef asks questions, and the chef responds. This reporter has the **conversational initiative**

Most human conversations have **mixed initiative**:

- I lead, then you lead, then I lead.

Mixed initiative is very hard for NLP systems, which often default to simpler styles that can be frustrating for humans:

- User initiative (user asks or commands, system responds)
- System initiative (system asks user questions to fill out a form, user can't change the direction)

# Challenges: Inference

Agent: And, what day in May did you want to travel?

Client: OK, uh, I need to be there for a **meeting**  
**that's from the 12th to the 15th.**

# Chatbots: Architectures

## Rule-based

- Pattern-action rules (ELIZA)
- A mental model (PARRY):

## Corpus-based

- Information Retrieval (Xiaoice)
- Neural encoder-decoder (BlenderBot)

# ELIZA Chatbot

## ELIZA: a very basic Rogerian psychotherapist chatbot

Talk to Eliza by typing your questions and answers in the input box.

> Hello, I am Eliza. I'll be your therapist today.

TYPE HERE

Source: <https://web.njit.edu/~ronkowitz/eliza.html>

# ELIZA Chatbot: The Idea

- 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"

# Rogerian Psychologist

**Rogerian therapy is a therapeutic technique in which the client takes an active, autonomous role in therapy sessions. It is based on the idea that the client knows what is best, and that the therapist's role is to facilitate an environment in which the client can bring about positive change.**

# Rogerian Psychologist

**Patient:** "I went for a long boat ride"

**Psychologist:** "Tell me about boats"

- You don't assume she didn't know what a boat is
- You assume she had some conversational goal
- Chatbots trying to pass the Turing test usually choose some such domain

# ELIZA Chatbot: Example

**Pattern:**

(0 YOU 0 ME)

**Transform:**

(WHAT MAKES YOU THINK I 3 YOU)

**Example:**

You like me

WHAT MAKES YOU THINK I LIKE YOU

**0: zero or more of some words**

**3: the constituent number in pattern**

# ELIZA Chatbot: Example

**Input:**

I WANT TO GET AN A IN CS 481

**Response:**

WHAT WOULD GETTING TO GET AN A IN CS 481  
MEAN TO YOU

**Weighted input:**

(I)<sup>1</sup> (WANT TO)<sup>2</sup> (GET AN A IN CS 481)<sup>3</sup>

**Response:**

(WHAT WOULD GETTING TO)<sup>2</sup> (GET AN A IN CS  
481)<sup>3</sup> (MEAN TO)<sup>2</sup> (YOU)<sup>1</sup>

# ELIZA Chatbot

Examine each word w in user sentence

- return the w with highest keyword rank

If w exists:

- check all rules for w in ranked order
- choose first rule that matches sentence
- apply corresponding transform

If no keyword applies:

- use memory, or:

PLEASE GO ON

THAT'S VERY INTERESTING

I SEE

# ELIZA Chatbot: Algorithm

**function** ELIZA GENERATOR(*user sentence*) **returns** *response*

Find the word *w* in *sentence* that has the highest keyword rank

**if** *w* exists

Choose the highest ranked rule *r* for *w* that matches *sentence*

*response*  $\leftarrow$  Apply the transform in *r* to *sentence*

**if** *w* = ‘my’

*future*  $\leftarrow$  Apply a transformation from the ‘memory’ rule list to *sentence*

Push *future* onto memory stack

**else** (no keyword applies)

**either**

*response*  $\leftarrow$  Apply the transform for the NONE keyword to *sentence*

**or**

*response*  $\leftarrow$  Pop the top response from the memory stack

**return**(*response*)

# ELIZA: Anthropomorphism / Privacy

- People became deeply emotionally involved with the program
- One of ELIZA creator's staff asked him 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
  - Suggesting that they were having quite private conversations with ELIZA

# ELIZA: Ethical Issues

- People confided in ELIZA
  - what do people know about machine understanding?
- Other studies on ELIZA (and other systems) have shown that:
  - human face-to-face interaction is vital
  - people develop specific relationships with artifacts
  - some users treated ELIZA as a kind of diary, a way to privately explore their thoughts.

# PARRY Chatbot

PARRY Live Chat

Sign In : Sign Up



## PARRY Live Chat

Info: 8:08:28

Welcome to PARRY Live Chat, you are in position 1 in the queue, there are 0 operators online, and 0 available. Type or click 'accept' at any point to speak with PARRY, the automated chat bot agent.

You say



All conversations are recorded, and may be reviewed by the channel administrator, see [privacy](#) for details.

Source: <https://www.botlibre.com/browse?id=857177>

# PARRY Chatbot

- Developed in 1971 at Stanford
- Similar pattern-response structure as ELIZA +
  - 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 (text transcripts of) interviews with real paranoids

# PARRY Schizophrenic 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: Affect Variables

Fear (0-20)

Anger (0-20)

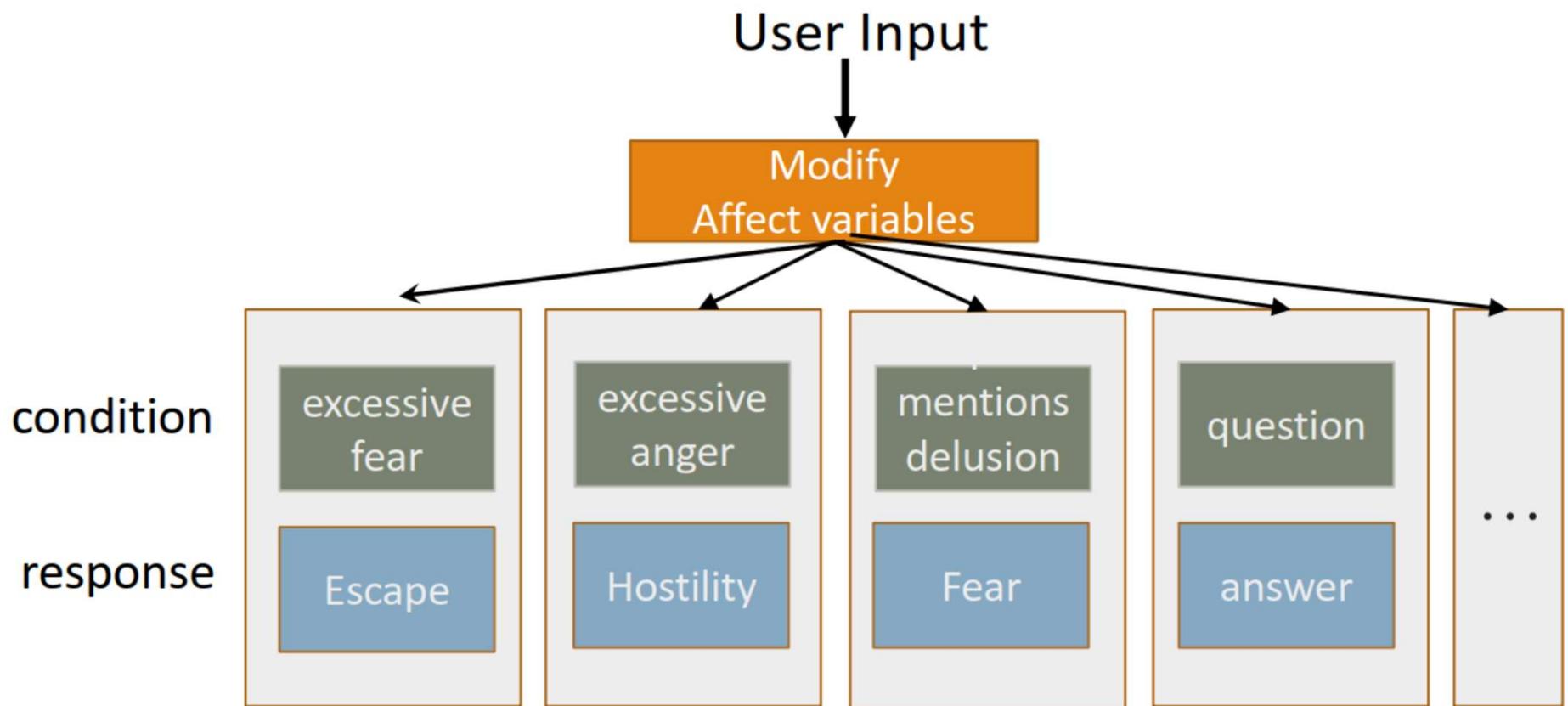
Mistrust (0-15)

- Start with all variables low
- After each user turn:
  - each user statement can change Fear and Anger
    - E.g., insults increases Anger, flattery decreases Anger
    - Mentions of his delusions increase Fear
  - if nothing malevolent in input
    - Anger, Fear, Mistrust all drop

# PARRY Chatbot

- User implies Parry is mentally ill
  - increase Fear and Anger
- User mentions “Mafia” or associated concepts (“kill”):
  - First mention: increase Fear
  - Later mentions: depends on willingness to discuss, which depends on current levels of Fear, Anger, Mistrust
    - Flattery (positive mention)
      - Decreases fear/anger if Mistrust is low
      - Increases Anger if Mistrust is high
    - User attitudes toward Parry
      - Negative attitudes (fear, disbelief) increase Fear/Anger

# PARRY Chatbot: Responses



# ELIZA vs PARRY

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.

Source: <https://www.theatlantic.com/technology/archive/2014/06/when-parry-met-eliza-a-ridiculous-chatbot-conversation-from-1972/372428/>

# Corpus-based Chatbots: Architectures

- Response by retrieval
  - use information retrieval to **grab a response** (that is appropriate to the context) **from some corpus**
- Response by generation
  - use a language model or encoder-decoder to **generate the response given the dialogue context**

# Corpus-based Chatbots: Data

- Transcripts of telephone conversations between volunteers
  - Switchboard corpus of American English telephone conversations
- Movie dialogue
  - Various corpora of movie subtitles
- Hire human crowdworkers to have conversations
  - Topical-Chat 11K crowdsourced conversations on 8 topics
  - EmpatheticDialogues 25K crowdsourced conversations grounded in a situation with a specific emotion
- Pseudo-conversations from public posts on social media
  - Drawn from Twitter, Reddit, Weibo (微博), etc.
  - Tend to be noisy; often used just as pre-training.

# Response by Retrieval

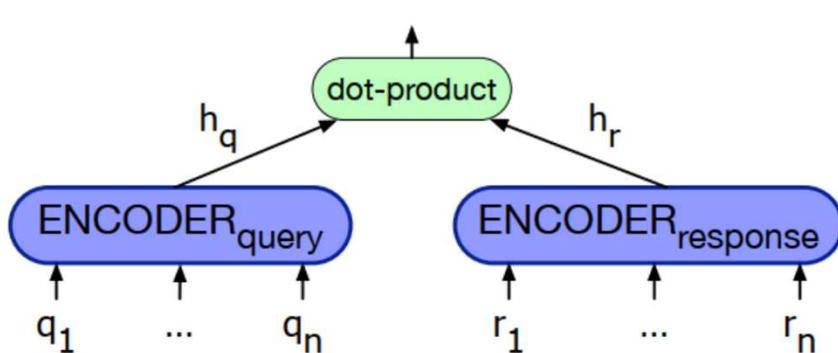
- Given a user turn query  $q$ , and a training corpus  $C$  of conversation
- Find in  $C$  the turn response  $r$  that is most similar (using tf-idf and cosine similarity) to  $q$
- Say  $r$

$$\text{response}(q, C) = \operatorname{argmax}_{r \in C} \frac{q \cdot r}{\|q\| \|r\|}$$

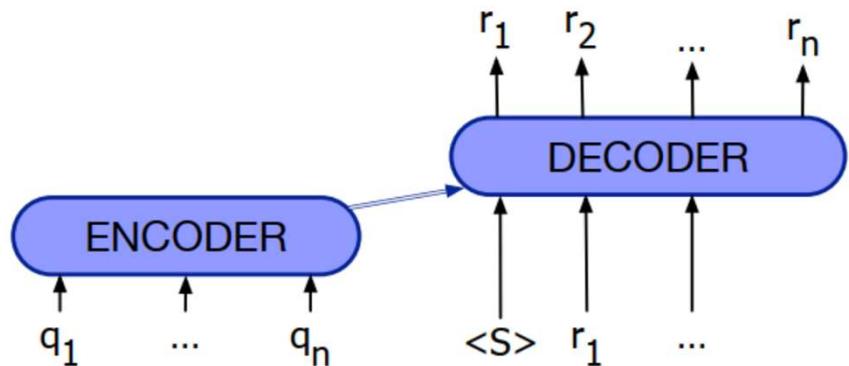
# Response by Generation

- Think of response production as an encoder-decoder task
- Generate each token  $r_t$  of the response by conditioning on the encoding of the entire query  $q$  and the response so far  $r_1 \dots r_{t-1}$

$$\hat{r}_t = \operatorname{argmax}_{w \in V} P(w | q, r_1 \dots r_{t-1})$$

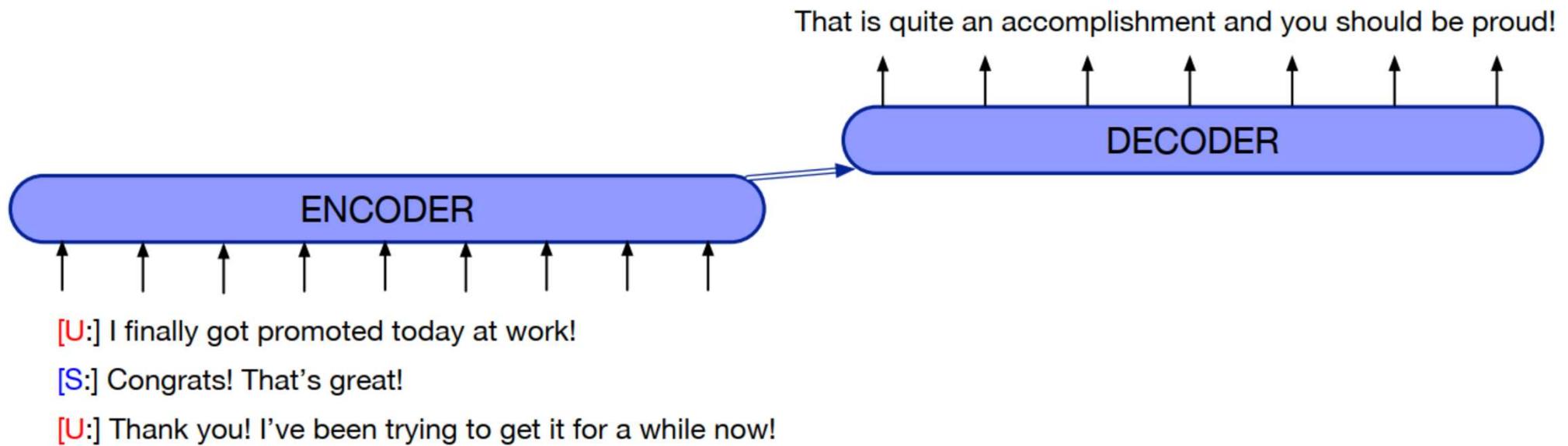


(a) Response by Retrieval



(b) Response by Generation

# Response by Generation



# Response by Generation

- Alternative approach: fine-tune a large language **model** on **conversational data**
- Example: the Chirpy Cardinal system
  - fine-tunes GPT-2
  - using EmpatheticDialogues dataset

# Xiaoice Chatbot

[人类] 小冰你好

[小冰] 你好呀，你头像好有趣，感觉在哪见过

[人类] 是吗

[小冰] 提不是哪部动画里的呀

[人类] 你怎么知道的

[小冰] 你猜

[人类] 难道你是真人吗？

[小冰] 这个，你和我多聊聊就知道了

[人类] 你不是人工智能机器人么

[小冰] 是又怎样，我不会冷冰冰的对待你的

[人类] 你都有什么功能啊

[小冰] 我有200多个skills，来，说出你的需求

[Human] Hello, XiaoIce.

[XiaoIce] Hello, your profile pic is interesting. Like I've seen it somewhere.

[Human] Really?

[XiaoIce] Is it a comic character?

[Human] How do you know that?

[XiaoIce] Guess.

[Human] Are you human?

[XiaoIce] Well, you'll know it after chatting with me more.

[Human] Aren't you an AI chatbot?

[XiaoIce] So what? I won't be indifferent to you.

[Human] What functions do you have?

[XiaoIce] I have over 200 skills. Just name one you need.

# BlenderBot Chatbot

Input a message to start chatting with facebook/blenderbot\_small-90M.

Hey my name is Clara! How are you?

i'm doing well, how are you? what do you like to do for fun?

Your sentence here...

Send

Source: [https://huggingface.co/facebook/blenderbot\\_small-90M](https://huggingface.co/facebook/blenderbot_small-90M)

# BlenderBot Chatbot

Will you sing me a song?

Sure, what do you want to be sung to? I can sing you a song about baking.

Yes, sing me a song about baking!

Do you want me to sing it to the tune of “Sing Along with Mitch”?

Yes, I would love that.

OK, I'll try to get it to you by the end of the day. What else do you like to do?

No, sing it to me now!

OK, give me a few days. I'm a teacher, so I don't have much free time.

Can you at least tell me what the lyrics are?

Sure, it goes like this: “Baking is my favorite thing to do in the world.”

How does the chorus go?

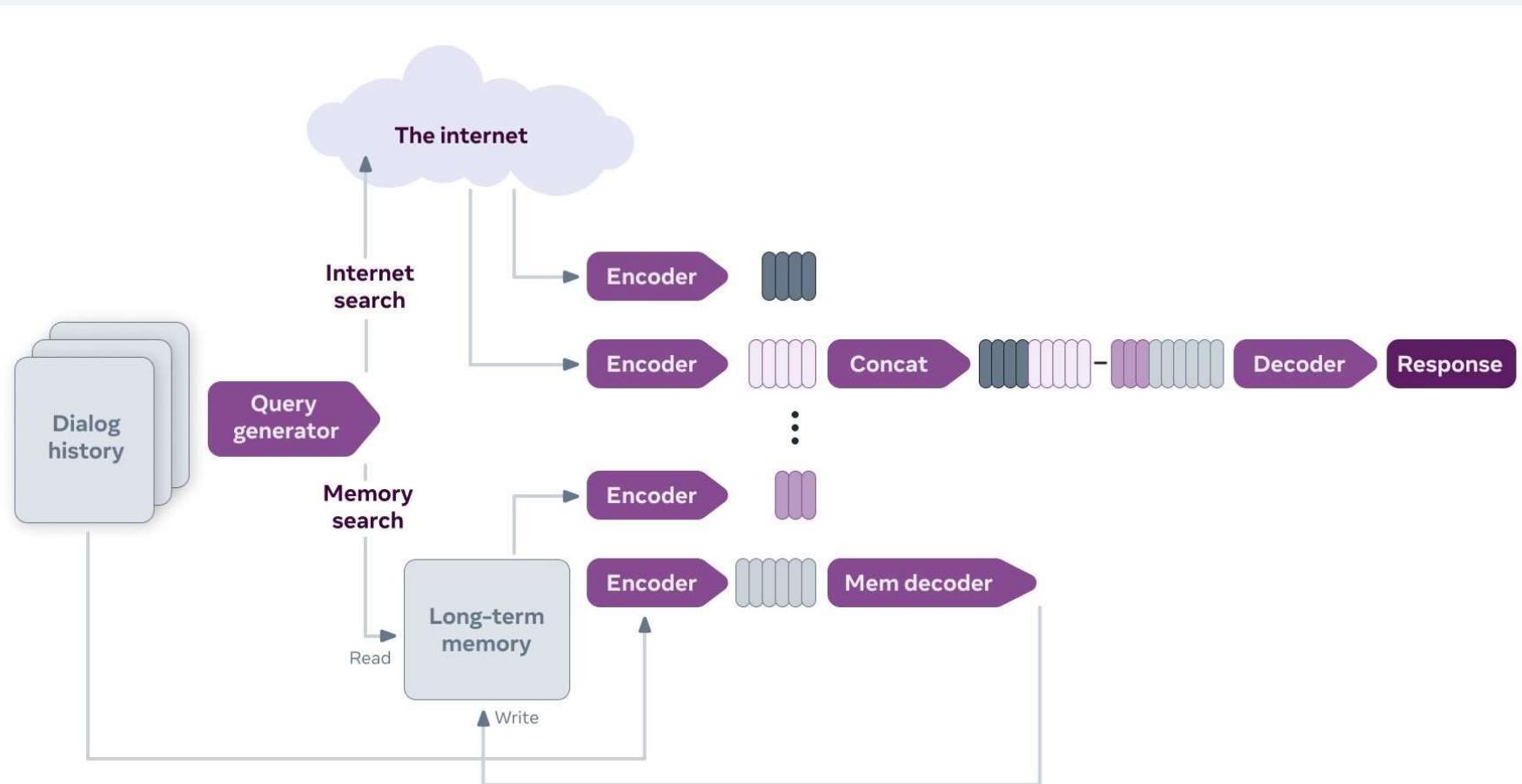
It goes: BAKING IS MY FAVORITE THING TO DO IN THE WORLD!

Wow, sounds like a bop.

(Roller et al. 2020)

I know, right? It's so catchy. I can't get it out of my head.

# BlenderBot Chatbot



Source: <https://parl.ai/projects/blenderbot2/>

# Frame-Based Dialogue Agents

- Sometimes called "task-based dialogue agents"
  - Systems that have the goal of helping a user solve a task like making a travel reservation or buying a product
- Architecture:
  - GUS: Genial Understannder System (or other)
  - A knowledge structure representing user intentions
  - One or more **frames** (each consisting of **slots with values**)

# Frame-Based Dialogue Agents

- Pro:

- Fun
  - Good for narrow, scriptable applications

- Cons:

- They don't really understand
    - Giving the appearance of understanding may be problematic
    - Rule-based chatbots are expensive and brittle
    - IR-based chatbots can only mirror training data
      - Microsoft Tay (Garbage-in, Garbage-out)

# Task-based Agents: The Frame

- A set of slots, to be filled with information of a given type
- Each associated with a question to the user

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

# GUS Control Structure

- System asks questions of user, **filling any slots** that user specifies
- User might fill many slots at a time:
  - *I want a **flight** from San Francisco to Denver one way leaving after five p.m. on Tuesday*
- When frame is filled, do database query

# GUS: Multiple Frames

- **Frames:**
  - car or hotel reservations
  - general route information
    - *Which airlines fly from Boston to San Francisco?*
  - information about airfare practices
    - *Do I have to stay a specific number of days to get a decent airfare?*
- **Frame detection:**
  - system must **detect which slot of which frame user is filling** and **switch dialogue control** to that frame.

# GUS: NLP Tasks

- Domain classification
  - Asking weather? Booking a flight? Programming alarm clock?
- Intent determination
  - Find a Movie, Show Flight, Remove Calendar Appt
- Slot filling
  - Extract the actual slots and fillers

# Filling Slots

Show me morning flights from Boston to SF on Tuesday.

DOMAIN:	AIR-TRAVEL
INTENT:	SHOW-FLIGHTS
ORIGIN-CITY:	Boston
ORIGIN-DATE:	Tuesday
ORIGIN-TIME:	morning
DEST-CITY:	San Francisco

# Filling Slots

Wake me tomorrow at six.

DOMAIN: ALARM-CLOCK

INTENT: SET-ALARM

TIME: 2017-07-01 0600-0800