

Introduction to NLP

By
Mirza Rahim Baig

Agenda

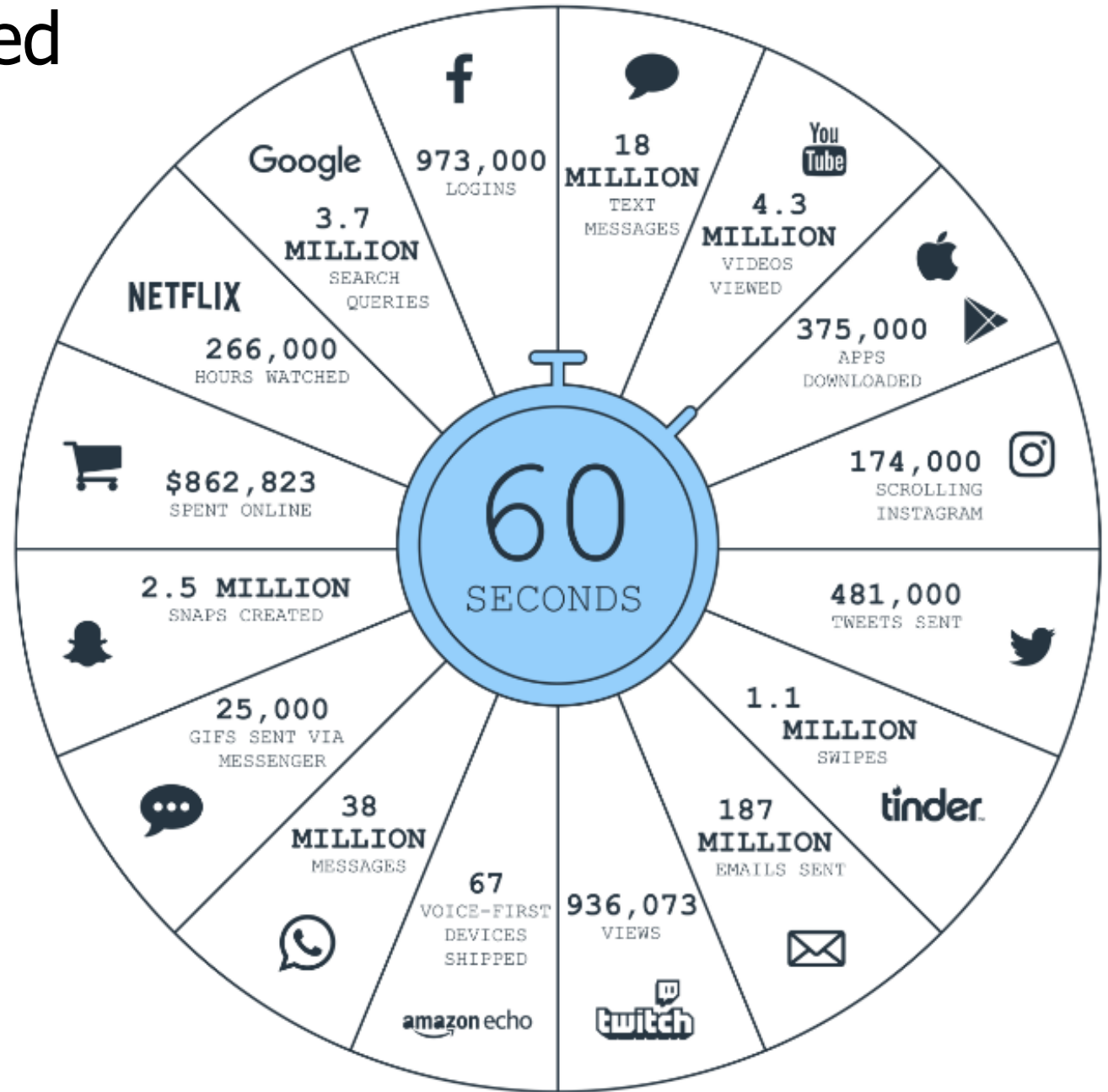
- Motivation for Natural Language processing
- Human language and why it is special
- Why NLP is hard
- Working with text - tasks in NLP



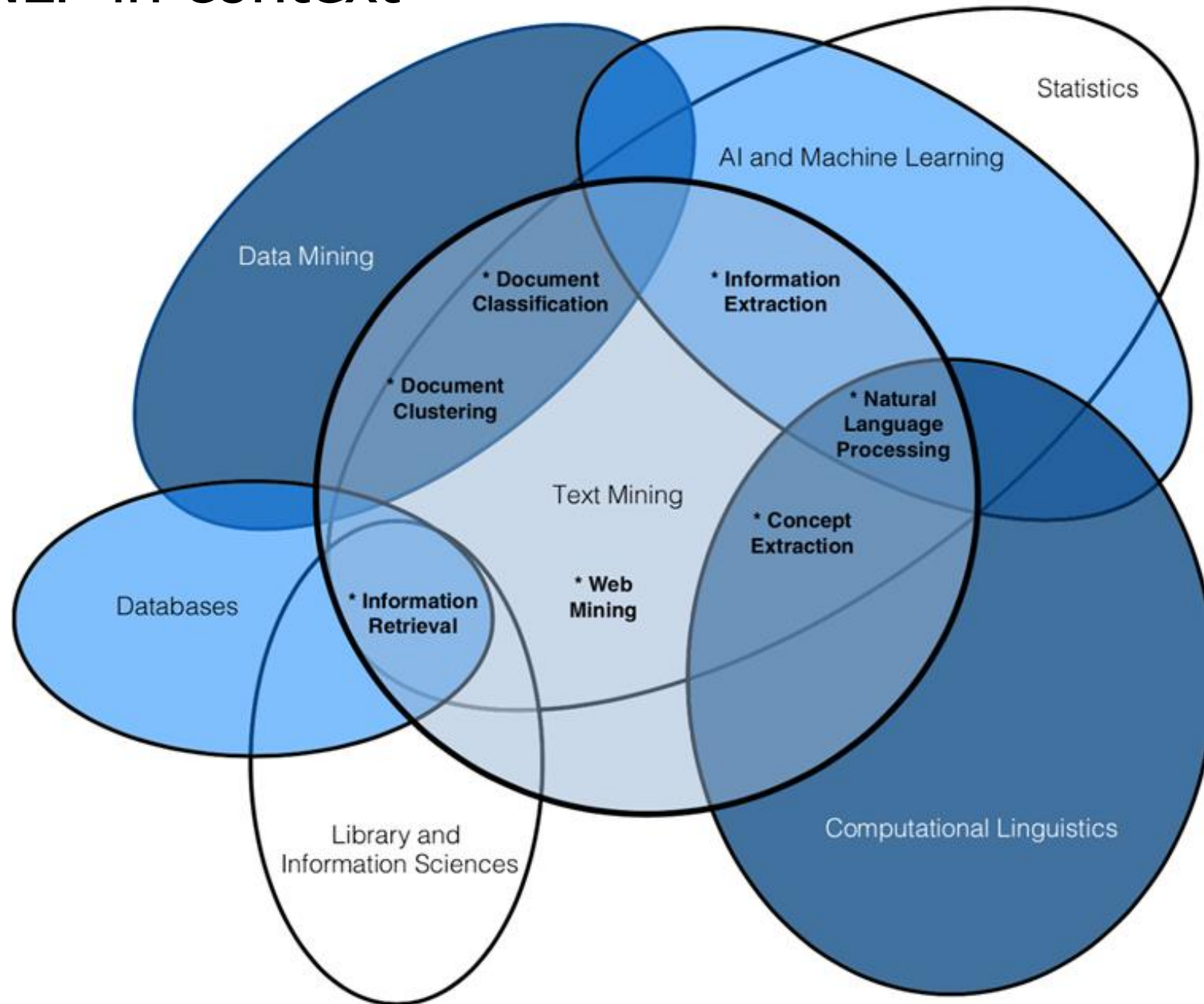
The "Internet minute" revisited

***Most of the data are unstructured;
most of which is text!***

Text data are severely underutilized



Putting NLP in context



Applications of NLP

1. Search (written and spoken)
2. Online advertisement matching
3. Automated /assisted translation
4. Sentiment analysis for marketing or finance/trading
5. Speech recognition
6. Chatbots/dialog agents
 - a. Automated customer support
 - b. Controlling devices
 - c. Ordering goods

Human language and why it is special

Human Language

Human language is a system specifically constructed to convey the speaker's meaning

- Not just an environmental signal, it is deliberate communication
- Using an encoding which little kids can learn very, very quickly

Human language is a discrete/symbolic/categorical signaling system

- Signals communicate ideas

A little bit more on language

Symbols and their meanings

5

(n) The cardinal number that is the sum of four and one

Schadenfruede

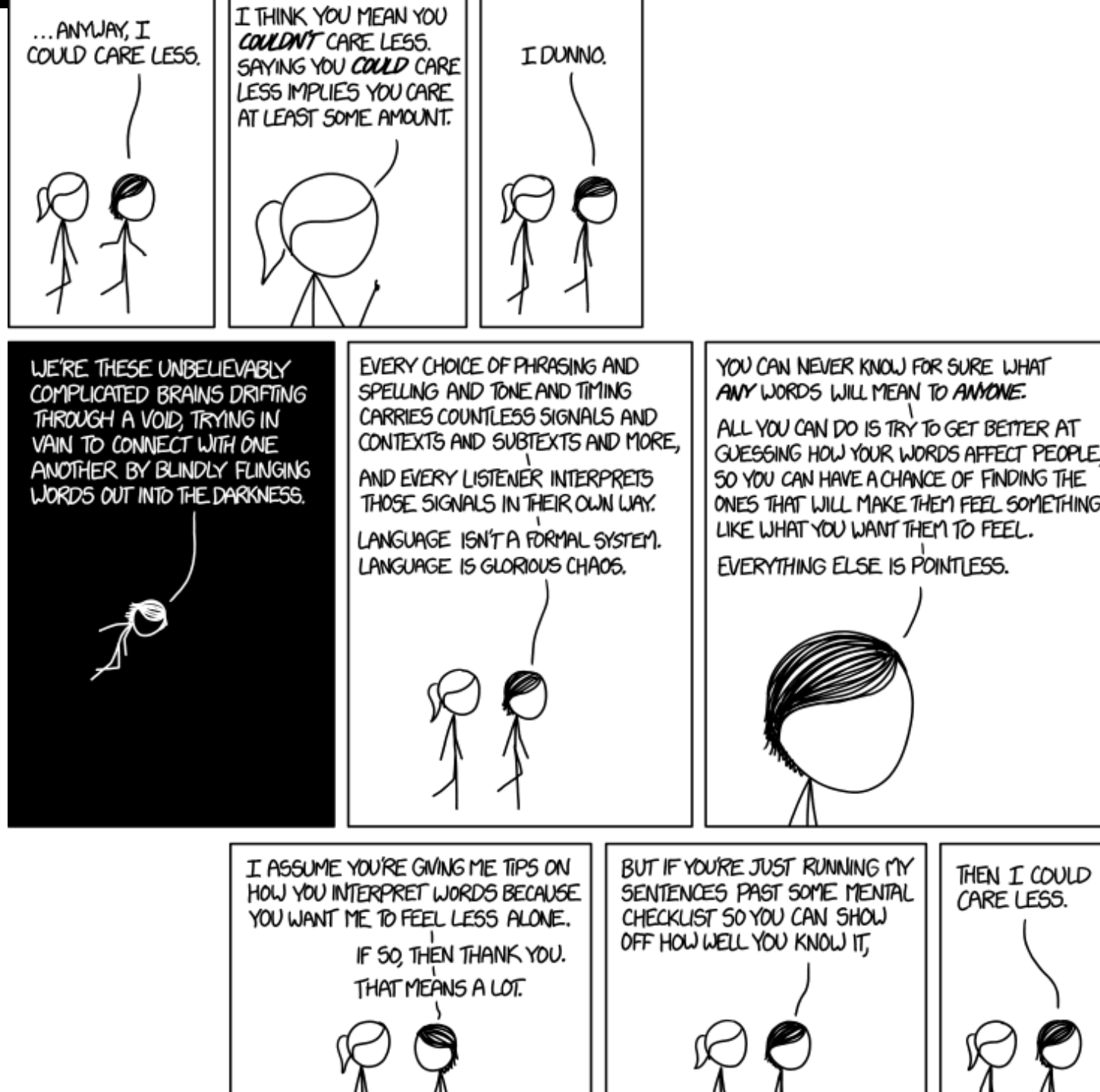
(n) Delight in another person's misfortune

sonder*

(n) the realization that each random passerby is living a life as vivid and complex as your own

* Non-standard

Language ...



... isn't simple

Why NLP is hard

Examples

Who does 'he' refer to in the following?

Rahim helped Pooja. He was kind.

Rahim helped Mohan. He was kind.

Rahim helped Kiran. He was kind.

Examples

What is the topic of discussion?

The stars are out tonight
at the 19th Annual Filmfare awards

Examples

Who had the knife?

San Jose cops kill man with knife

Close

Text

Paper

Translate

Listen

San Jose cops kill man with knife

Ex-college football player, 23, shot 9 times allegedly charged police at fiancée's home

By Hamed Aleaziz and Vivian Ho

A man fatally shot by San Jose police officers while allegedly charging at them with a knife was a 23-year-old former football player at De Anza College in Cupertino who was distraught and depressed, his family said

Thursday.

Police officials said two officers opened fire Wednesday afternoon on Phillip Watkins outside his fiancée's home because they feared for their lives. The officers had been drawn to the home, officials said, by a 911 call reporting an armed home invasion

that, it turned out, had been made by Watkins himself.

But the mother of Watkins' fiancée, who also lives in the home on the 1300 block of Sherman Street, said she witnessed the shooting and described it as excessive. Faye Buchanan said the confrontation happened

shortly after she called a suicide intervention hotline in hopes of getting Watkins medical help.

Watkins' 911 call came in at 5:01 p.m., said Sgt. Heather Randol, a San Jose police spokeswoman. "The caller stated there was a male breaking into his home armed with a knife," Randol said. "The caller also stated he was locked in an upstairs bedroom with his children and request-

ed help from police."

She said Watkins was on the sidewalk in front of the home when two officers got there. He was holding a knife with a 4-inch blade and ran toward the officers in a threatening manner, Randol said.

"Both officers ordered the suspect to stop and drop the knife," Randol said. "The suspect continued to charge the officers with the knife in his hand. Both officers, fear-

ing for their safety and defense of their life, fired at the suspect."

On the police radio, one officer said, "We have a male with a knife. He's walking toward us."

"Shots fired! Shots fired!" an officer said moments later.

A short time later, an officer reported, "Male is down. Knife's still in hand."

Buchanan said she had been prompted to call the

Shoot continues on D8

Back Continue

Some more examples of ambiguity

Word level ambiguity:

- "design" can be a noun or a verb (ambiguous POS)
- "root" has multiple meanings (ambiguous sense)

Syntactic ambiguity:

- "A man saw a boy with a telescope." (PP Attachment)

Anaphora resolution:

- "John persuaded Bill to buy a TV for himself"

Presupposition:

- "He has quit smoking" implies that he smoked before.

And we haven't even talked of poetry yet!

Why NLP is hard

Biggest problem -

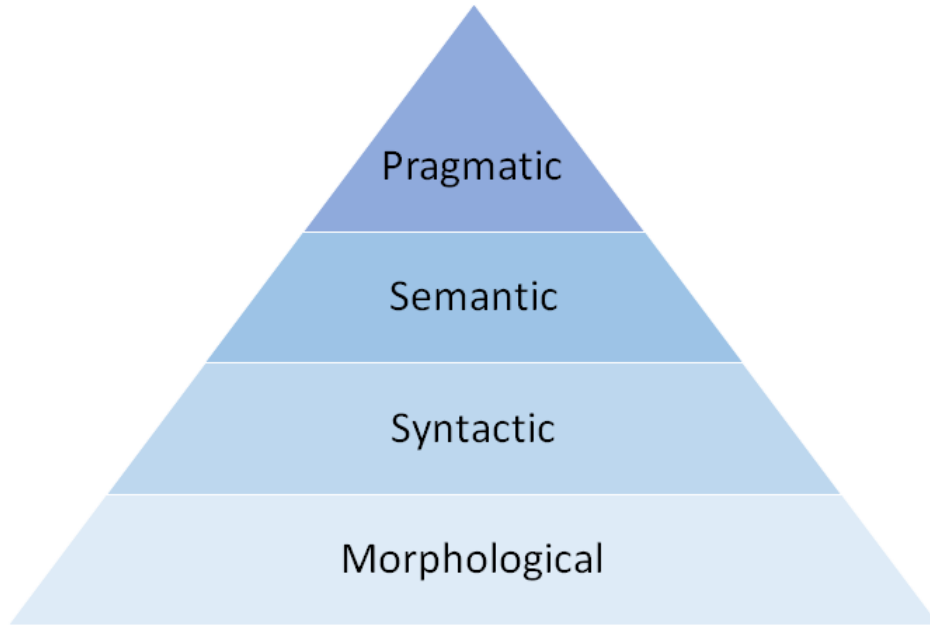
Humans use language efficiently by **not** saying most things

- We leave out most details, and the user figures out the rest
- Minimal messages transmitted, listener fills the rest using his world knowledge, common sense, contextual understanding
- We keep a lot of ambiguities, which we assume the hearer/reader knows how to resolve.

Ambiguity is unambiguously problematic!

Tasks in NLP - working with natural language

Natural Language Processing Pyramid



Pragmatic: discourse and contextual awareness

Semantic: the meaning of the words, sentences

Syntactic: structure of the language, relationship between words, and the roles they play

Morphology: words, their forms, variations

Rule based methods

- Regular expressions
- Context free grammars
- ...

Probabilistic modeling and machine learning

- Likelihood maximization
- Classifiers
- ...

Deep Learning

- Recurrent Neural Networks
- Convolutional Neural Networks
- ...

Remark: Deep Learning vs. Traditional NLP

“Oh, for *** sake, DL people, leave language alone and stop saying you solve it.”**

- Yoav Goldberg

Question: If Deep learning is so great, why do we even need to study traditional NLP?

- Perform good in many tasks
- To not be blinded by the hype
- Can help to further improve DL models

Coming up next:

The foundations required for NLP



THANK YOU

All product details and company names used or referred in this work are copyright and trademarks or registered trademarks of their respective holders. Use of them in this work does not imply any affiliation with or endorsement by them.

This work contains a variety of intellectual property rights including trademark and copyrighted material. Unless stated otherwise, Manipal Global Education Services Pvt Ltd ("Company"), owns the intellectual property for all the information provided on this work, and some material is owned by others which is clearly indicated, and other material may be in the public domain. Except for material which is unambiguously and unarguably in the public domain, permission is not given for any commercial use or sale of this work or any portion or component hereof. You may view or download information for personal use only. Any unauthorized access to, review, publish, adapt, copy, share, reproduction, dissemination or other use of the information contained herein is strictly prohibited.

All material on this site is subject to copyright under Indian law and through international treaties, and applicable law in other countries. Company respects the intellectual property rights of others. If you believe your copyright has been violated in such a way that it constitutes a copyright infringement or a breach of a contract or license, we request you to notify our designated representative on the contact column of the website.