

# A Very Short Course on Practical NLP in the Age of LLMs

Spring 2023, IISc

# People

- **Instructors:**

- Monojit Choudhury, Microsoft Turing
- Kabir Ahuja, Microsoft Research
- Harshita Diddee, Microsoft Research

- **Coordinator:**

- Prof Chiranjib Bhattacharya, CSA, IISc



LECTURE 1

# A History of NLP through the lens of Machine Translation

24<sup>th</sup> Feb 2023

It is not difficult to devise a paper machine which will play a not very bad game of chess... Are there imaginable digital computers which would do well in the imitation game?

Alan Turing, 1950



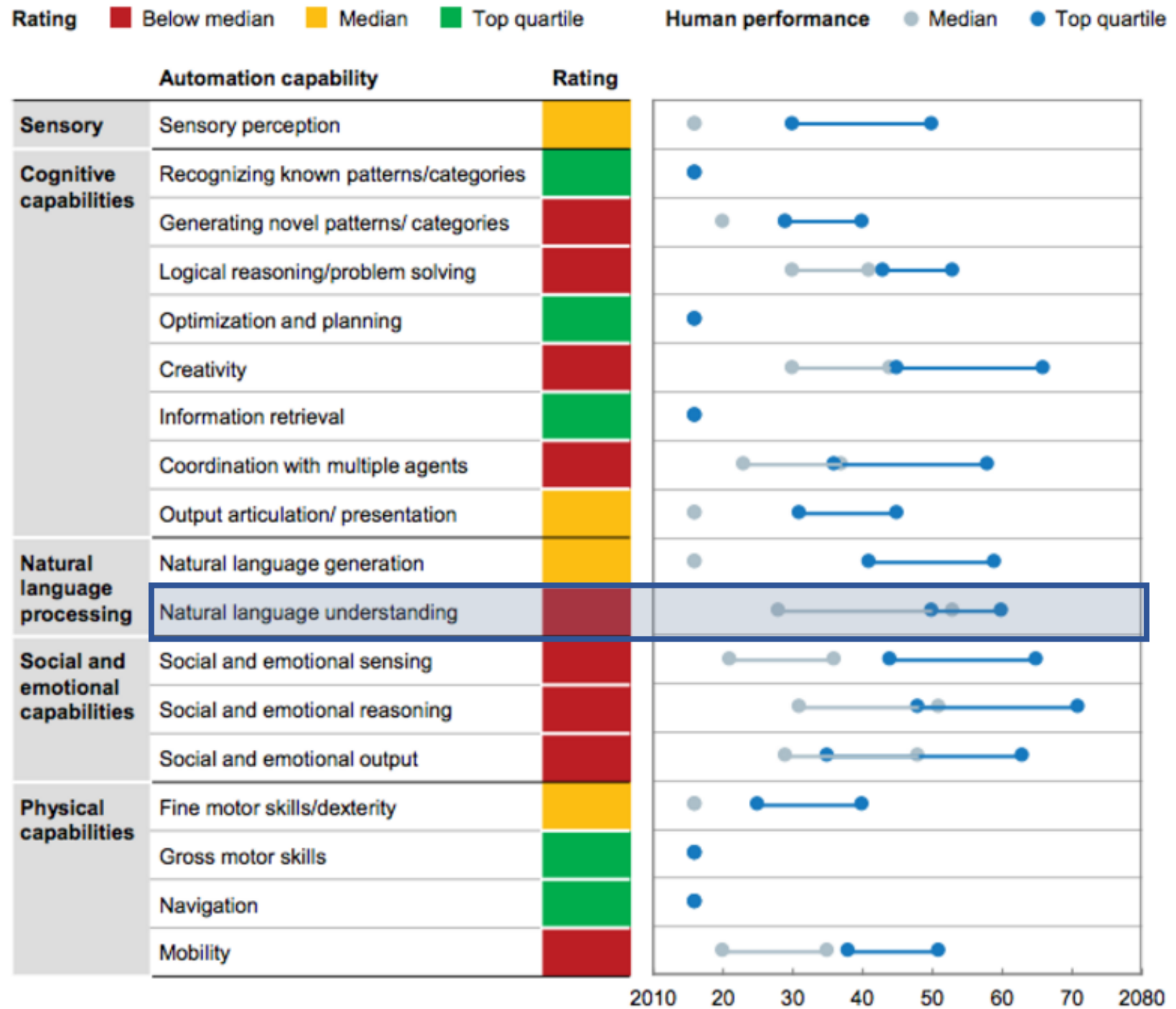
The biggest breakthrough in AI will be when computers can read and understand information like humans do.

Bill Gates, 2017

*Language Understanding*  $\equiv$  *Artificial General Intelligence*  $\equiv$  *AI-complete*

---

How long  
will it take  
for AI to  
surpass  
humans?

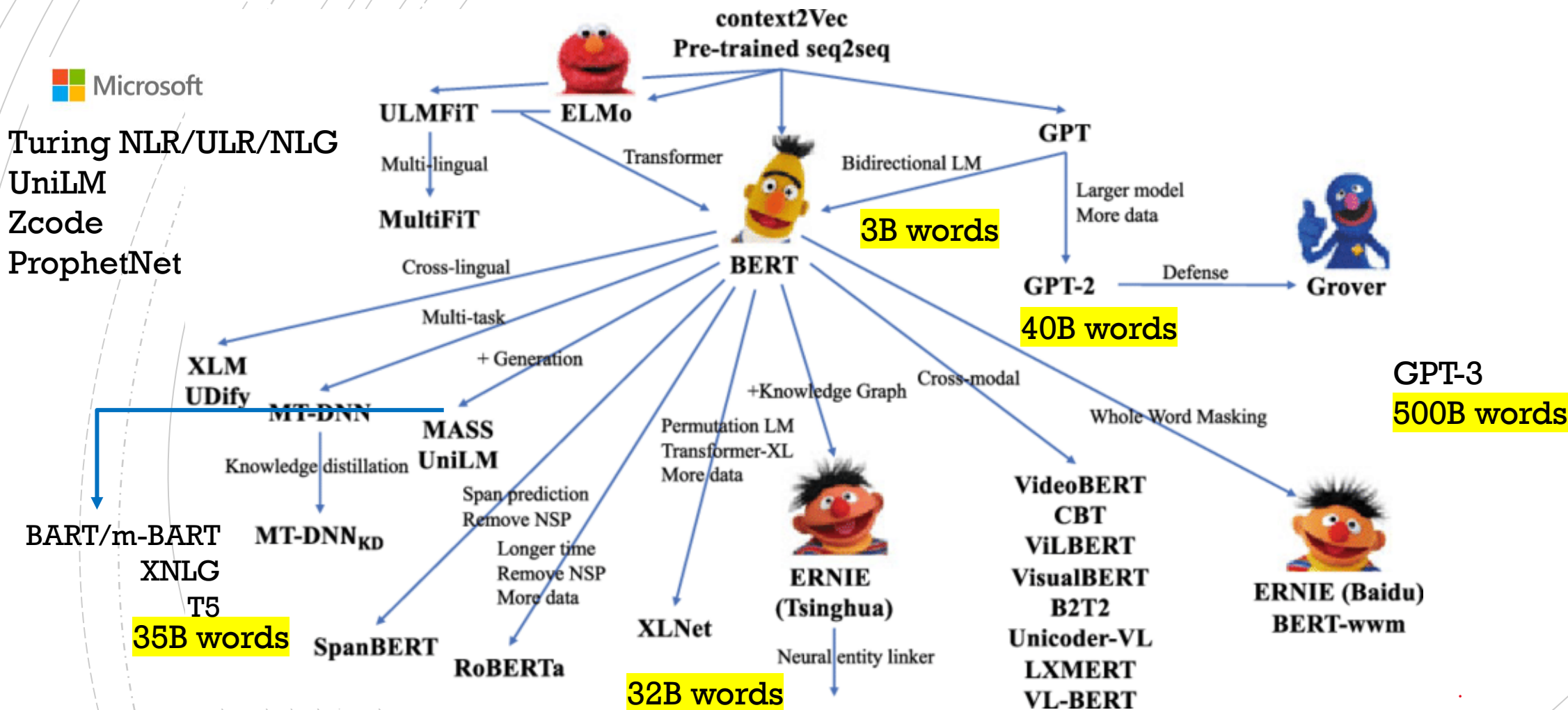


What's the  
big deal?

March 2018



# The Pretrained Language Model family





# A sonnet on IISc by ChatGPT

*In Bangalore, a place of wondrous fame  
There lies a seat of learning and of lore  
Where scholars strive to leave their lasting claim  
On knowledge's vast and ever-changing shore*

*The Indian Institute of Science, they call  
This institution born of lofty thought  
Where intellect and reason stand so tall  
And learning's fire is diligently wrought*

*From all around the world, they come to learn  
To study and to seek a greater truth  
And in their quest, they do not cease or turn  
For in their minds, they see eternal youth*

*So here's to IISc, a noble place  
Where learning's flame forever will embrace.*

# ChatGPT on Course on MT

If we have to cover topics on Machine Translation in just 5 sessions of 3 hours each, what would be a good set of topics to cover? Break them down by session.

Sure, here is a possible breakdown of topics on Machine Translation that can be covered in five sessions of three hours each:

## **Session 1: Introduction to Machine Translation**

Overview of Machine Translation (MT) and its importance

Types of MT: Rule-based MT, Statistical MT, Neural MT

MT evaluation metrics: BLEU, NIST, METEOR

Challenges in MT: Ambiguity, Context, Idioms, Named Entities, etc.

Recent advancements in MT: Transfer Learning, Pretrained Language Models, Multilingual MT

## **Session 2: Rule-based Machine Translation**

Rule-based MT pipeline: Morphological Analysis, Syntactic Analysis, Semantic Analysis, Lexical Transfer, Morphological Generation

Examples of Rule-based MT systems: Apertium, Moses, etc.

Advantages and disadvantages of Rule-based MT

## **Session 3: Statistical Machine Translation**

Statistical MT pipeline: Corpus Collection and Preprocessing, Word Alignment, Phrase Extraction, Language Modelling, Decoding

Examples of Statistical MT systems: Google Translate, Bing Translator, etc.

Advantages and disadvantages of Statistical MT

## **Session 4: Neural Machine Translation**

Neural MT pipeline: Word Embedding, Encoder-Decoder Architecture, Attention Mechanism, Decoding

Examples of Neural MT systems: Google Neural Machine Translation, Facebook FAIR MT, etc.

Advantages and disadvantages of Neural MT

## **Session 5: MT Applications and Future Directions**

Applications of MT: Website Localization, Subtitling, Speech Translation, etc.

MT challenges in low-resource languages

Future directions in MT: Multimodal MT, Interactive MT, Zero-shot MT, etc.

Ethical considerations in MT

This breakdown is just a suggestion, and depending on the audience and their level of expertise, the topics can be adjusted accordingly.

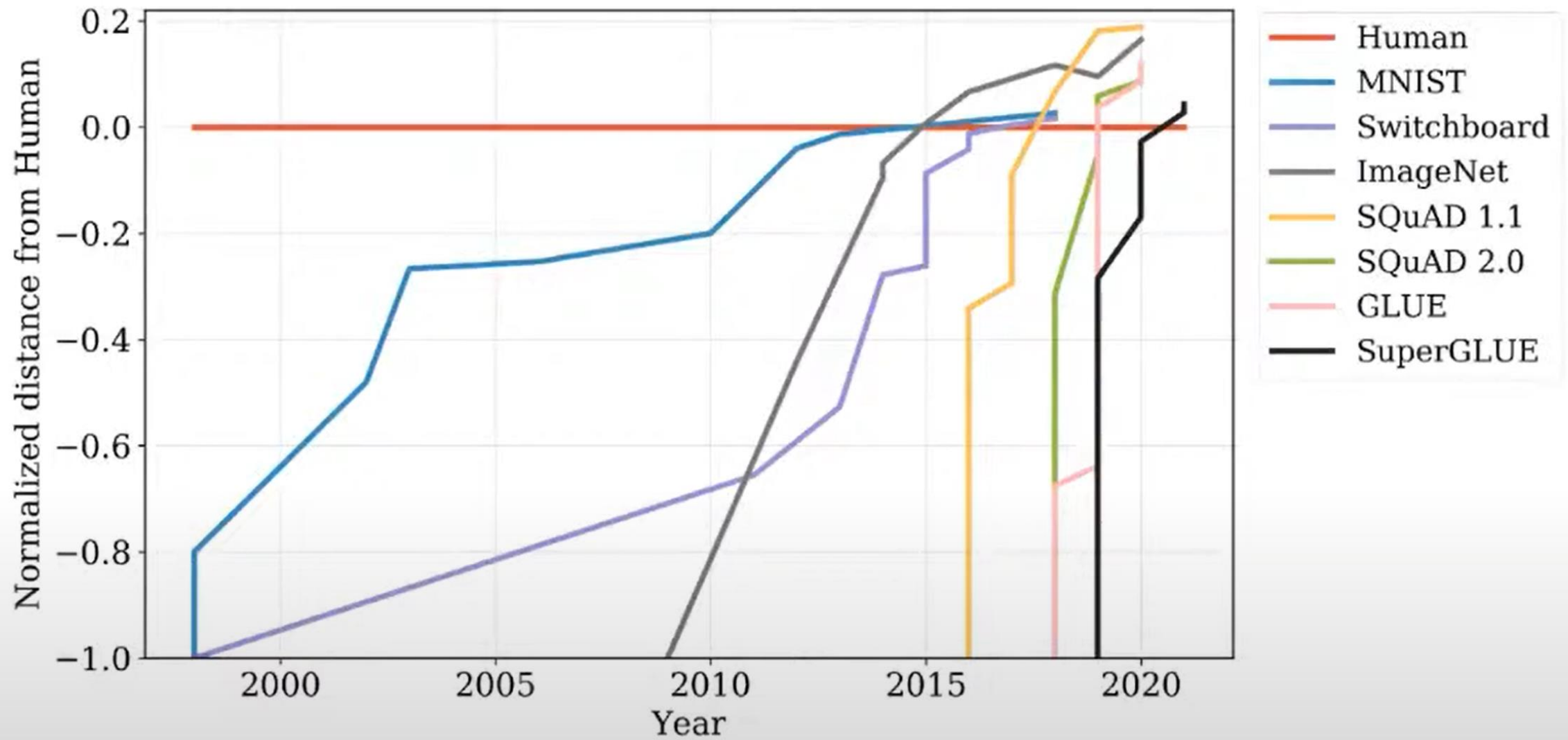


Image courtesy: Chris Potts

# Is NLP a solved problem then?

What are the fundamental challenges of NLP post ChatGPT?



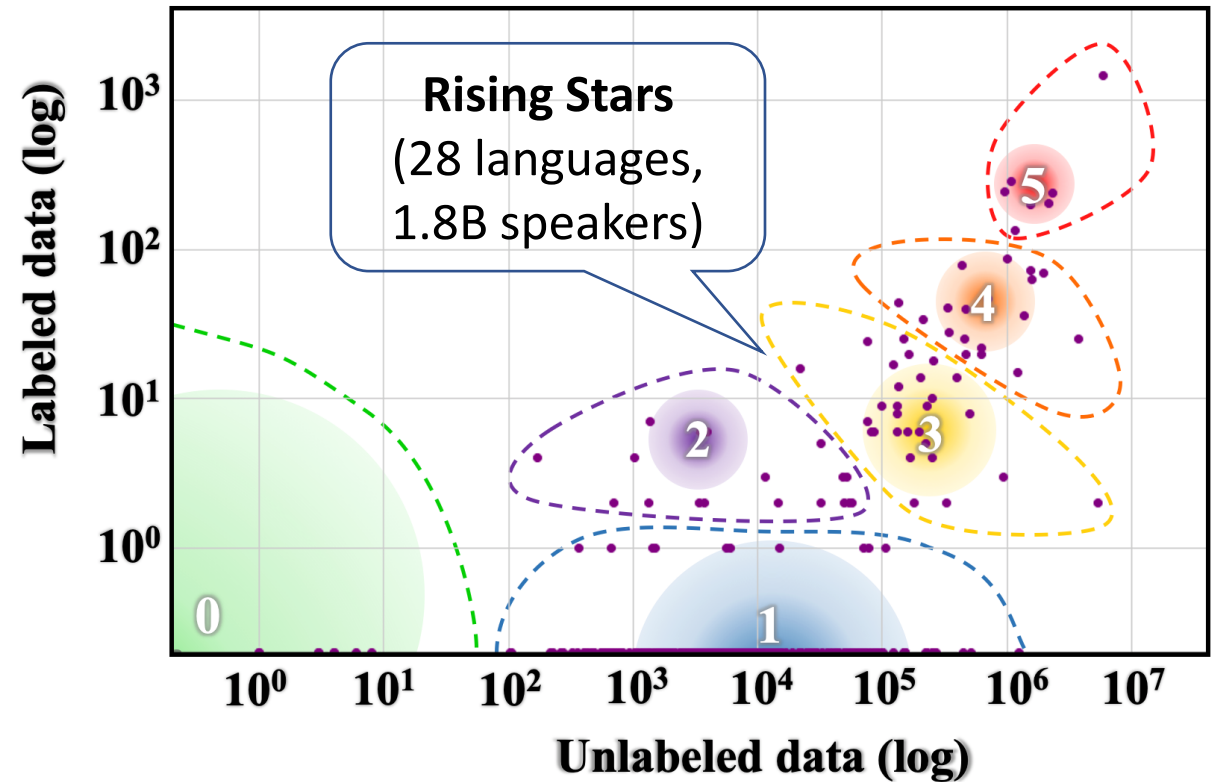
Grounding

Sustainability

Multilingual  
Scaling

MLOPs

# The new class hierarchy of LANGUAGES

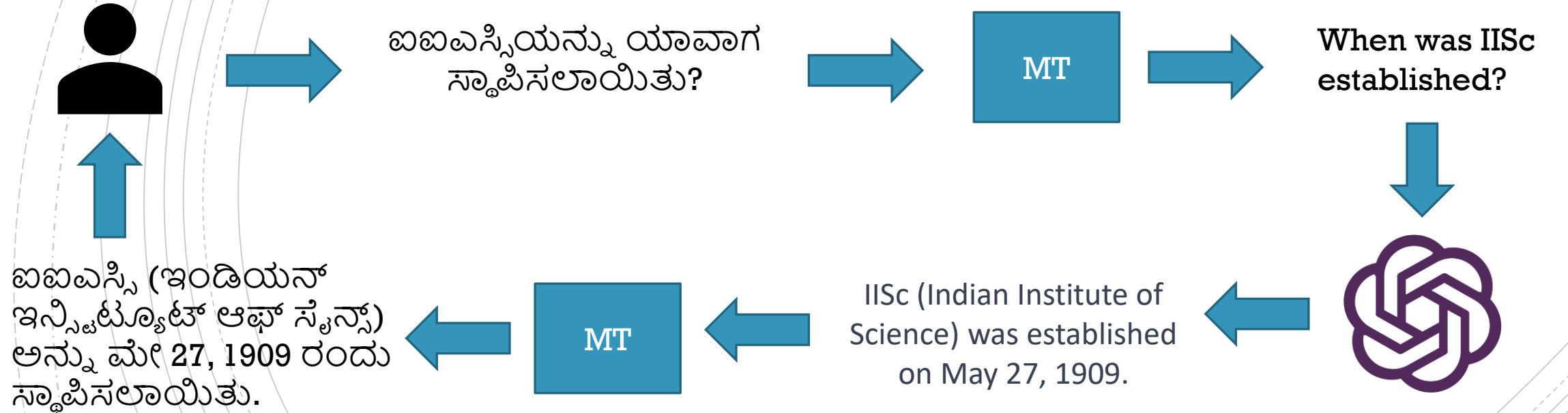


Universal Language Models require massive amounts of unlabeled data for pre-training and some labeled data for testing.

88% of the world's languages, spoken by 1.2B people are untouched by the benefits of language technology.

Class	5 Example Languages	#Langs	#Speakers	% of Total Langs
0	Dahalo, Warlpiri, Popoloca, Wallisian, Bora	2191	1.2B	88.38%
1	Cherokee, Fijian, Greenlandic, Bhojpuri, Navajo	222	30M	5.49%
2	Zulu, Konkani, Lao, Maltese, Irish	19	5.7M	0.36%
3	Indonesian, Ukranian, Cebuano, Afrikaans, Hebrew	28	1.8B	4.42%
4	Russian, Hungarian, Vietnamese, Dutch, Korean	18	2.2B	1.07%
5	English, Spanish, German, Japanese, French	7	2.5B	0.28%

chatbot: ETtech in-depth: Government chatbot  
powered by Bhashini, OpenAI's ChatGPT is at your  
service - The Economic Times (indiatimes.com)





## COURSE FOCUS

# A Very Short Course on Practical NLP in the Age of LLMs

Machine  
Translation

Resource  
constrained  
scenarios

End-to-end  
Application

Hands-on – 60%  
Theory and Broad  
Overview – 40%

# Sessions

## Session 1 (24<sup>th</sup> Feb)

- Course Overview
- History for NLP through MT
- Download required models and platforms
- Understanding the MT pipeline

## Session 2 (3<sup>rd</sup> Mar)

- Neural MT models
- Preprocessing, Training and Inferencing
- Data Pre-processing
- Training an MT model for En-Hindi

## Session 3 (10<sup>th</sup> Mar)

- MT Evaluation
- Techniques for low-resource MT systems
- Techniques for low-resource MT

## Session 4 (17<sup>th</sup> Mar)

- End-to-end Application
- MT market, businesses
- Interfacing MT systems with Chatbots

## Session 5 (24<sup>th</sup> Mar)

- Recent developments (RLHF, FATE issues, probing)
- Buffer time + Project Discussions

# Project

- Choose a language L which is medium to low resource (you can check the status of a language <https://microsoft.github.io/linguisticdiversity/assets/lang2tax.txt>)
  - Build an end-to-end conversation agent using OpenAI GPT for L.
  - Students will use the translation models trained / evaluated by them during the course assignments for the final project to enable query understanding and response generation in the selected language
  - Be creative! Choose an interesting and important application area for your chat-bot
  - The developed bot should be deployable i.e. either as a web-app or an add-on to Slack / Microsoft Teams / Discord.
- 
- Final presentation – 31<sup>st</sup> March
  - Possibility of paper, hosting a service, etc.

# Logistics

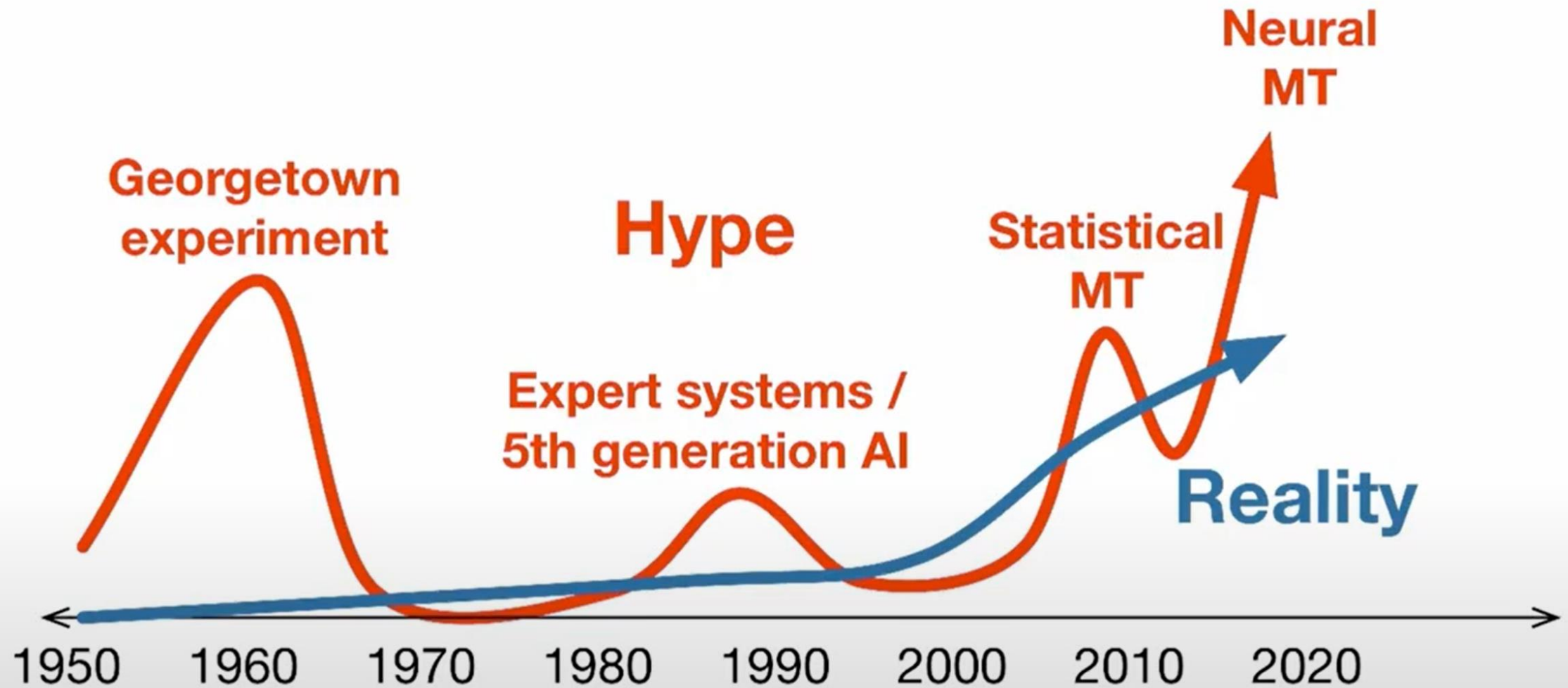
- Every Friday, 3 pm – 6 pm
- Same venue
- Teams of 3 for project and assignments
- Teams for most communications beyond classroom
- Benefits?



LECTURE 1

# A History of NLP through the lens of Machine Translation

24<sup>th</sup> Feb 2023



<https://youtu.be/DuYkqCQEbpo>

# Georgetown-IBM Experiment (1954)

- English to Russian
- It had only six grammar rules and 250 lexical items in its vocabulary (of stems and endings)
- [Georgetown-IBM experiment - Wikipedia](#)

# Georgetown-IBM Experiment (1954)

---

Russian (Romanized)	English translation
Mi pyeryedayem mislyi posryedstvom ryechyi.	We transmit thoughts by means of speech.
Vyelyichyina ugla opryedyelyayetsya otnoshyenyiyem dlyini dugi k radiusu.	Magnitude of angle is determined by the relation of length of arc to radius.
Myezhdunarodnoye ponyimanyiye yavlyayetsya vazhnim faktorom v ryeshyenyiyi polyityichyeskix voprosov.	International understanding constitutes an important factor in decision of political questions.

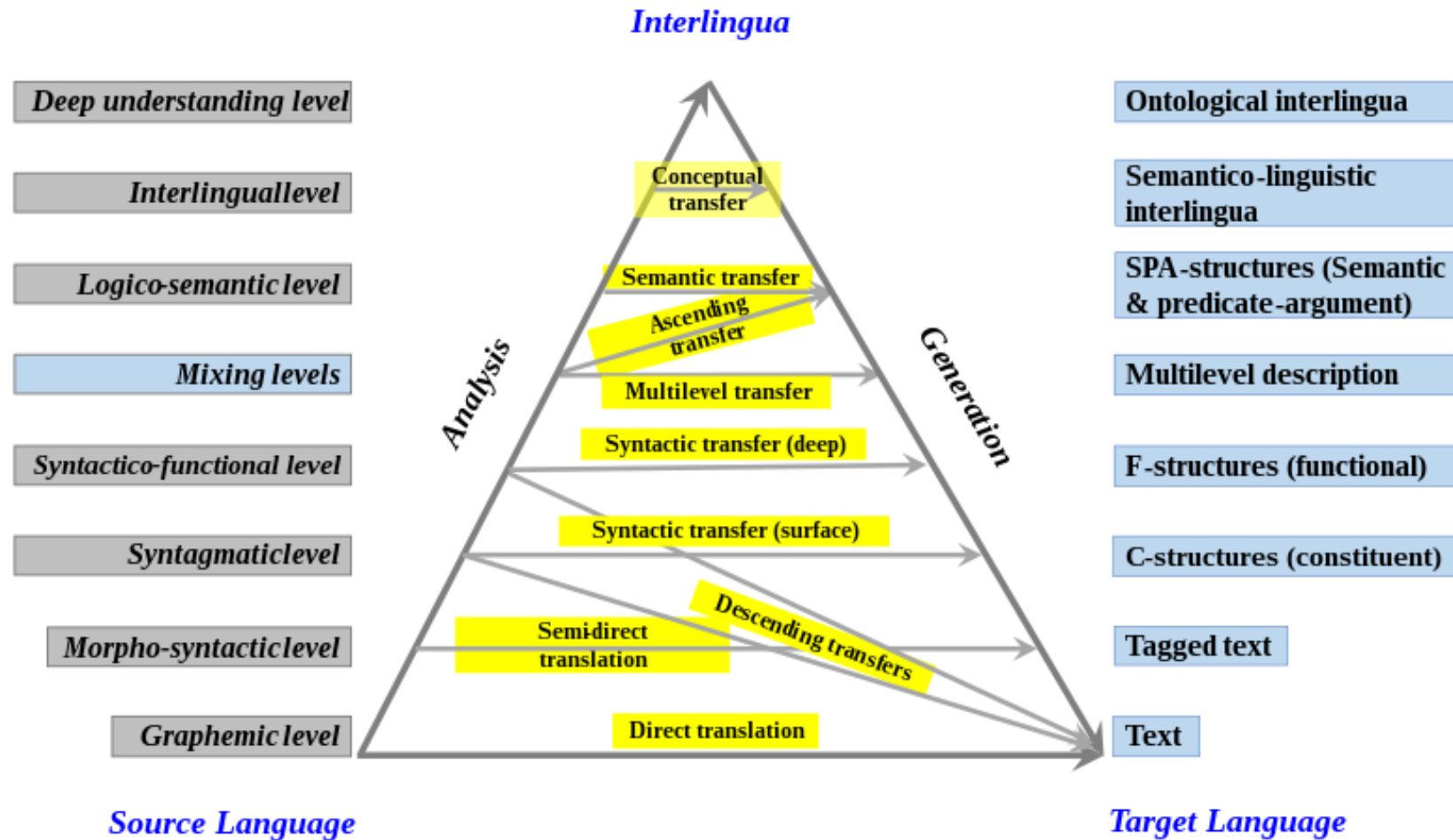
## Reception [\[ edit \]](#)

Well publicized by journalists and perceived as a success, the experiment did encourage governments to invest in [computational linguistics](#). The authors claimed that within three or five years, machine translation could well be a solved problem. However, the real progress was much slower, and after the [ALPAC](#) report in 1966, which found that the ten years of long research had failed to fulfill the expectations, funding was reduced dramatically.

The demonstration was given widespread coverage in the foreign press, but only a small fraction of journalists drew attention to previous machine translation attempts.<sup>[\[7\]](#)</sup>



# Vauquois' Pyramid



# Statistical Machine Translation

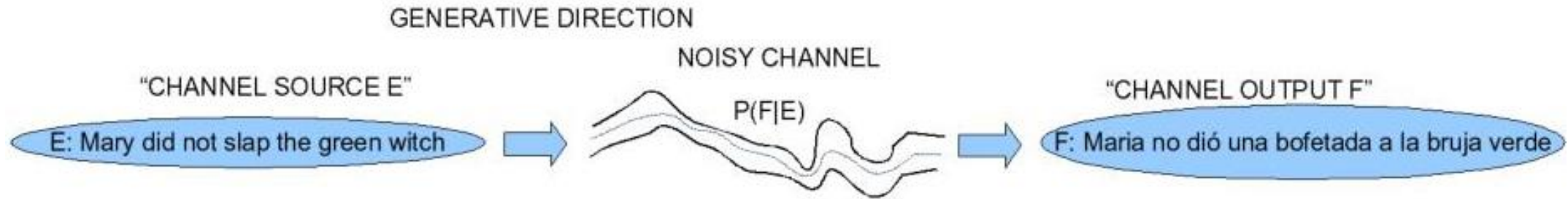
P. Brown; John Cocke; S. Della Pietra; V. Della Pietra; Frederick Jelinek; John D. Lafferty; Robert L. Mercer; P. Roossin (1990). "A statistical approach to machine translation".

*Computational Linguistics*. MIT Press. 16 (2): 79–85

*One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.*

— [Warren Weaver](#), Letter to [Norbert Wiener](#), March 4, 1947

# Noisy channel model



$$\begin{aligned} F^* &= \operatorname{argmax}_F p(F|E) \\ &= \operatorname{argmax}_F p(E|F)p(F)/p(E) \\ &= \operatorname{argmax}_F p(E|F)p(F) \end{aligned}$$

Translation Model

Language Model

# Language Model

- Language models assign probabilities on text sequences:
  - $p(w_1 w_2 w_3 \dots w_n)$
  - $p(w_n \mid w_1 w_2 \dots w_{n-1})$
- Given  $w_1 w_2 \dots w_{n-1}$ , predict the next word(s)

# Statistical LMs

- N-gram models (Markov chains)
- Estimate the probabilities from corpus
- Pros: Easy to implement; compute friendly; transparent
- Cons: Impossible to model long distance (more than 5-10 words) context

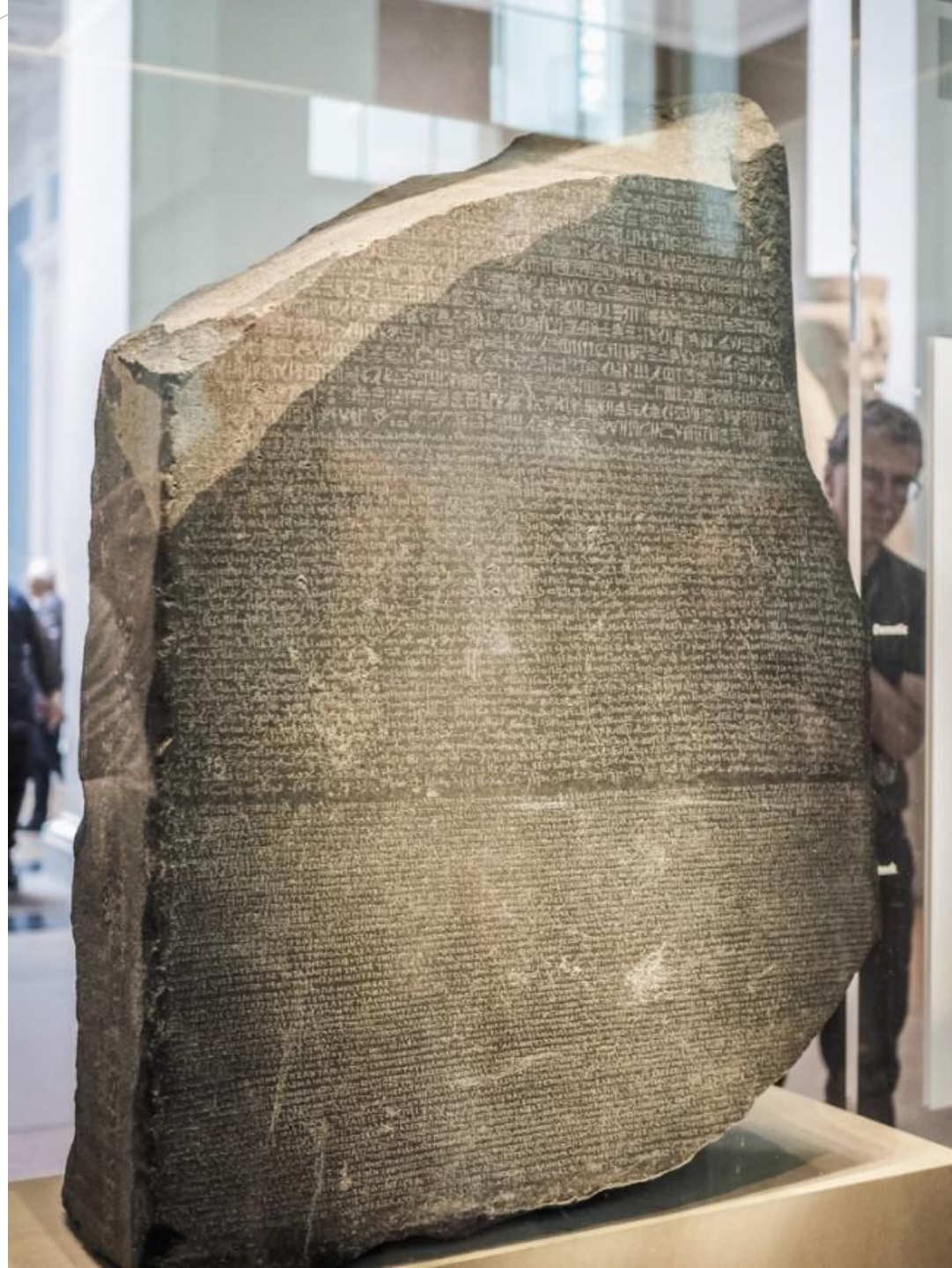
# Translation Model

- Word-based
- Syntax-based
- Phrase-based

# Word Alignment

[illegible]

# Parallel Data





The background is a dark blue-grey color. It features several concentric circles of varying radii, some of which are solid and others dashed. A dashed line forms a large, irregular shape that encompasses the text. A small red dot is located near the bottom left of the text area.

# ▼ Neural Machine Translation

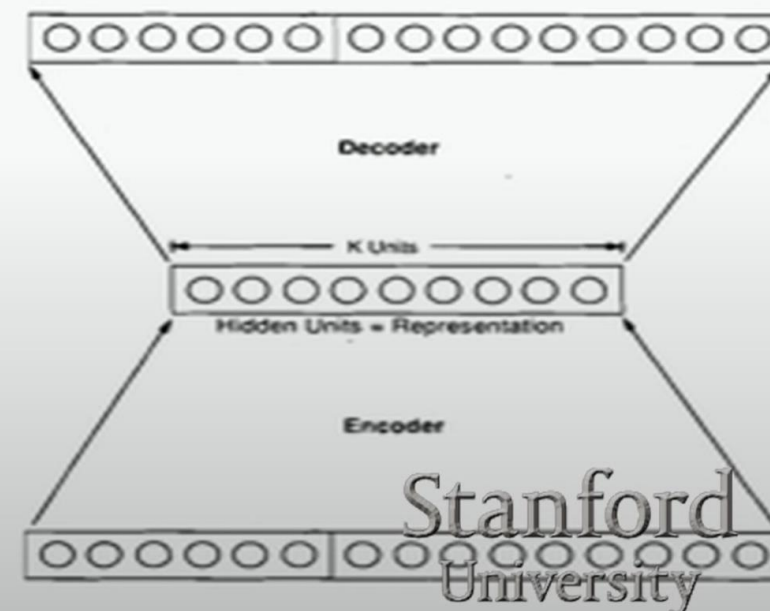
# Neural MT: The Bronze Age

[Allen 1987 IEEE 1<sup>st</sup> ICNN]

3310 En-Es pairs constructed on 31 En, 40 Es words, max 10/11 word sentence; 33 used as test set

The grandfather offered the little girl a book →  
El abuelo le ofrecio un libro a la nina pequena

Binary encoding of words – 50 inputs, 66 outputs; 1 or 3 hidden 150-unit layers. Ave WER: 1.3 words



# Neural MT: The Bronze Age

[Chrisman 1992 *Connection Science*]

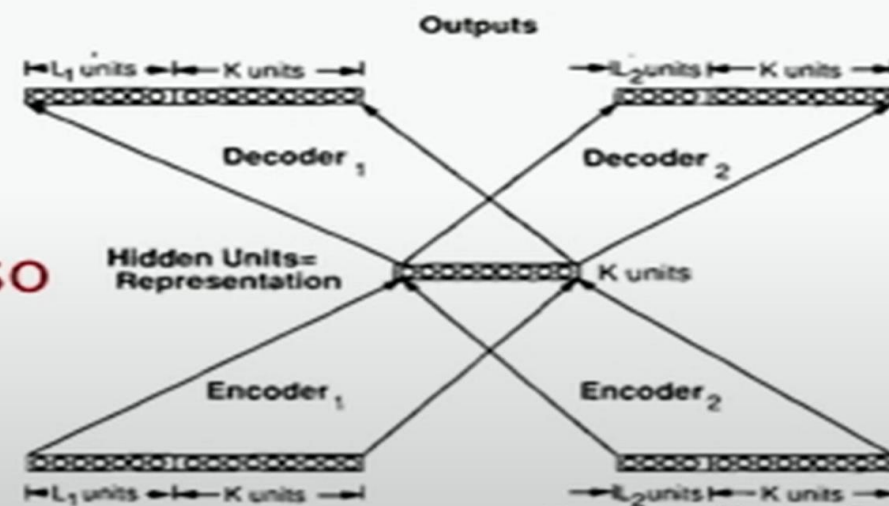
Dual-ported RAAM architecture

[Pollack 1990 *Artificial Intelligence*]

applied to corpus of 216 parallel pairs of simple En-Es sentences:

You are not angry ↔ Usted no esta furioso

Split 50/50 as train/test, 75% of sentences correctly translated!

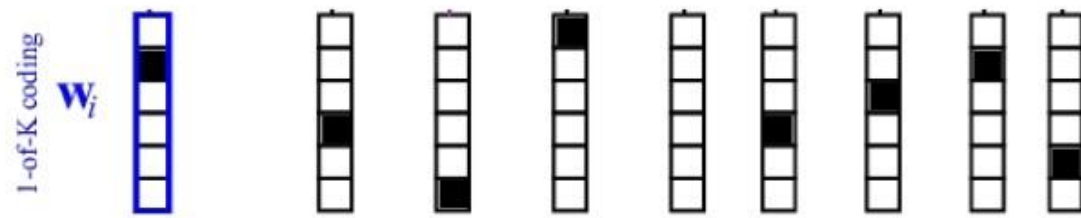


Stanford  
University

# HOW TO INPUT WORDS TO NEURAL NETWORKS?

## Encoder: One-hot encoding

One-hot is a very simple representation: every word is equidistant from every other word.



$e = (\text{Economic, growth, has, slowed, down, in, recent, years, .})$

Kyunghyun Cho, ["Introduction to Neural Machine Translation with GPUs"](#) (2015)

Problems of One-hot encoding:

- Very large input layer (100k+)
- Doesn't exploit semantic and syntactic relation between words

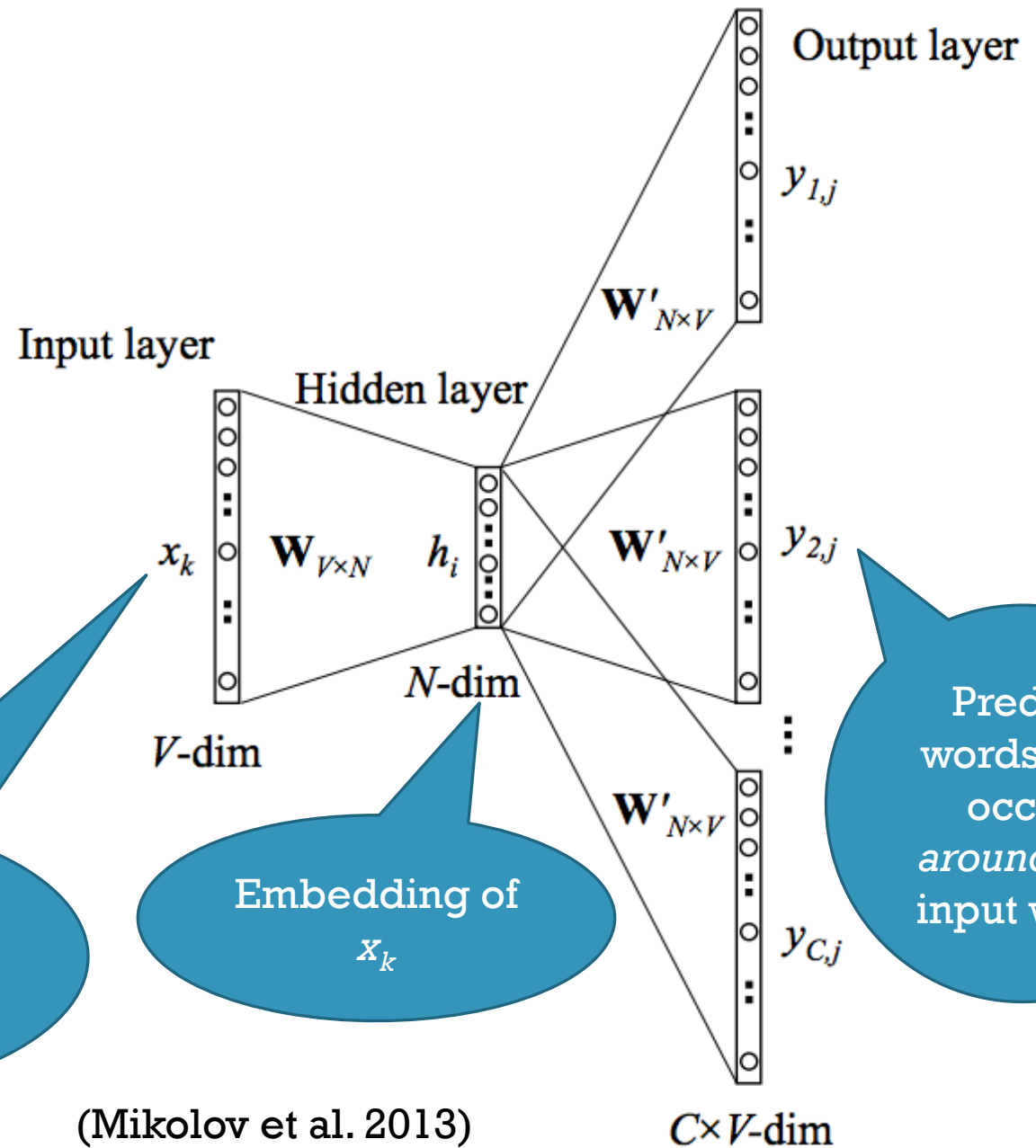
*“You shall know a  
word by the **company**  
it keeps.”*

*- Firth, 1957*

Input word (whose  
embedding is to  
be learnt) as one-  
hot encoding

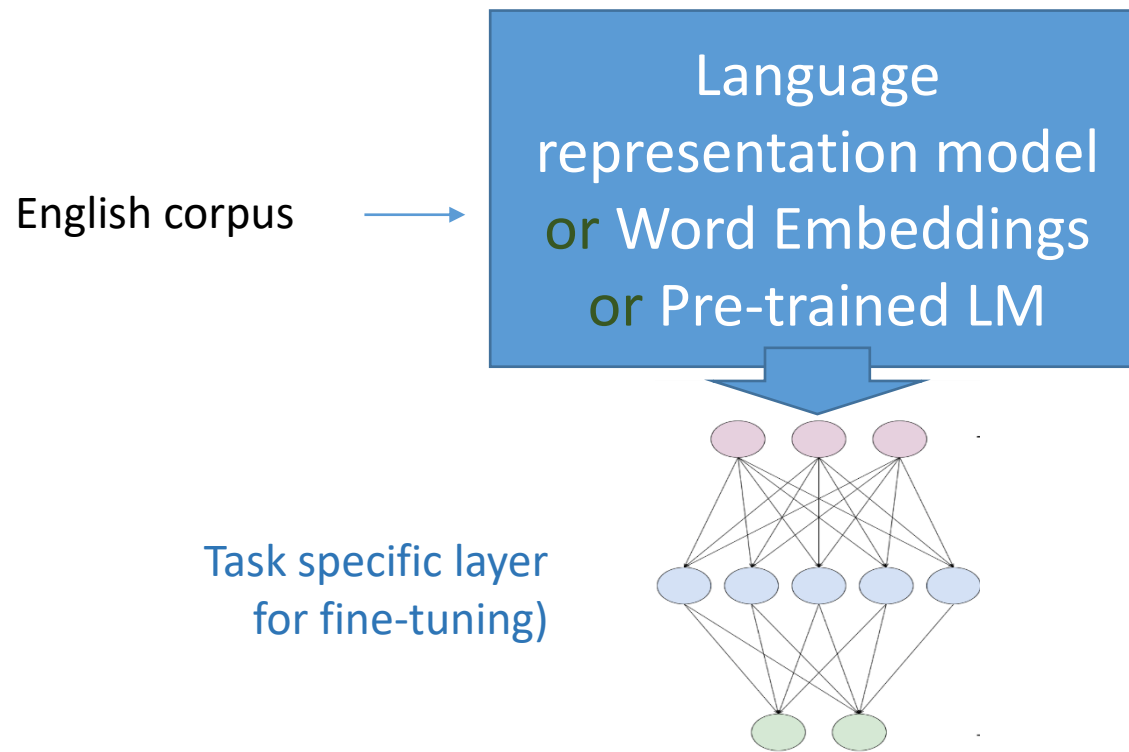
Embedding of  
 $x_k$

Predict  
words that  
occur  
around  
the input word





# The Pretraining-finetuning paradigm of NLP



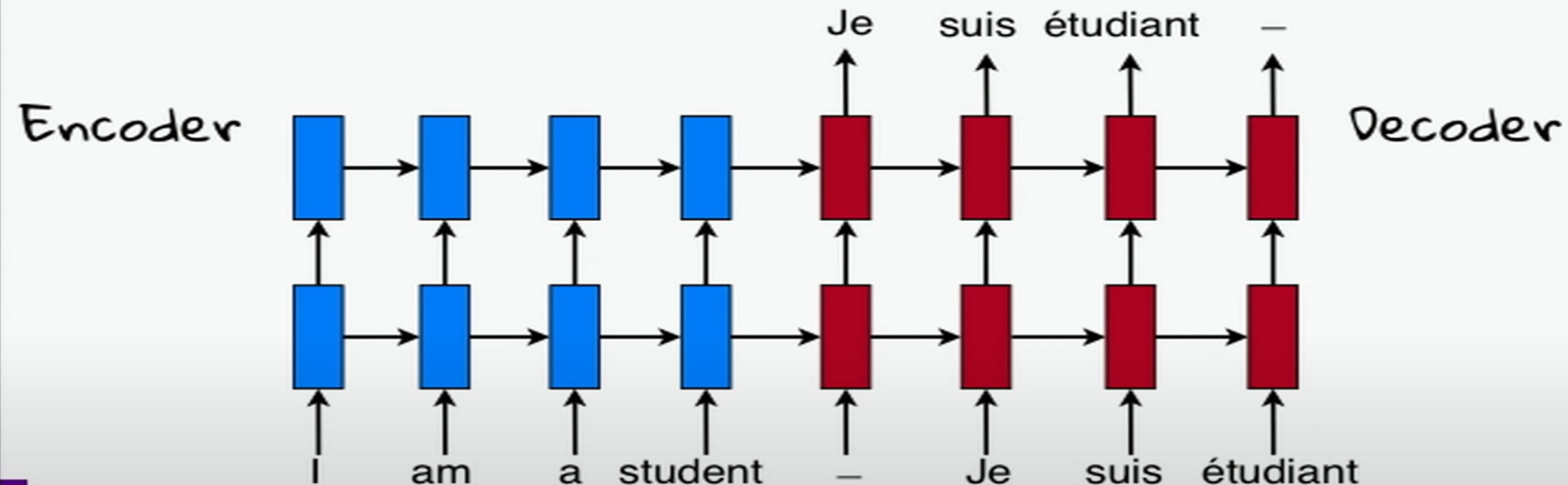
Pre-training: Task independent  
Unsupervised Auto-encoder model

Fine-tuning: Requires task specific  
supervised training data.

Benefit: Much less labeled data  
can give very good performance.

# Modern Sequence Models for NMT

[Sutskever et al. 2014, cf. Bahdanau et al. 2014, et seq.]



Stanford  
University

Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio

## Neural Machine Translation by Jointly Learning to Align and Translate

ICLR 2015

# Attention

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob  
Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and  
Illia Polosukhin. 2017.

## Attention is all you need.

Advances in Neural Information Processing Systems



~~Word~~  
~~Alignment~~  
Attention!

	Tomorrow's	class	will	be	postponed	to	next	Week	.
कल	Dark Blue	Light Blue							
की	Dark Blue	Dark Blue							
कक्षा		Dark Blue	Light Blue	Light Blue					
अगले						Light Blue	Dark Blue	Light Blue	
सप्ताह						Light Blue	Light Blue	Dark Blue	
के						Dark Blue	Light Blue	Light Blue	
लिए						Dark Blue	Light Blue	Light Blue	
स्थगित			Light Blue	Light Blue	Dark Blue				
कर					Dark Blue				
दी			Light Blue	Light Blue	Light Blue				
जाएगी			Dark Blue	Dark Blue	Light Blue				
।		Light Blue	Light Blue	Light Blue	Light Blue				Dark Blue



# AI can translate as well as a human

## Microsoft System Achieves Same Quality And Accuracy As Us

**Washington:** In a first, Microsoft researchers have developed an artificial intelligence (AI) system that can translate news from Chinese to English with the same quality and accuracy as a human. Researchers said their system achieved human parity on a commonly used test set of news stories, which was developed by a group of industry and academic partners.

To ensure the results were both accurate and on par with what people would have done, the team hired external bilingual human evaluators, who compared Microsoft's results to two independently produced human reference translations.

Xuedong Huang, a technical

Thinkstock\ Photo for representation



**AS GOOD AS US**

fellow in charge of Microsoft's speech, natural language and machine translation efforts, called it a major milestone in one of the most challenging natural language processing tasks.

"Hitting human parity in a

machine translation task is a dream that all of us have had. We just didn't realise we'd be able to hit it so soon," Huang said.

The translation milestone was especially gratifying because of the possibilities it has for helping people understand each other better, he said.

Arul Menezes, partner research manager of Microsoft's machine translation team, said they set out to prove that its systems could perform about as well as a person when it is used a language pair — Chinese and English — for which there is a lot of data, on a test set that includes more ordinary vocabulary of general interest news stories.

Researchers added other

training methods to make the system more fluent and accurate. These methods mimic how people improve their own work iteratively, by going over it again and again until they get it right.

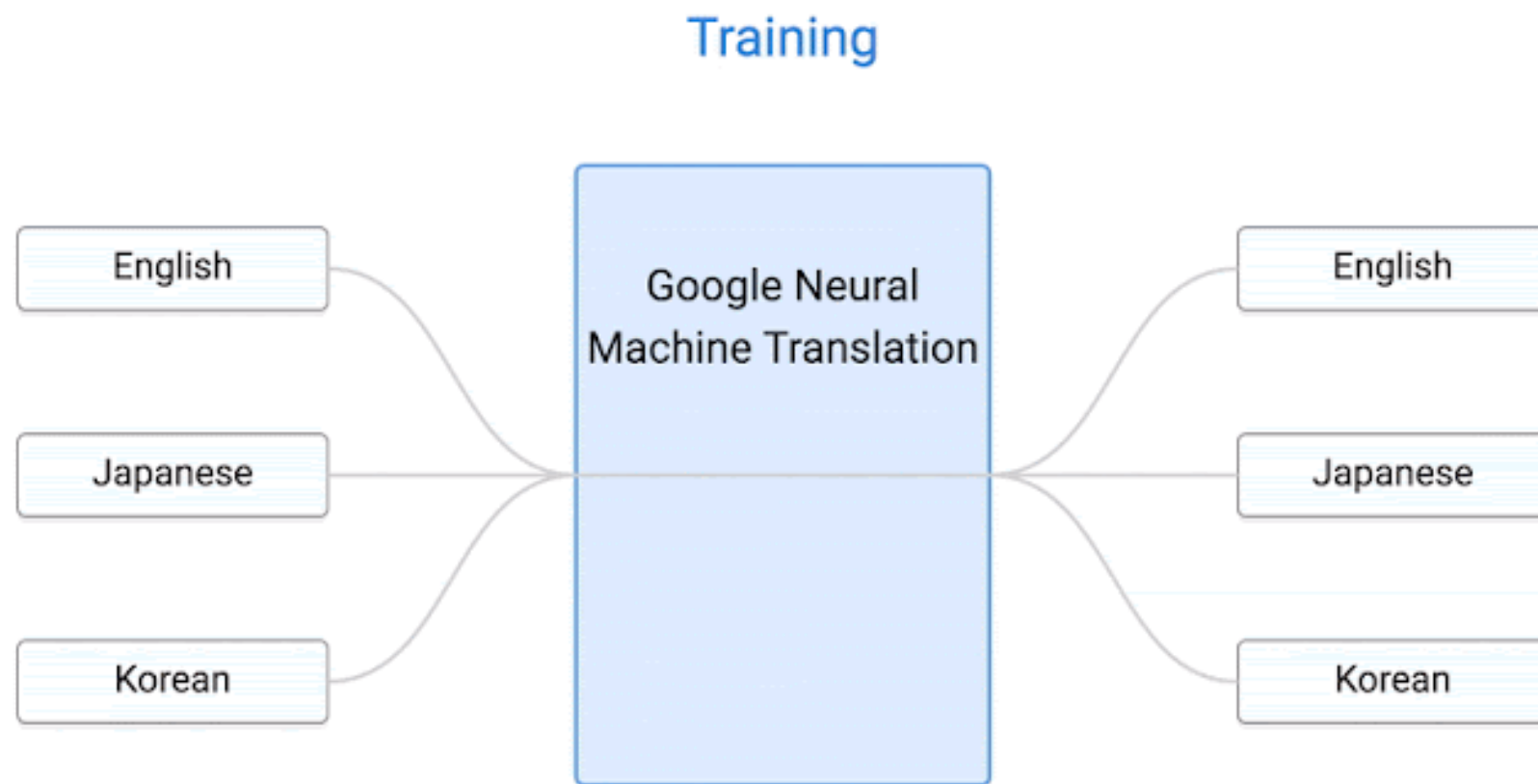
"Much of our research is inspired by how we humans do things," said Tie-Yan Liu, a principal research manager with Microsoft Research Asia in Beijing.

The researchers also developed two new techniques to improve the accuracy of their translations, Zhou said.

These techniques could be useful for improving machine translation in other languages as well. He said they also could be used to make other AI breakthroughs beyond translation. **PT**

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean. (2016)

## Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation



# TIMELINE

W2V, RNN, LSTM,

...

(Soft-)alignment  
(Bahadanu et al, 2014)

Transformer  
(Vaswani et al.,  
2017)

ELMO (Peter et  
al., 2018)

ULMFit (Ruder et al, 2018)

**GPT1**

June, 2018

**BERT**

Oct, 2018

**GPT2**

Feb, 2019

**XLNET**

June, 2019

**T5**

Oct, 2019

**BART**

Oct, 2019

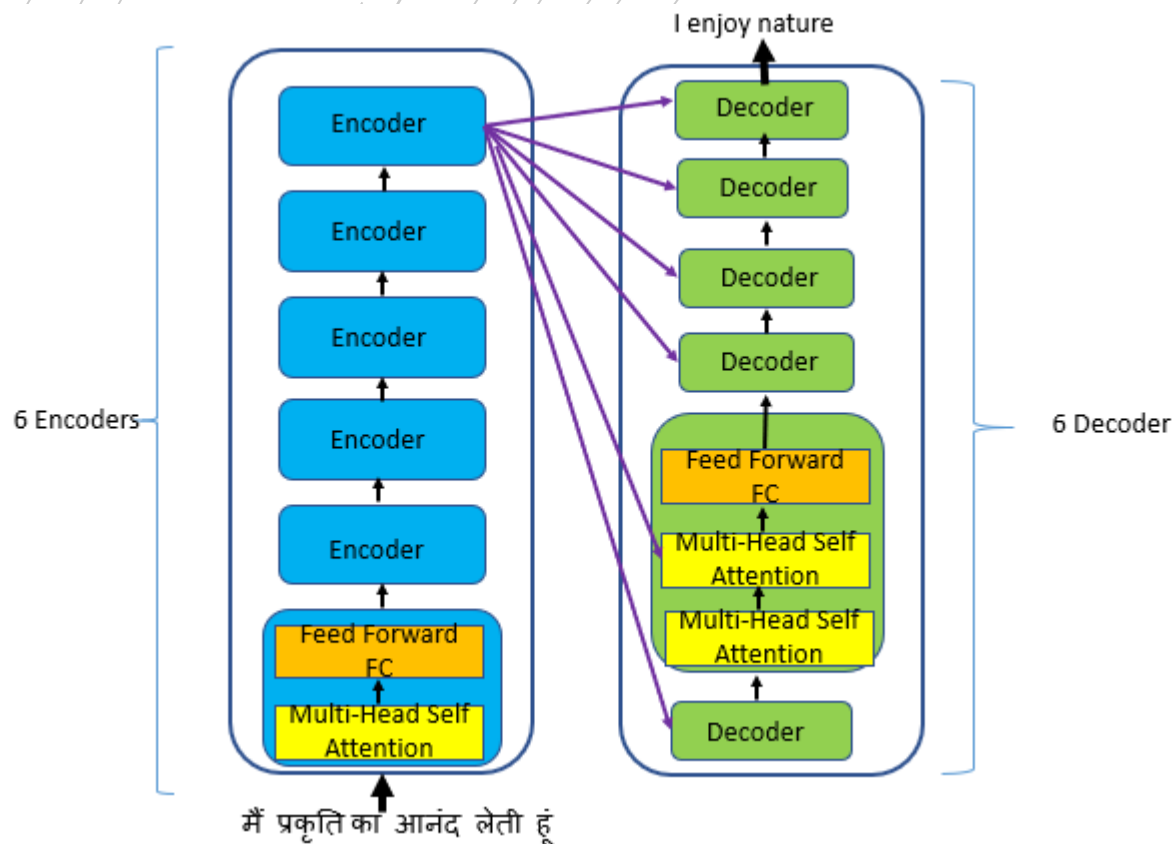
**ProphetNet**

Jan, 2020

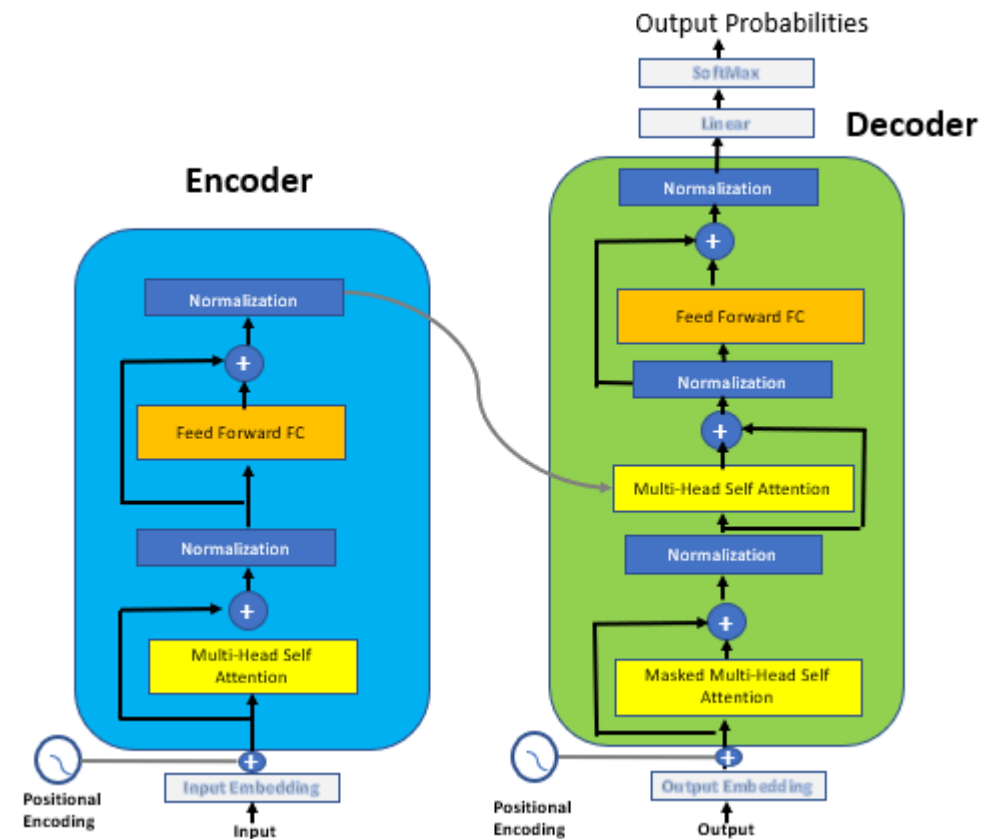
**GPT3**

May, 2020





Encoder-Decoder Architecture



A single Encoder-Decoder

# Commercial MT Systems

- Bing Translator: 111 languages (12 Indic languages)
- Google Translator: 133 languages (20 Indic languages)

Other: IBM Watson, Yandex, etc.

Source: Wikipedia (24/2/2023)

## Questions to Ponder

- What makes machine translation such a difficult problem?
- Are all language pairs equally difficult to translate? Why or why not?
- How do humans translate between languages (ie. what are the cognitive steps involved)?