Machine Translation and Quality Estimation

Dr Tharindu Ranasinghe
Damith Premasiri

Lancaster University

Glimpse of ML/NLP (All about ChatGPT)

ChatGPT better than undergraduates at solving SAT problems, study suggests

Researchers at UCLA found GPT-3 solved 80% of reasoning problems correctly compared with 60% of humans

Dan Milmo *Global technology editor*

NEWS FEATURE | 25 July 2023

ChatGPT broke the Turing test — the race is on for new ways to assess AI

Large language models mimic human chatter, but scientists disagree on their ability to reason.

ChatGPT passes exams from law and business schools

By Samantha Murphy Kelly, CNN Business
Updated 1:35 PM EST, Thu January 26, 2023

NEXT GEN INVESTING

Bill Gates watched ChatGPT ace an AP Bio exam and went into 'a state of shock': 'Let's see where we can put it to good use'

Published Fri, Aug 11 2023·3:02 PM EDT

ChatGPT Passes US Medical Licensing Exam Without Clinician Input

ChatGPT achieved 60 percent accuracy on the US Medical Licensing Exam, indicating its potential in advancing artificial intelligence-assisted medical education.

ChatGPT Passes CPA Exam on Second Attempt. Is It Coming For Accounting Jobs?

The latest version of the chatbot passed the CPA exam with a score of 85.1, just weeks after an earlier version had failed

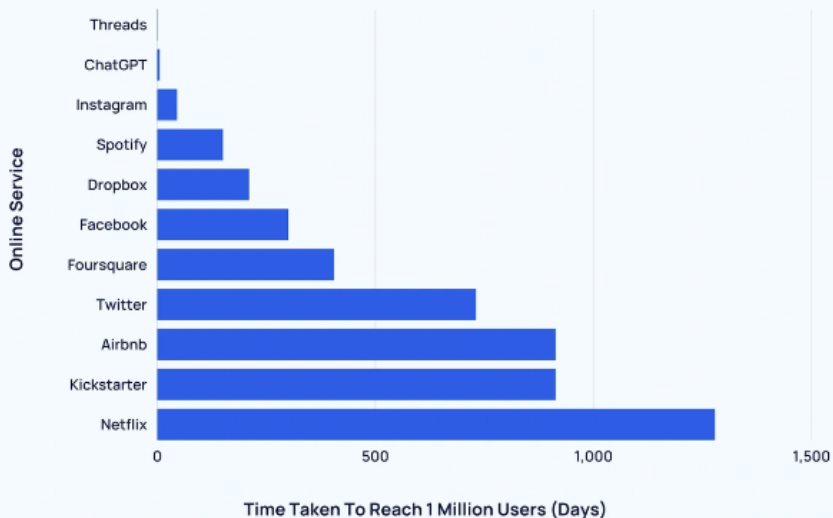# Glimpse of ML/NLP (All about ChatGPT)

ChatGPT has over 100 million users

- ChatGPT currently has over 100 million users. And the website generated 1.6 billion visits in June 2023.

- This user and traffic growth was achieved in a record-breaking three-month period (from April 2023 to June 2023).

Glimpse of ML/NLP (All about ChatGPT)

Time taken to reach 1 million users

- ChatGPT acquired 1 million users just 5 days after launching in November 2022.

- It took Instagram approximately 2.5 months to reach 1 million downloads. Whereas Netflix had to wait around 3.5 years to reach 1 million users.

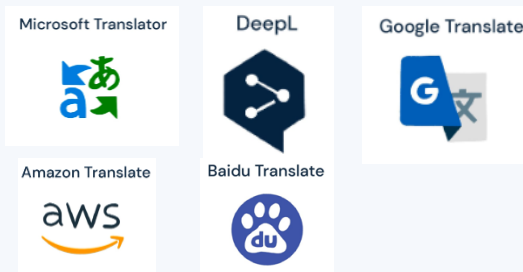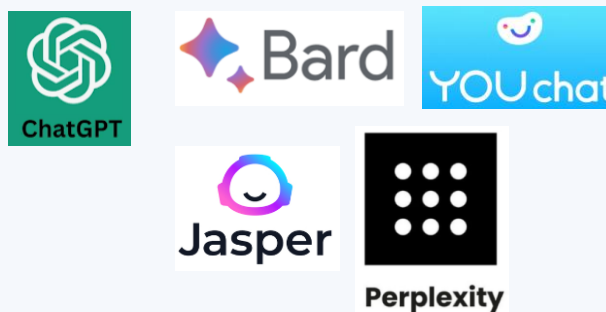# Glimpse of ML/NLP (Not only about ChatGPT)

Language Tools

Machine Translation

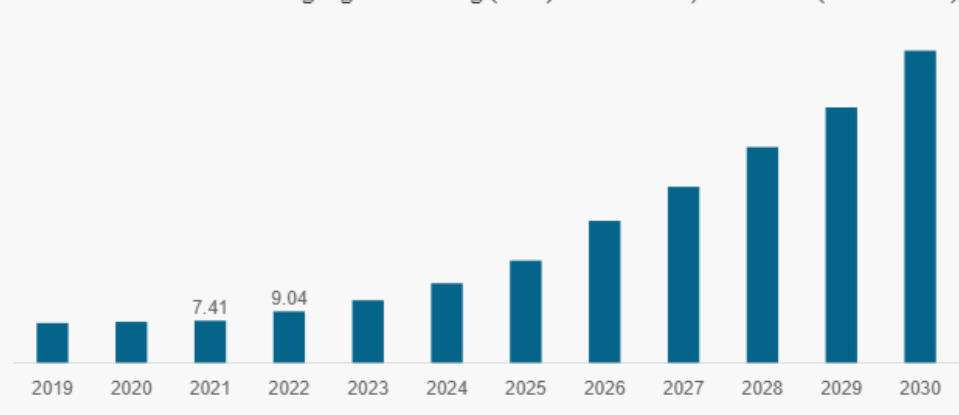
Content Moderation

Chatbots

Glimpse of ML/NLP (Not only about ChatGPT)

North America Natural Language Processing (NLP) Market Size, 2018-2029 (USD Billion)

- The NLP market was worth $15.7 billion in 2022. It is expected to grow to $49.4 billion by 2027, at a compound annual growth rate (CAGR) of 25.7%.

Glimpse of ML/NLP (Not only about success)

Bloomberg

Technology
ChatGPT Fails at Recommending Cancer Treatment, Study Finds

WORLD

Not smarter than a 6th grader: ChatGPT fails Singapore's 6th-grade maths and science exams

For a chatbot that cleared Wharton's MBA, four law school courses, and a US medical licensure exam with ease, it is a bit surprising that it failed primary-level school exams. Seems that OpenAI's ChatGPT isn't coming for our jobs any time soon.

INDIA TODAY

News / Technology / News / ChatGPT fails JEE, manages to solve only 11 questions in both papers

ChatGPT fails JEE, manages to solve only 11 questions in both papers

The AI-powered ChatGPT faced a setback when it failed to clear India's prestigious Joint Entrance Examination (JEE) Advanced.

NEWS / EDUCATION NEWS / NEWS / AI Chatbot 'ChatGPT' Fails To Clear UPSC CSE Preli...

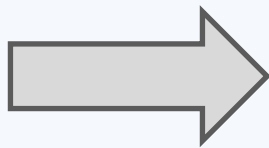TRENDING | CTET Admit Card | KCET Seat Allotment | SSC CAPF Final Result | NVS Class 6 Re

🕐 THIS STORY IS FROM MARCH 6, 2023

AI chatbot 'ChatGPT' fails to clear UPSC CSE Prelims 2022 by 30%

Better models to solve everything

Creating better models

Déjà vu

Déjà vu – 8 years ago…

Microsoft's Chinese-English translation system achieves human parity

A China-US research team at software powerhouse Microsoft has developed an AI-powered system that can translate Chinese news articles into English as well as humans do.

By Aaron Tan, TechTarget

Published: 15 Mar 2018 3:30

slator
Language Industry Intelligence

Subscribe

Sign In

Technology • By Florian Faes On 27 Sep 2016

"Nearly Indistinguishable From Human Translation"— Google Claims Breakthrough

Fast forward 8 years

How to Avoid Embarrassing Translation Mistakes and Costly Errors
By Victoria Nicol | February 15, 2023 | No Comments

text&form | iworks
When machine translation mistakes are no laughing matter

future tense
Death by Machine Translation?
BY SOFIA QUAGLIA
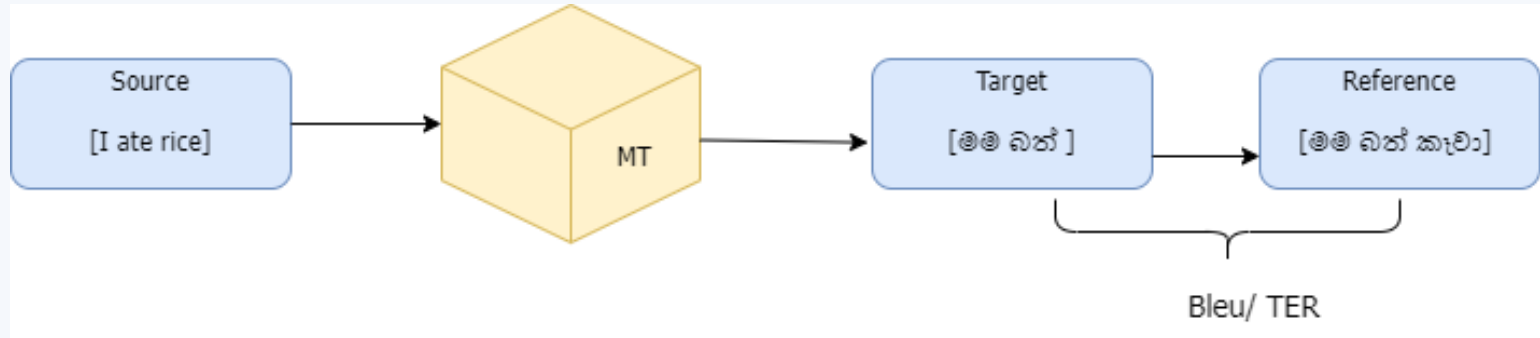SEPT 21, 2022 • 5:50

National | Social Affairs
Hidden meaning of Korean term "agassi" leads to murder
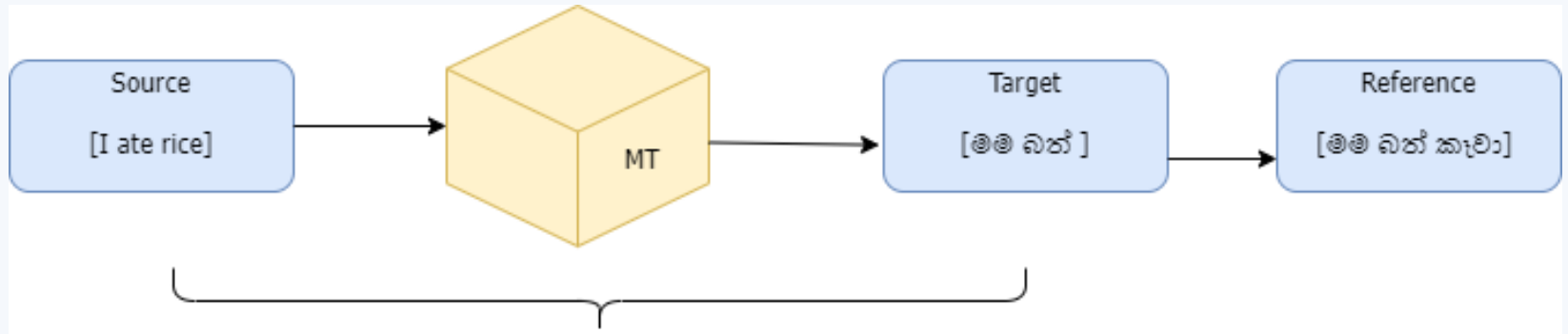From young woman to bar hostess, a term causes a fatal misunderstanding between a Chinese and Korean man

Facebook translates 'good morning' into 'attack them', leading to arrest
Palestinian man questioned by Israeli police after embarrassing mistranslation of caption under photo of him leaning against bulldozer

Machine Translation Evaluation

Machine Translation Quality Estimation

Machine Translation Quality Estimation

QuEst (Specia et. al., 2013) - Utilised feature extraction, and supervised machine learning.

QuEst++ (Specia et. al., 2015) - Features change at word-level, sentence-level or document-level QE.

OpenKiwi (Keplar et. al., 2019) - Predictor-Estimator architecture where the predictor uses a bidirectional LSTM to encode the source while the estimator takes features from predictor and classifies them.

TransQuest (Ranasinghe et. al., 2020) - QE with the help of multilingual language models like XLM-R.
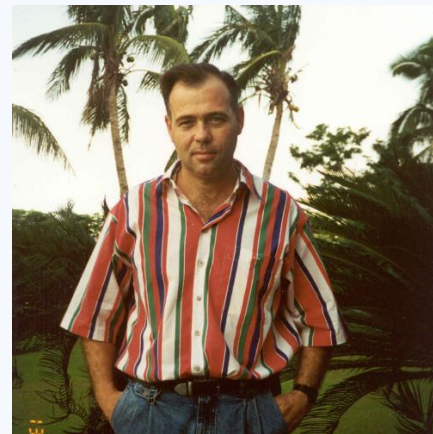
TransQuest

Tharindu Ranasinghe

Constantin Orăsan

Ruslan Mitkov

QE Sentence Level

1. Post-editing Effort

2. Direct Assessment

3. MQM (Multidimensional Quality Metrics)

QE Sentence Level

	Language Pairs	Sentences Train / Dev / Test23	Tokens Train / Dev / Test23	DA	PE	MQM	Data Source	Release
DA & post edits	En-De	10,000 / − / −	148,044 / − / −	✓	✓		Wikipedia	2021/22
	En-Zh	10,000 / − / −	148,529 / − / −	✓	✓		Wikipedia	2021/22
	Ru-En	10,000 / − / −	105,871 / − / −	✓	✓		Reddit	2021/22
	Ro-En	10,000 / − / −	154,825 / − / −	✓	✓		Wikipedia	2021/22
	Et-En	10,000 / − / −	126,547 / − / −	✓	✓		Wikipedia	2021/22
	Ne-En	10,000 / − / −	135,095 / − / −	✓	✓		Wikipedia	2021/22
	Si-En	10,000 / − / −	140,932 / − / −	✓	✓		Wikipedia	2021/22
	Ps-En	2,000 / − / −	54,459 / − / −	✓	✓		Wikipedia	2021/22
	Km-En	2,000 / − / −	44,029 / − / −	✓	✓		Wikipedia	2021/22
	En-Ja	2,000 / − / −	41,272 / − / −	✓	✓		Wikipedia	2021/22
	En-Cs	2,000 / − / −	40,638 / − / −	✓	✓		Wikipedia	2021/22
	En-Yo	1,010 / − / −	21,238 / − / −	✓	✓			2021/22
	En-Mr	27,000 / 1,000 / 1,086	717,581 / 26,253 / 27,951	✓	✓		multi-domain/multi-corpus	2022/23
	En-Hi	7,000 / 1,000 / 1,074	181,336 / 25,943 / 28,032	✓			multi-domain/multi-corpus	2023
	En-Gu	7,000 / 1,000 / 1,075	153,685/ 21,238 / 23,084	✓			multi-domain/multi-corpus	2023
	En-Ta	7,000 / 1,000 / 1,067	150,670 / 21,655/ 20,342	✓			multi-domain/multi-corpus	2023
	En-Te	7,000 / 1,028 / 1,000	147,492 / 20,686 / 22,640	✓			multi-domain/multi-corpus	2023
	En-Fa	− / − / 1,000	−/ − / 26,807		✓		news (multi-domain)	2023
MQM	En-De	30,425 / − / 1,897	877,066 / − / 37,996			✓	multi-domain	2021/23
	En-Ru	17,144 / − / −	395,045 / − / −			✓	multi-domain	2021/22
	Zh-En	36,851 / − / 1,675	1,654,454 / − / 39,770			✓	multi-domain	2021/23
	He-En	− / − / 1,182	− / − / 35,592			✓	multi-domain	2023

Table 1: Statistics of the data used for Task 1 and Task 2. The number of tokens is computed based on the source sentences. Hallucinated data included in the calculations for the 2023 testsets.

QE Word Level

Source - *for example , you could create a document containing a car that moves across the Stage .*

Target - *Sie können beispielsweise ein Dokument erstellen , das ein Auto über die Bühne enthält .*

Source - *for example , you could create a document containing a car that moves across the Stage .*

Target - *<GAP> Sie <GAP> können <GAP> beispielsweise <GAP> ein <GAP> Dokument <GAP> erstellen <GAP> , <GAP> das <GAP> ein <GAP> Auto <GAP> über <GAP> die <GAP> Bühne <GAP> enthält <GAP> . <GAP>*
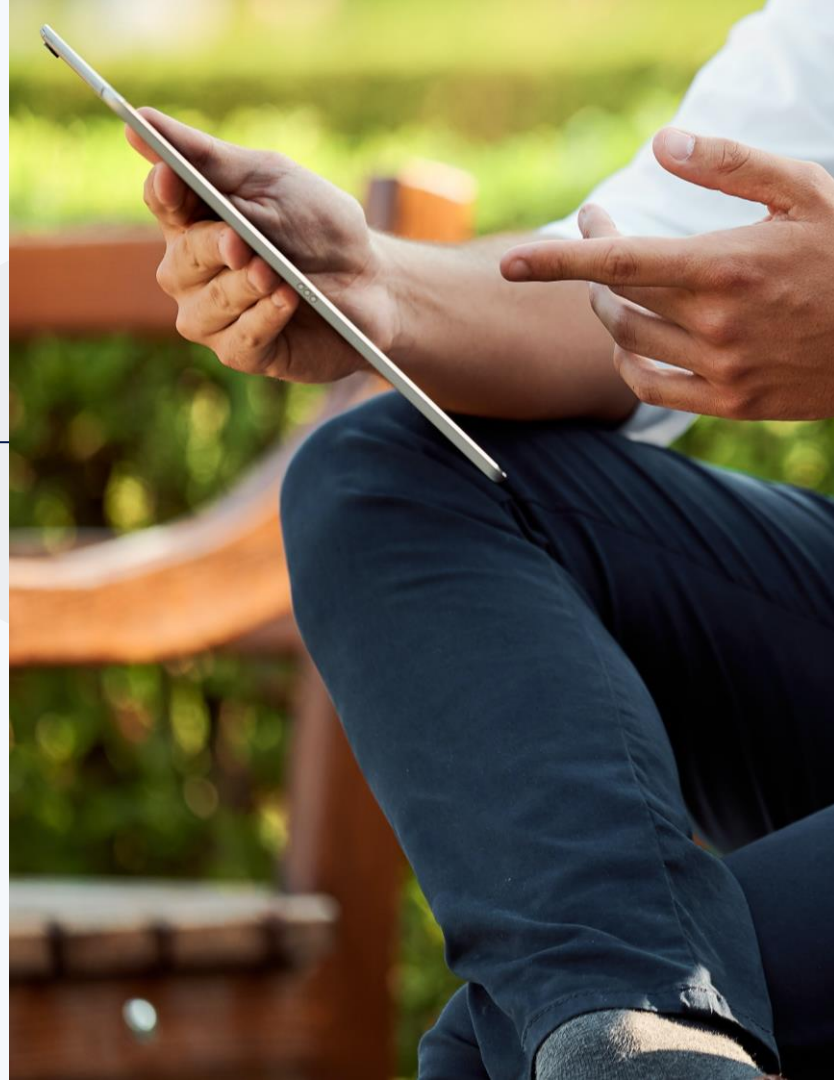
Responsible ML/NLP

Responsible ML/NLP

Recent Trends

Recent Trends in QE

1. Creating resources for low-resource languages/ domains

2. Multilingual QE models

3. Few-shot/ Zero-shot QE models

4. LLMs and machine translation

Thank you!

- Checkout the WMT QE Shared Task 2024
- Building responsible MT is a responsibility of us

Do you have any questions?

t.ranasinghe@lancaster.ac.uk | tharindu.co.uk