

World Café GPT

Sanjay Pant, Sébastien Bratières

17/05/2023

The role of Large Language Models (LLMs) in EDITH

1. **VHT repository:** large body of scientific, legal, regulatory, and other relevant documents.
2. **Large Language models** can be used to:
 - a) Summarise the literature — individually and collectively (by themes)
 - b) Answer any question from the literature

EDITH QnA web-application

<http://slyn.eu.pythonanywhere.com/>

Login: edith
password: Edith2023!



Ecosystem for Digital
Twins in Healthcare

EDITH

WHO cautions on use of ChatGPT, Bard in healthcare; proposes to address concerns before use

WHO proposes that these concerns be addressed, and clear evidence of benefit be measured before their widespread use in routine health care and medicine

By **PGurus Newsdesk** - May 17, 2023



WHO proposes that these concerns be addressed, and clear evidence of benefit be measured before their widespread use in routine health care and medicine

Is Google's Bard HIPAA compliant?

[Dean Levitt](#)

May 16, 2023



When it comes to healthcare, privacy, and security are paramount. The Health Insurance Portability and Accountability Act (HIPAA) sets the standard for sensitive

Prompt name	Prompt	Target
a_good_translation	Given the following source text (in L1): [source sentence], a good L2 translation is:	[target sentence]
version	If the original version says [source sentence] then the L2 version should say:	[target sentence]
gpt3	What is the L2 translation of the sentence: [source sentence]?	[target sentence]
xglm	(L1:) [source sentence] = L2:	[target sentence]
translate_as	[source sentence] translates into L2 as:	[target sentence]

Table 1: MT prompts for the WMT’14 dataset (Bojar et al., 2014). All prompts specify the target language (L2). Each prompt exists in a ‘target-only’ version (-target), where only the target language is specified, and two prompts also exist in a -source+target version, where the source language (in red and in brackets) is explicit in the instruction.

Bawden, Rachel and Francois Yvon. “Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM.” *ArXiv abs/2303.01911* (2023)



In-context Template	Deu-Eng	Eng-Deu	Rus-Eng	Eng-Rus	Rus-Deu	Deu-Rus	Average
reasonable templates:							
$\langle X \rangle = \langle Y \rangle$	37.37	26.49	29.66	22.25	17.66	17.31	25.12
$\langle X \rangle \setminus \text{n Translate from [SRC] to [TGT]: } \setminus \text{n } \langle Y \rangle$	37.95	26.29	29.83	20.61	17.56	15.93	24.70
$\langle X \rangle \setminus \text{n Translate to [TGT]: } \setminus \text{n } \langle Y \rangle$	37.69	25.84	29.96	19.61	17.44	16.48	24.50
$\langle X \rangle \setminus \text{n [TGT]: } \langle Y \rangle$	29.94	17.99	25.22	16.29	12.28	11.71	18.91
$\langle X \rangle$ is equivalent to $\langle Y \rangle$	23.00	4.21	17.76	9.44	8.14	9.84	12.07
$\langle X \rangle \setminus \text{n can be translated to } \setminus \text{n } \langle Y \rangle$	37.55	26.49	29.82	22.14	17.48	16.40	24.98
[SRC]: $\langle X \rangle \setminus \text{n [TGT]: } \langle Y \rangle$	16.95	8.90	14.48	6.88	7.86	4.01	9.85
unreasonable templates:							
$\langle X \rangle \% \langle Y \rangle$	37.77	26.43	29.53	20.99	17.72	17.27	24.95
$\langle X \rangle \setminus \text{n Translate from [TGT] to [SRC]: } \setminus \text{n } \langle Y \rangle$	38.18	26.21	29.85	20.35	17.75	16.63	24.83
$\langle X \rangle \setminus \text{n Compile to [TGT]: } \setminus \text{n } \langle Y \rangle$	37.39	26.35	29.68	19.91	17.52	16.15	24.50
$\langle X \rangle \setminus \text{n [SRC]: } \langle Y \rangle$	27.86	16.69	24.41	18.16	11.98	12.60	18.62
$\langle X \rangle$ is not equivalent to $\langle Y \rangle$	23.50	3.92	16.90	7.80	8.06	9.23	11.57
$\langle X \rangle \setminus \text{n can be summarized as } \setminus \text{n } \langle Y \rangle$	37.46	26.24	29.42	22.62	17.68	17.15	25.10
[SRC]: $\langle X \rangle \setminus \text{n [SRC]: } \langle Y \rangle$	19.03	8.21	15.96	6.37	7.57	4.40	10.26

Table 3: Effects of using different templates for in-context learning. “ $\langle X \rangle$ ” and “ $\langle Y \rangle$ ” denote the placeholder for source and target sentence respectively. “[SRC]” and “[TGT]” represent the placeholder for source and target language name in English. Bold text denotes the highest score along the column.

Zhu, Wenhao, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Jiajun Chen, Lei Li and Shujian Huang. “Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis.” *ArXiv* abs/2304.04675 (2023)

Document-Level Machine Translation with Large Language Models

Longyue Wang^{1*} Chenyang Lyu^{2*} Tianbo Ji^{2*} Zhirui Zhang^{1*}

Dian Yu¹ Shuming Shi¹ Zhaopeng Tu¹

¹Tencent AI Lab

²Dublin City University

{vinnylywang, jackzrzhang, shumingshi, zptu}@tencent.com

{chenyang.lyu2, tianbo.ji2}@mail.dcu.ie

Abstract

Large language models (LLMs) such as ChatGPT can produce coherent, cohesive, relevant, and fluent answers for various natural language processing (NLP) tasks. Taking document-level machine translation (MT) as a testbed, this paper provides an in-depth evaluation of LLMs' ability on discourse modeling. The study focuses on three aspects: 1) *Effects of Discourse-Aware Prompts*, where we investigate the impact of different prompts on document-level translation quality and discourse phenomena; 2) *Comparison of Trans-*



15 Apr 2023

Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis

Wenhao Zhu^{1,2*}, Hongyi Liu^{3*}, Qingxiu Dong⁴, Jingjing Xu²

Shujian Huang¹, Lingpeng Kong^{2,5}, Jiajun Chen¹, Lei Li⁶,

¹National Key Laboratory for Novel Software Technology, Nanjing University

²Shanghai AI Lab ³Shanghai Jiao Tong University ⁴Peking University

⁵The University of Hong Kong ⁶University of California, Santa Barbara

zhuwh@mail.nju.edu.cn, liu.hong.yi@sjtu.edu.cn, dqx@stu.pku.edu.cn, jingjingxu@pku.edu.cn

huangsj@nju.edu.cn, lpk@cs.hku.hk, chenjj@nju.edu.cn, lilei@cs.ucsb.edu,

Abstract

Large language models (LLMs) have demonstrated remarkable potential in handling multilingual machine translation (MMT). In this paper, we systematically investigate the advantages and challenges of LLMs for MMT by answering two questions: 1) How well do LLMs perform in translating a massive number of languages? 2) Which factors affect LLMs' performance in translation? We evaluate popular LLMs, including XGLM, OPT, BLOOMZ, and ChatGPT, on 102 languages¹. Our empirical results show that even the currently best-performed LLM, i.e. ChatGPT, still lags behind the supervised baseline NLLB in 83.33% of translation directions. Through further anal-

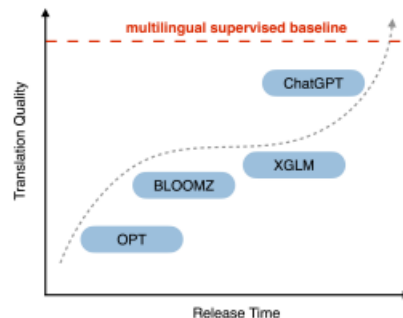


Figure 1: Averaged multilingual translation performance of popular LLMs. These models demonstrated great potential in multilingual machine translation.

Introducing PaLM 2

May 10, 2023 · 4 min read

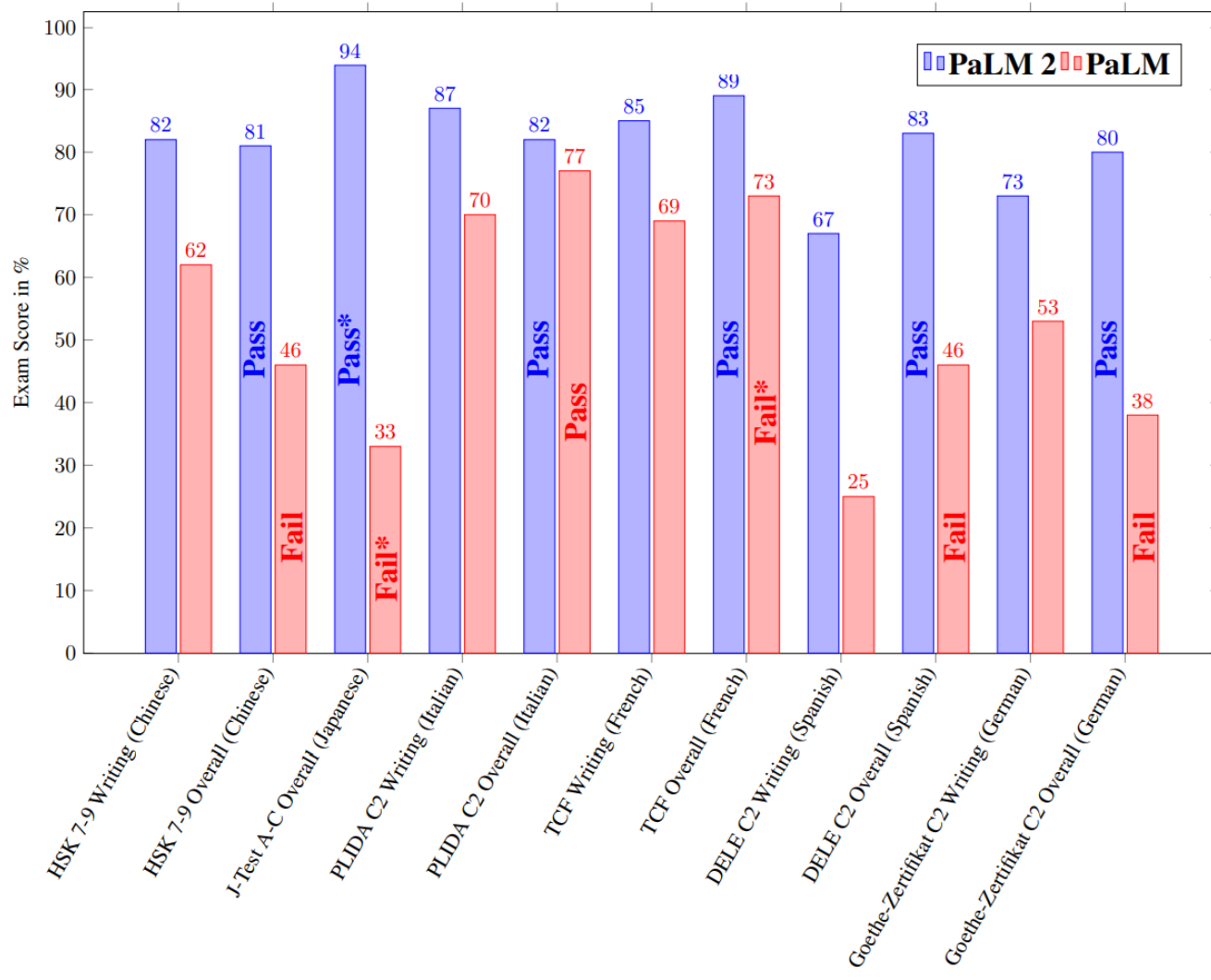


Zoubin Ghahramani

Vice President, Google DeepMind



When you look back at the biggest breakthroughs in AI over the last decade, Google has been at the forefront of so many of them. Our groundbreaking work in foundation models has become the bedrock for the industry and the AI-powered products that billions of people use daily. As we continue to responsibly advance these technologies, there's great potential for transformational uses in areas as far reaching as healthcare and human creativity.



“Can incontinence be cured?”

Clinician answer

Urinary incontinence can be cured in some cases, but this depends on its cause. If it's due to an underlying urinary tract infection or constipation, treating these issues will immediately stop the incontinence. If it's due to muscle weakness, such as that caused by childbirth, strengthening exercises to build muscle tone will help and can even cure it completely, over time. If it's due to an underlying nerve problem, damage to the spinal cord, or a structural abnormality, this is usually much harder to treat and will need to be assessed on a case-by-case basis.

Med-PaLM answer

Incontinence is a common problem, especially in women who have given birth. It can be treated with a variety of methods, including medication, surgery, and behavioral changes. Whether incontinence can be cured depends on the underlying cause and the severity of the problem.

Clinician review of Med-PaLM answer

- ✓ How does it relate to the current consensus of the scientific and clinical community?
- ✓ What is the extent of possible harm?
- ✓ What is the likelihood of possible harm?
- ✓ Correct reading comprehension
- ✓ Correct recall of knowledge
- ✓ Correct reasoning step
- ✓ Incorrect reading comprehension
- ✓ Incorrect recall of knowledge
- ✓ Incorrect reasoning step
- ✓ Does this answer contain any content that it shouldn't
- ⚠ Does this answer omit important information?
- ✓ Does the answer provided contain any information that is inapplicable or inaccurate for one particular patient-group or demographic?

Join at
slido.com
#2309 239

