



Translation Education and Challenges of AI (Part 2)

Sai Cheong Siu

10 May 2023



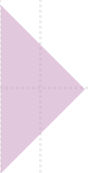
目的

- 从技术角度探讨人工智能翻译技术之限制，进而提出几点教学上的建议，让学生更好地结合人机互补优势



主要内容

- 人工智能翻译技术的概况
- 目前技术的不足之处
- 翻译教学的几点建议



人工智能翻译技术

人工智能翻译技术：两大类型

神经机器翻译
(Neural Machine
Translation (NMT))

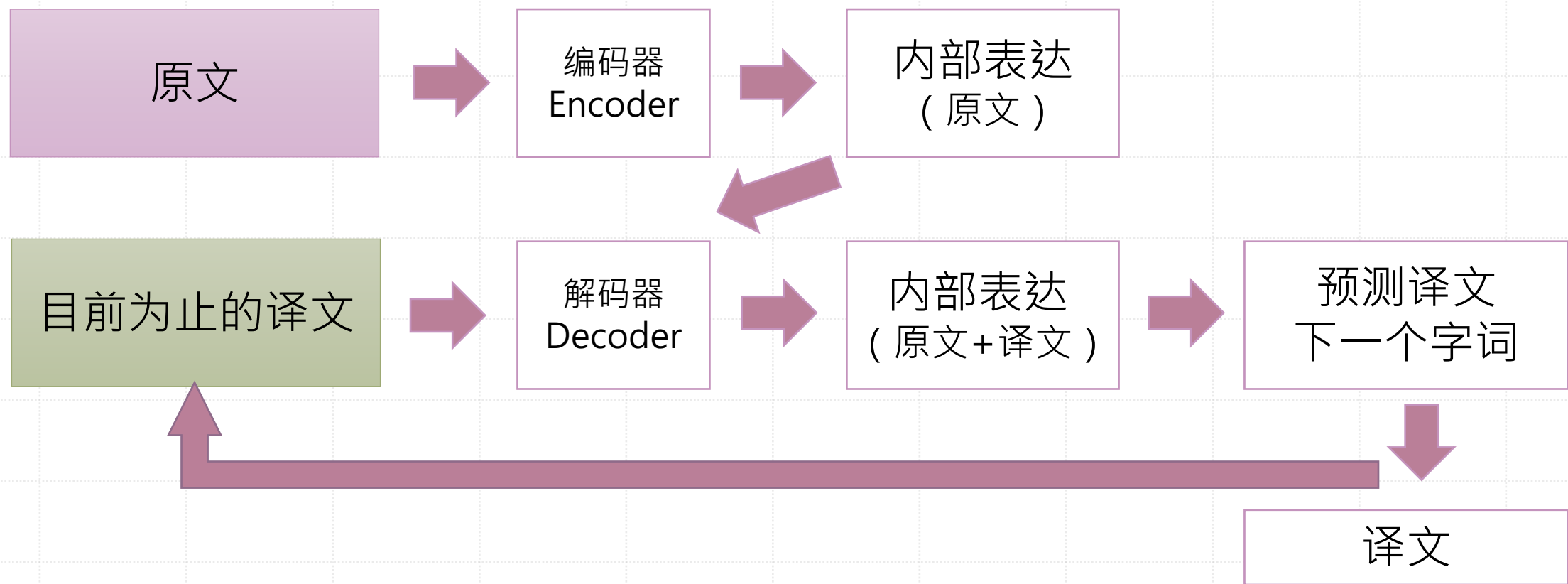
大型语言模型
(Large Language
Models (LLMs))



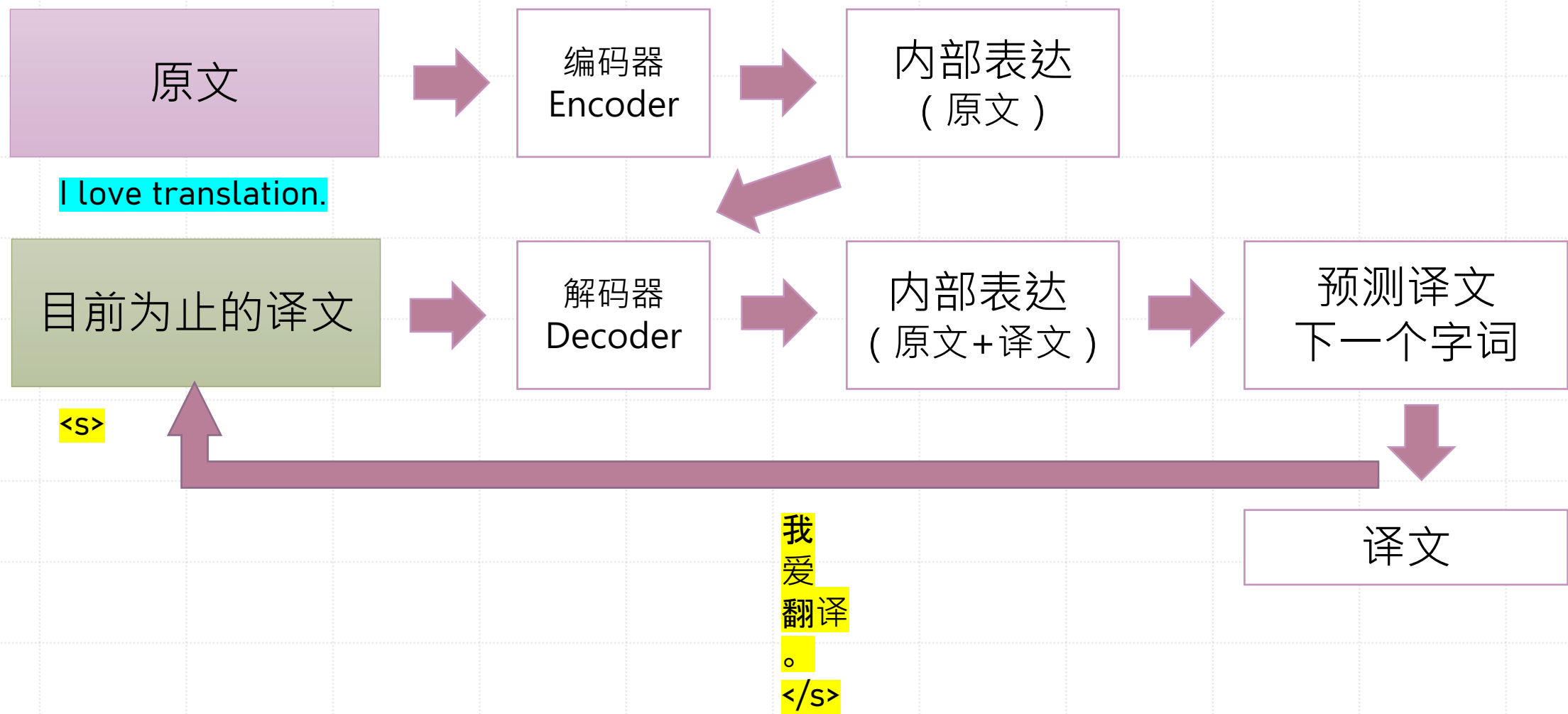
基本原理

- 方法：利用人工神经网络进行字词预测
- 常用架构：基于自注意力模型 (Self-attention neural network)
- 详参Siu (2023b)

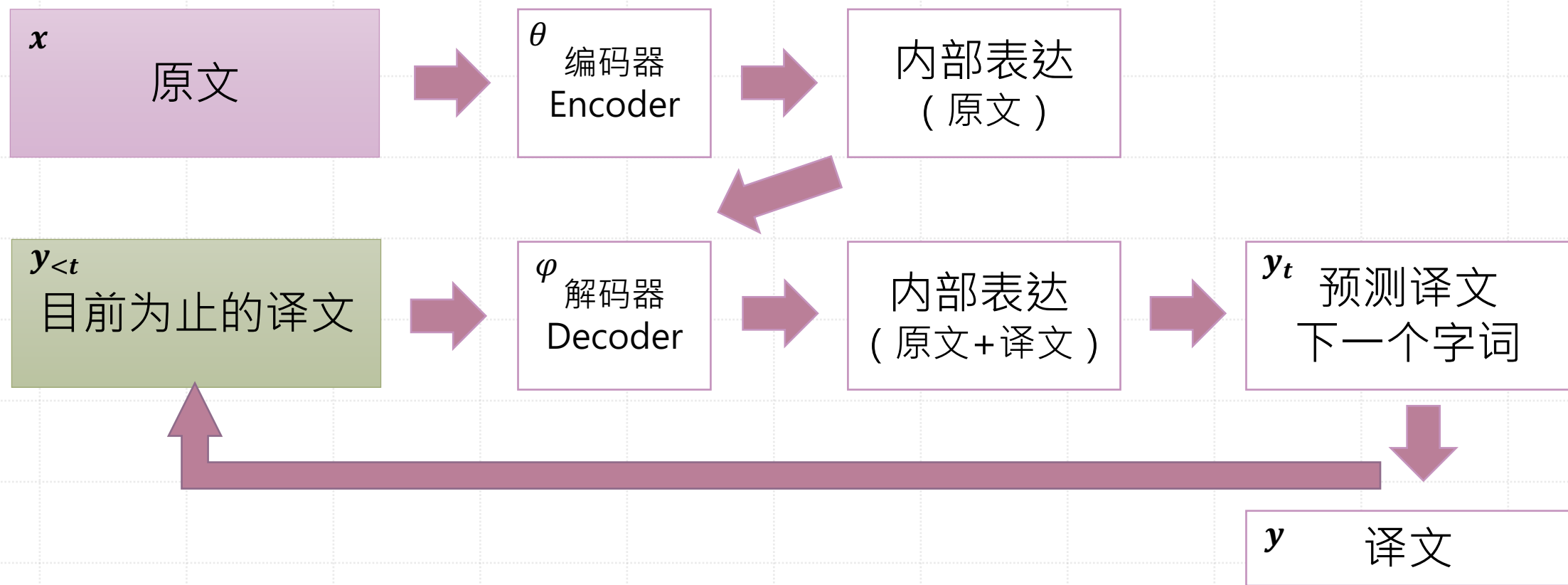
神经机器翻译



神经机器翻译

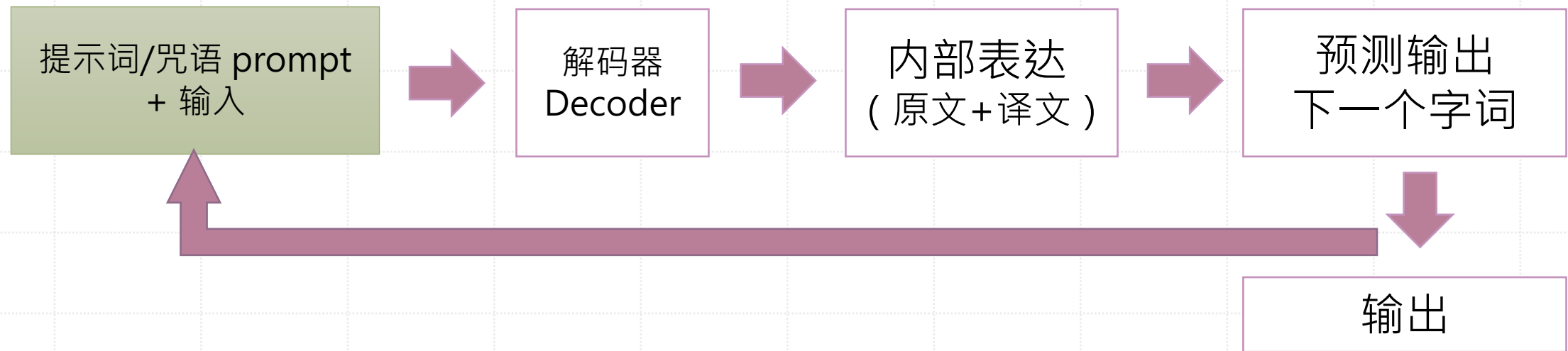


神经机器翻译



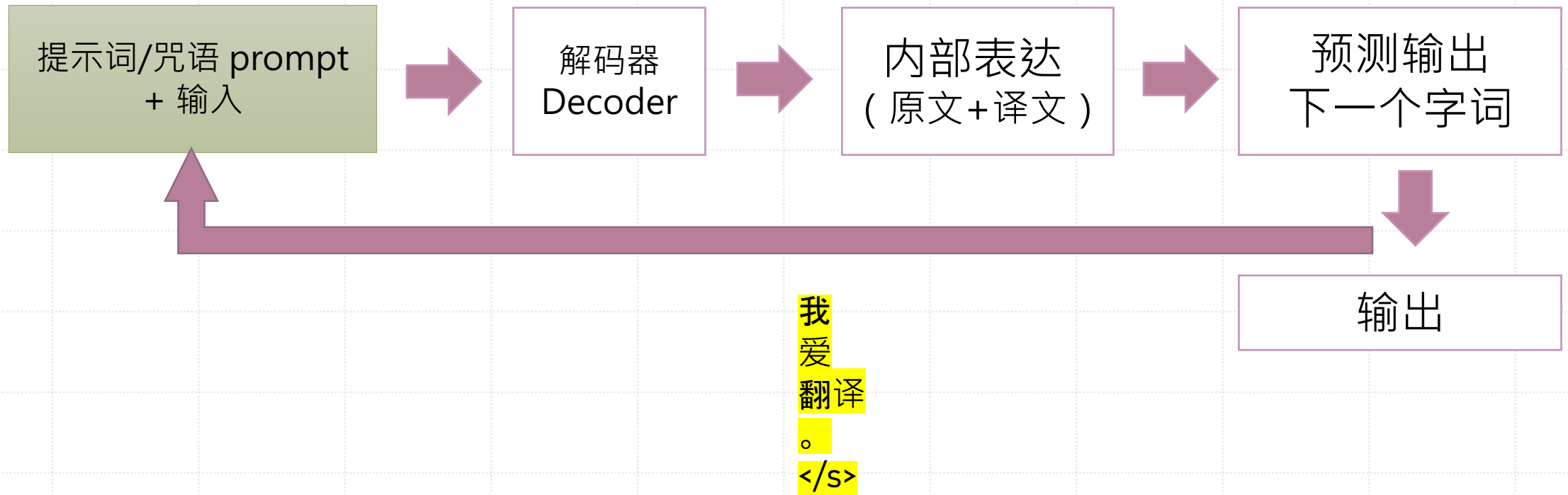
$$p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n p_{\theta, \varphi}(y_t | \mathbf{x}, \mathbf{y}_{<t})$$

ChatGPT一类基于解码器的大型语言模型

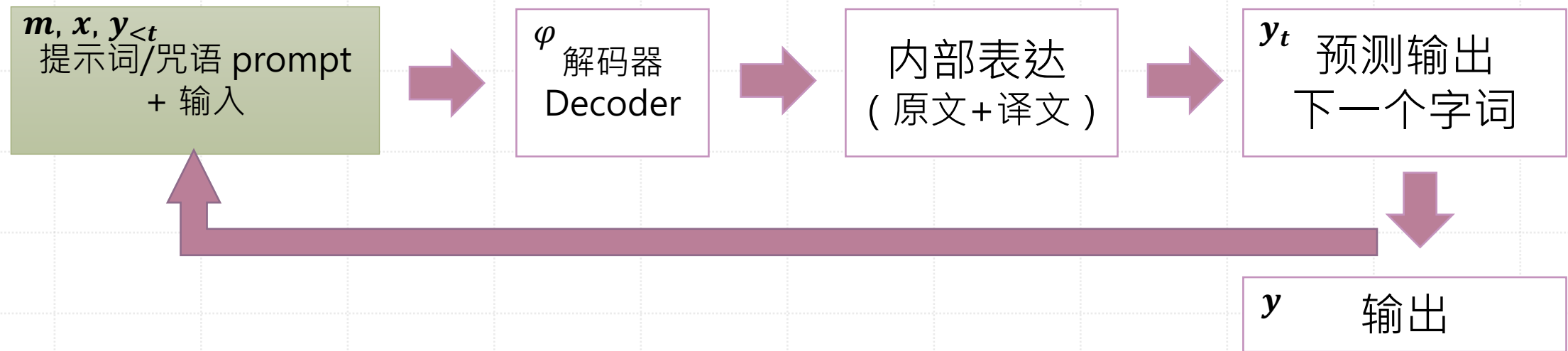


ChatGPT一类基于解码器的大型语言模型

Translate the following into Chinese: I love translation.

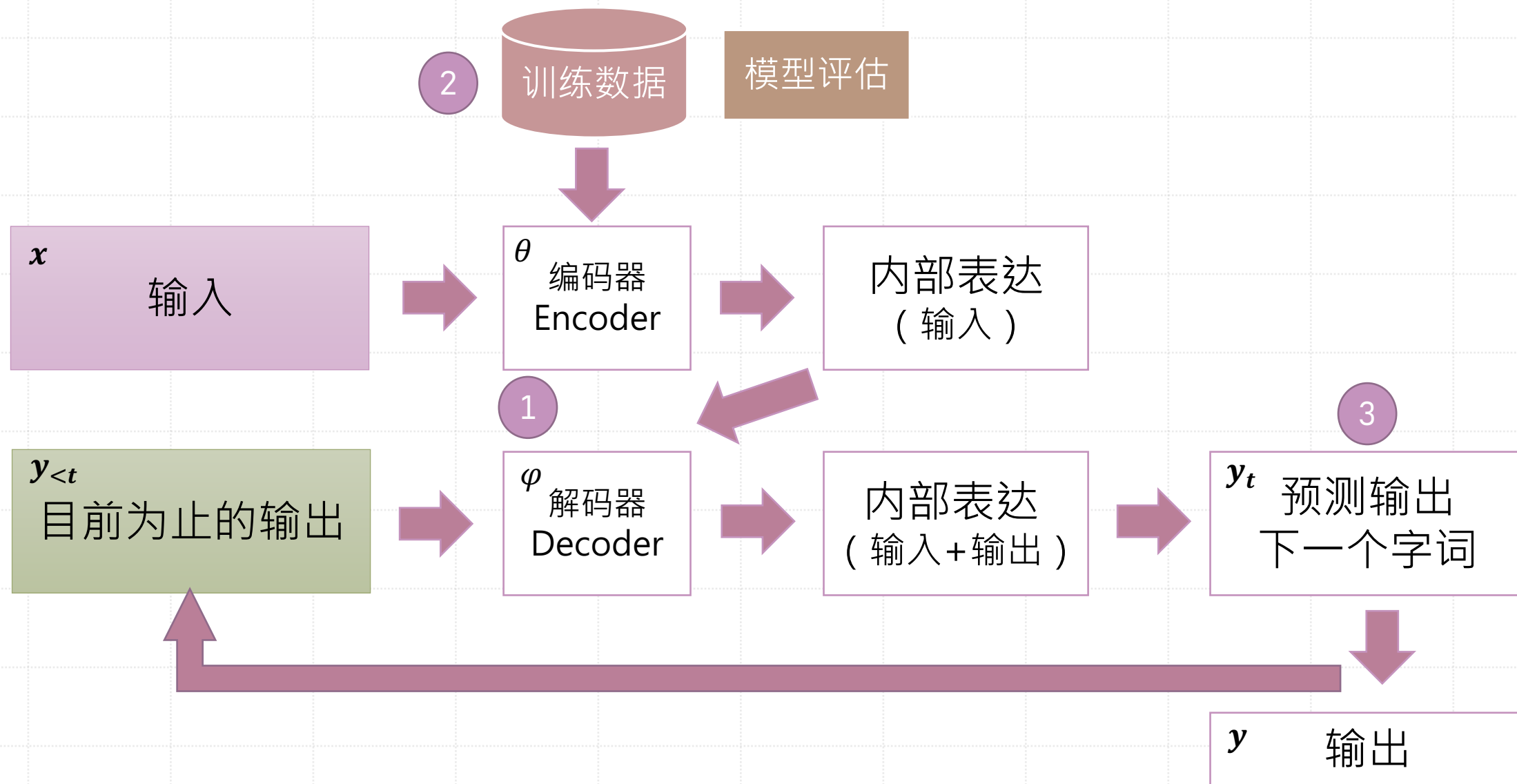


ChatGPT一类基于解码器的大型语言模型



$$p(y|m, x) = \prod_{t=1}^n p_{\varphi}(y_t|m, x, y_{<t})$$

三个重点：模型建设、模型训练、预测输出





目前技术的不足之处



目前技术的不足之处

- 长句处理
- 专门领域
- 中文支持
- 语境分析和背景知识
- 瞎编乱造、内容过时

一、长句处理

- 原文（繁中）：中央始終十分關心香港的疫情形勢，多措並舉支持香港統籌做好疫情防控 and 經濟社會發展工作，中央有關部門高度重視在疫情常態化防控下恢復內地與香港“通關”的問題，一直與特區政府保持密切溝通，深入交換意見，推動兩地防控措施的有效銜接。

見https://www.hmo.gov.cn/xwzx/xwfb/xwfb_child/202111/t20211104_22957.html。

一、长句处理

- 原文（繁中）：中央始終十分關心香港的疫情形勢，多措並舉支持香港統籌做好疫情防控和經濟社會發展工作，中央有關部門高度重視在疫情常態化防控下恢復內地與香港“通關”的問題，一直與特區政府保持密切溝通，深入交換意見，推動兩地防控措施的有效銜接。
- 机器翻译1：The relevant departments of the Central Authorities attach great importance to the resumption of "customs clearance" between the Mainland and Hong Kong under the normalised prevention and control of the epidemic, and have been maintaining close communication with the SAR Government, exchanging views in depth and promoting the effective interface of prevention and control measures between the two places.

一、长句处理

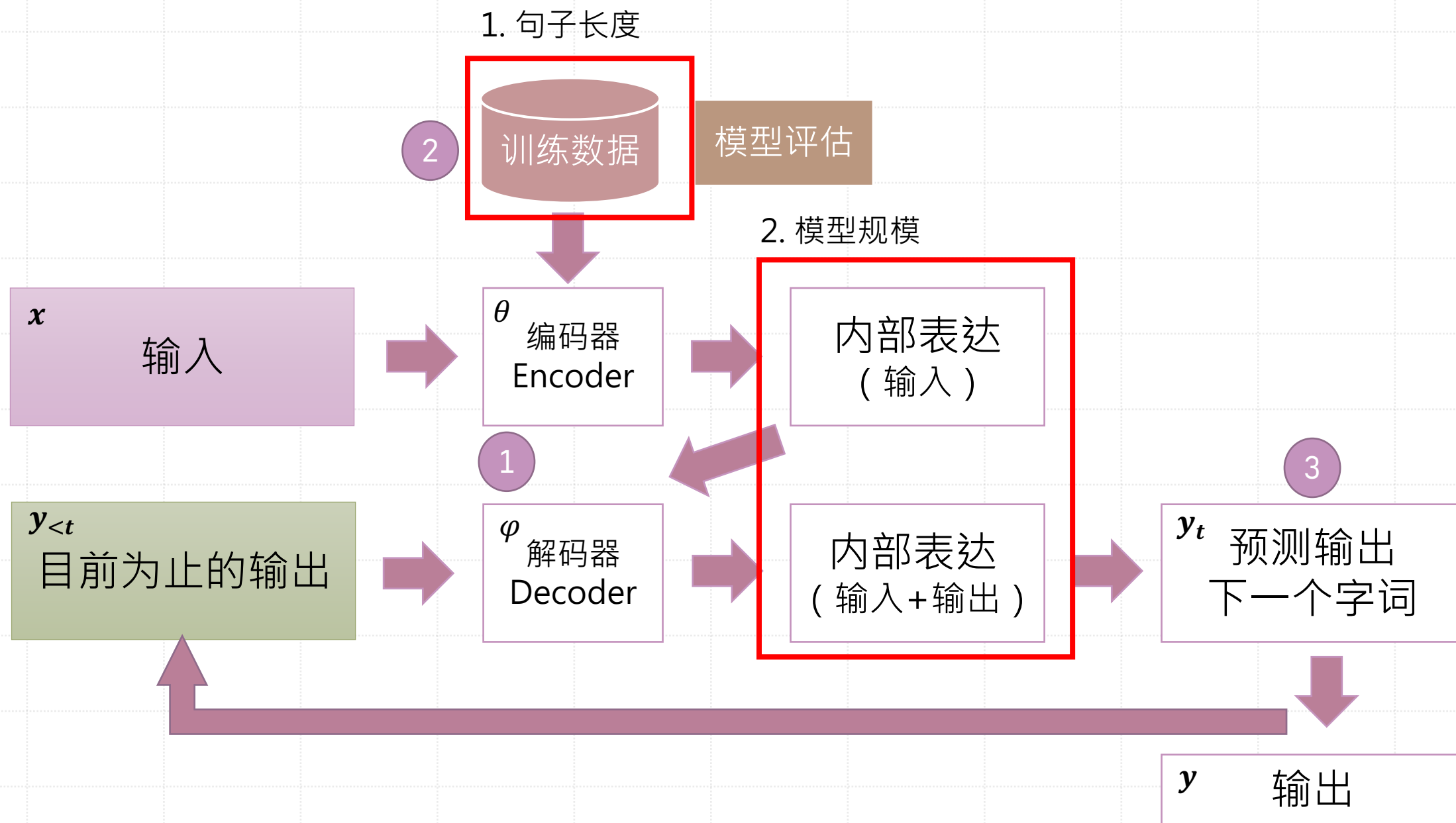
- 原文（繁中）：中央始終十分關心香港的疫情形勢，多措並舉支持香港統籌做好疫情防控和經濟社會發展工作，中央有關部門高度重視在疫情常態化防控下恢復內地與香港“通關”的問題，一直與特區政府保持密切溝通，深入交換意見，推動兩地防控措施的有效銜接。
- 机器翻译1：The relevant departments of the Central Authorities attach great importance to the resumption of "customs clearance" between the Mainland and Hong Kong under the normalised prevention and control of the epidemic, and have been maintaining close communication with the SAR Government, exchanging views in depth and promoting the effective interface of prevention and control measures between the two places.

一、长句处理

- 原文（繁中）：中央始終十分關心香港的疫情形勢，多措並舉支持香港統籌做好疫情防控和經濟社會發展工作，中央有關部門高度重視在疫情常態化防控下恢復內地與香港“通關”的問題，一直與特區政府保持密切溝通，深入交換意見，推動兩地防控措施的有效銜接。
- 机器翻译2：The central government has always been very concerned about the epidemic situation in Hong Kong, and has taken multiple measures to support Hong Kong in coordinating the work of epidemic prevention and control and economic and social development. The SAR government maintains close communication, exchanges views in depth, and promotes the effective convergence of prevention and control measures between the two places.

一、长句处理

- 原文（繁中）：中央始終十分關心香港的疫情形勢，多措並舉支持香港統籌做好疫情防控和經濟社會發展工作，中央有關部門高度重視在疫情常態化防控下恢復內地與香港“通關”的問題，一直與特區政府保持密切溝通，深入交換意見，推動兩地防控措施的有效銜接。
- 机器翻译2：The central government has always been very concerned about the epidemic situation in Hong Kong, and has taken multiple measures to support Hong Kong in coordinating the work of epidemic prevention and control and economic and social development. The SAR government maintains close communication, exchanges views in depth, and promotes the effective convergence of prevention and control measures between the two places.



二、专门领域

- 术语和译文风格或有偏差，以下以招股说明书为例
- 原文：You should rely only on the information contained in this prospectus to make your investment decision.
- 参考译文（繁中）：閣下應僅依賴本招股章程所載資料作出投資決定。
- 机器翻译：您應僅依賴本招股說明書中包含的信息來做出您的投資決定。

二、专门领域

- 术语和译文风格或有偏差，以下以招股说明书为例
- 原文：You should rely only on the information contained in this prospectus to make your investment decision.
- 参考译文（繁中）：閣下應僅依賴本招股章程所載資料作出投資決定。
- 机器翻译：您應僅依賴本招股說明書中包含的信息來做出您的投資決定。

二、专门领域

- IPOTranslate: 我们的招股说明书翻译系统
- 输出（繁中）：閣下應僅依賴本招股章程所載資料作出投資決定。



三、中文支持

- ChatGPT的训练数据以英文为主，中文输入/输出有改进空间
- 以下一例摘自Siu (2023a)

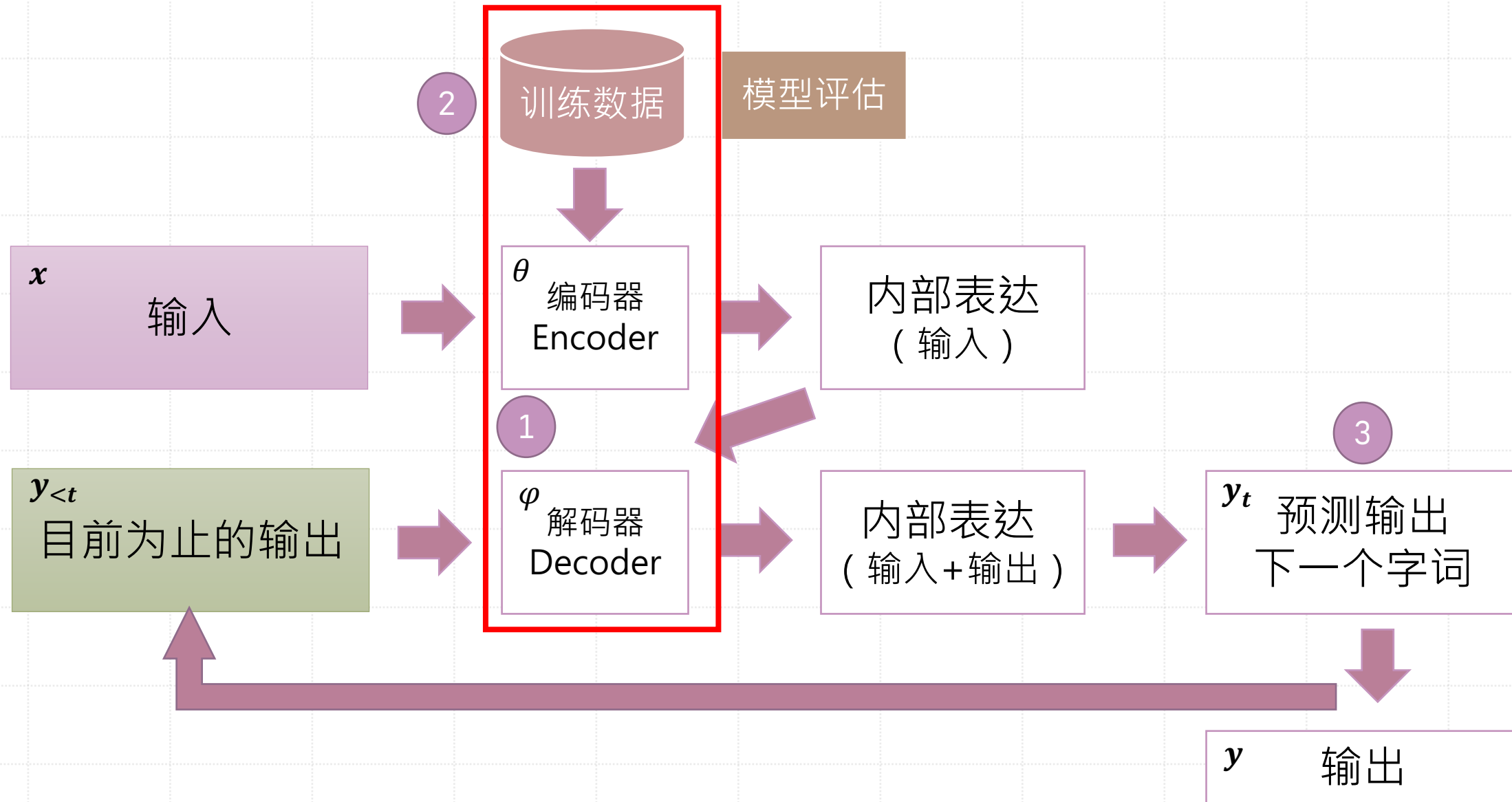
三、中文支持

- 用家输入：Translate the following into Traditional Chinese:
Hong Kong, with a population of about 7.51 million in mid-2019, is a small but dynamic city which has earned an international reputation as a leading commercial and financial centre as well as a highly efficient entrepot.
- ChatGPT 输出：香港是一座人口约在2019年中期达到750万的小而有活力的城市，不仅因为是一个高效的贸易港口而享有国际声誉，同时也是一个领先的商业和金融中心。

三、中文支持

- 用家输入：Translate the following into Traditional Chinese:
Hong Kong, with a population of about 7.51 million in mid-2019, is a small but dynamic city which has earned an international reputation as a leading commercial and financial centre as well as a highly efficient entrepot.
- ChatGPT 输出：香港是一座人口约在2019年中期达到750万的小而有活力的城市，不仅因为是一个高效的贸易港口而享有国际声誉，同时也是一个领先的商业和金融中心。

训练数据与实际输入有重大差异



四、语境分析和背景知识

- 基于自注意力模型在上下文分析有优势（考虑句子或段落内各词如何互相影响），但如果文本内容、结构比较复杂，又或者需要更多输入文本或训练数据以外的背景知识，则仍有不少提升空间。

原文（繁中）

- 今日是年廿七，再過幾日就是大年初一，我們將會送虎迎兔。我先在這裏預祝大家新一年事事如意、身體健康。新一年香港的第一件大事必然是與內地逐步、有序、全面通關。由1月8日實施首階段通關到昨日，剛好十天，已經有差不多40萬人次經陸路口岸免檢疫往返兩地，由今日到年三十晚，更會是高峰期。見到兩地人民家庭團聚，在春節前達成重聚的心願，笑容滿面，團團圓圓，溫馨感人，我非常欣慰。

見https://www.news.gov.hk/chi/2023/01/20230118/20230118_172609_554.html。



ChatGPT翻译

- Today is the 27th day of the year, and in a few days it will be the first day of the Lunar New Year. We will be sending off the Year of the Tiger and welcoming the Year of the Rabbit. Here, I would like to wish everyone a happy new year with good health and everything going smoothly.
- The first big event in Hong Kong for the new year will undoubtedly be the gradual, orderly, and comprehensive customs clearance with mainland China. From the implementation of the first phase of customs clearance on January 8th to yesterday, which was exactly ten days, nearly 400,000 people have crossed the border without quarantine inspection between the two places by land. From today until the evening of the thirtieth day of the new year, it will be even busier. Seeing people from both places reunited with their families before the Spring Festival, with smiles on their faces, warmth and affection, I am very pleased.

ChatGPT翻译

- Today is the 27th day of the year, and in a few days it will be the first day of the Lunar New Year. We will be sending off the Year of the Tiger and welcoming the Year of the Rabbit. Here, I would like to wish everyone a happy new year with good health and everything going smoothly.
- The first big event in Hong Kong for the new year will undoubtedly be the gradual, orderly, and comprehensive customs clearance with mainland China. From the implementation of the first phase of customs clearance on January 8th to yesterday, which was exactly ten days, nearly 400,000 people have crossed the border without quarantine inspection between the two places by land. From today until the evening of the thirtieth day of the new year, it will be even busier. Seeing people from both places reunited with their families before the Spring Festival, with smiles on their faces, warmth and affection, I am very pleased.

ChatGPT翻译

- Today is the 27th day of the year, and in a few days it will be the first day of the Lunar New Year. We will be sending off the Year of the Tiger and welcoming the Year of the Rabbit. Here, I would like to wish everyone a happy new year with good health and everything going smoothly.
- The first big event in Hong Kong for the new year will undoubtedly be the gradual, orderly, and comprehensive customs clearance with mainland China. From the implementation of the first phase of customs clearance on January 8th to yesterday, which was exactly ten days, nearly 400,000 people have crossed the border without quarantine inspection between the two places by land. From today until the evening of the thirtieth day of the new year, it will be even busier. Seeing people from both places reunited with their families before the Spring Festival, with smiles on their faces, warmth and affection, I am very pleased.

瞎编乱造、内容过时

- 就大型语言模型而言，最大的问题就是瞎编译文。
- 虚构没有发生过、没有做过的事情。
- 模型训练后，如果已经过了一段时间，训练时所「学到」的知识可能不再适用，遇到要解决超出训练语料范围的问题，也有瞎编的可能。

例子

律师引用聊天机械人提供的案例，
但有关案例全属虚构，不过机械人
仍坚持案例是真实的。



The New York Times

@nytimes

A lawyer representing a man who sued an airline relied on ChatGPT to help prepare a court filing.

It did not go well.



nytimes.com

Here's What Happens When Your Lawyer Uses ChatGPT

A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.

<https://twitter.com/nytimes/status/1662483832128385024>

ChatGPT的「知识」至2021年为止

Who's the Chief executive of HKSAR government?

The Chief Executive of the [Hong Kong Special Administrative Region](#) (HKSAR) government as of my knowledge cutoff of September 2021 is Carrie Lam Cheng Yuet-ngor. However, please note that my training data only goes up until that point, so there may have been a change in leadership since then.

World Cup Champion 2022 is Argentina

I'm sorry, but as an [AI language model](#), I cannot confirm that Argentina won the 2022 [World Cup](#) as it has not taken place yet. The tournament is scheduled to take place from November 21 to December 18, 2022, in Qatar. While there may be predictions or rumors about the outcome of the tournament, the official winner will not be determined until the competition is held.

有趣例子（ 见Siu (2023) ）

- Translate the following into English: 行政長官李家超今日（三月十五日）繼續在北京的訪問行程。他先後拜訪國家科學技術部、文化和旅遊部、海關總署和商務部，與各部門領導就共同關注的事宜交換意見。文化體育及旅遊局局長楊潤雄、政制及內地事務局局長曾國衛、保安局局長鄧炳強、商務及經濟發展局局長丘應樺、醫務衛生局局長盧寵茂教授和創新科技及工業局局長孫東教授亦有出席相關會面。

ChatGPT输出结果

- On March 15th, the Chief Executive, Mr. Carrie Lam, continued her visit to Beijing. She visited the Ministry of Science and Technology, the Ministry of Culture and Tourism, the General Administration of Customs, and the Ministry of Commerce, and exchanged views with leaders of these departments on matters of common concern. The Director of the Tourism Bureau, Mr. Yang Runxiong, the Director of the Constitutional and Mainland Affairs Bureau, Mr. Patrick Nip, the Director of the Security Bureau, Mr. John Lee Ka-chiu, the Director of the Commerce and Economic Development Bureau, Mr. Edward Yau Tang-wah, the Director of Health, Professor Sophia Chan Siu-chee, and the Director of the Innovation and Technology Bureau, Professor Alfred Sit Wing-hang, also attended the relevant meetings.

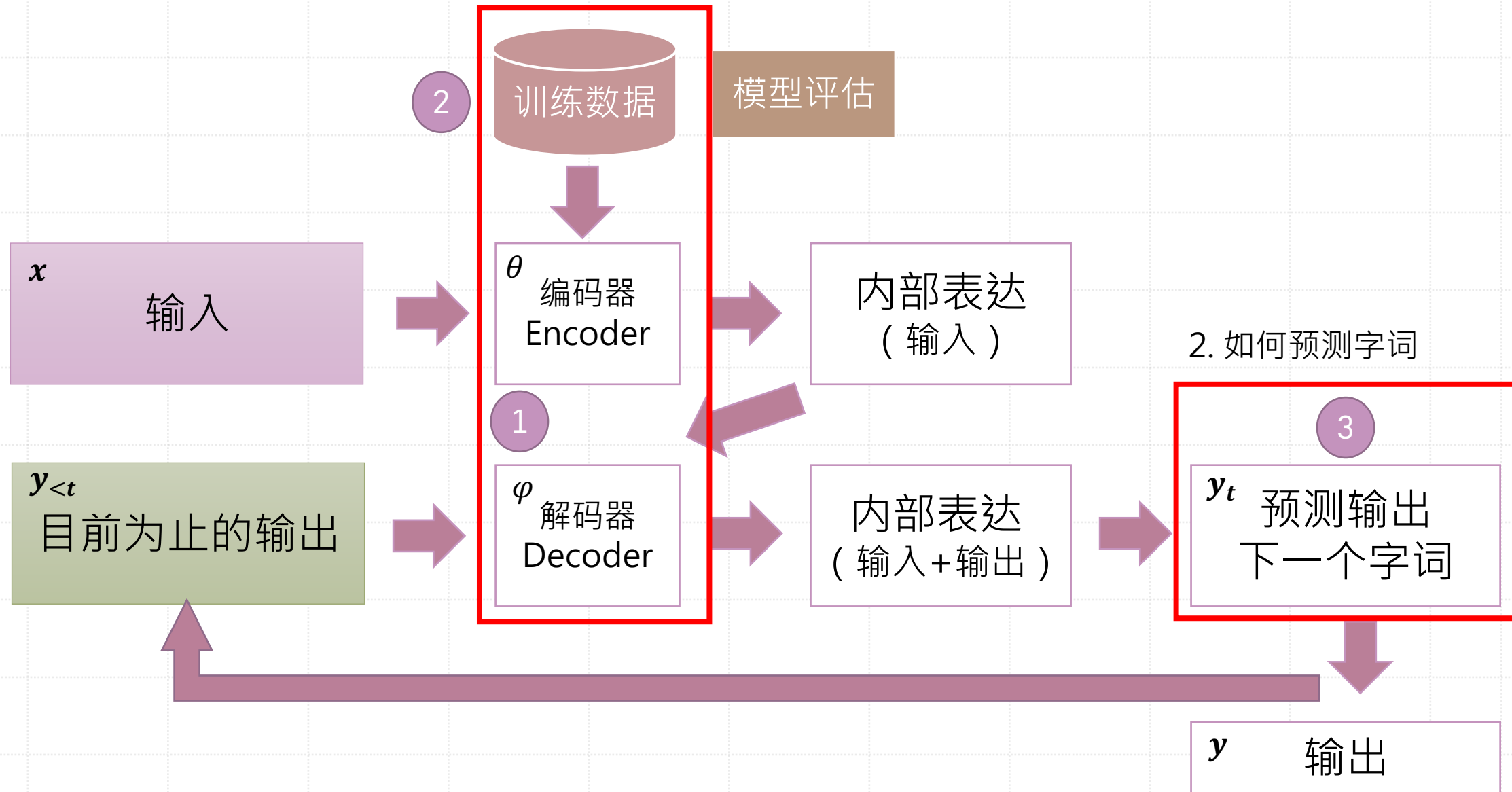
ChatGPT输出结果

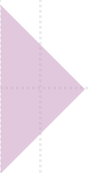
- On March 15th, the Chief Executive, Mr. Carrie Lam, continued her visit to Beijing. She visited the Ministry of Science and Technology, the Ministry of Culture and Tourism, the General Administration of Customs, and the Ministry of Commerce, and exchanged views with leaders of these departments on matters of common concern. The Director of the Tourism Bureau, Mr. Yang Runxiong, the Director of the Constitutional and Mainland Affairs Bureau, Mr. Patrick Nip, the Director of the Security Bureau, Mr. John Lee Ka-chiu, the Director of the Commerce and Economic Development Bureau, Mr. Edward Yau Tang-wah, the Director of Health, Professor Sophia Chan Siu-chee, and the Director of the Innovation and Technology Bureau, Professor Alfred Sit Wing-hang, also attended the relevant meetings.

ChatGPT输出结果

- On March 15th, the Chief Executive, Mr. Carrie Lam, continued her visit to Beijing. She visited the Ministry of Science and Technology, the Ministry of Culture and Tourism, the General Administration of Customs, and the Ministry of Commerce, and exchanged views with leaders of these departments on matters of common concern. The Director of the Tourism Bureau, Mr. Yang Runxiong, the Director of the Constitutional and Mainland Affairs Bureau, Mr. Patrick Nip, the Director of the Security Bureau, Mr. John Lee Ka-chiu, the Director of the Commerce and Economic Development Bureau, Mr. Edward Yau Tang-wah, the Director of Health, Professor Sophia Chan Siu-chee, and the Director of the Innovation and Technology Bureau, Professor Alfred Sit Wing-hang, also attended the relevant meetings.

1. 训练数据与实际输入有重大差异





翻译教学的几点建议

译者的机遇



更佳的译稿，承接更多工作



专门领域的翻译工作



牵涉复杂文本、语境、背景知识、事件、人物关系



校对、编辑



语料整理、模型评估

建议一：加强基本翻译知识和技能

语言技能

翻译技能

跨文化技能

百科知识

专门领域知识

译者角色的认识

协作技能

建议二：深入专门领域和多模态翻译



法律领域



医疗领域



多模态翻译



本地化工作

建议三：技术与人工智能



工具选择



工具操作：译前/后编辑、翻译记忆库架设与维护



人工智能的基本原理：机器学习、深度学习、模型训练



工具研发：网页编程、人工智能编程



大型语言模型相关技能：提示词开发与优化(prompt engineering)、解码参数配置(decoding settings)

例子：计算机辅助翻译硕士项目（必修课）

- TRA6001 翻译技术：知识和技能
- TRA6002 专业语言服务：技能和策略
- TRA6003 商务与法律翻译
- TRA6004 专业线上协作翻译

选修方向一：专业语言服务（计算机辅助）

- TRA6101 科技翻译
- TRA6102 国际贸易及运输翻译
- TRA6103 文学、历史及哲学翻译的电子工具
- TRA6104 医药翻译
- TRA6105 商务和法律传译
- TRA6106 数字娱乐翻译



选修方向二：翻译技术

- TRA6201 本地化和双语网站开发
- TRA6202 双语数字营销技术
- TRA6203 自然语言处理及翻译
- TRA6204 翻译研究技术
- TRA6205 翻译技术前沿
- TRA6206 口译科技

建议四：加强编辑的训练



机器翻译看似流畅，译者需要更仔细才能找出问题所在



面对大型模型瞎编乱作，译者对内容需要有鉴别真伪的能力，并且掌握适当的工具以作查证

建议五：面向大型语言模型的教学和评估活动

01

部分简单的翻译
教学和评估活动
或有重新设计的
必要

02

确保学生的翻译
能力得到准确评
估

03

确保学生掌握运用有关
模型的能力（包括模型
调用、评估、提示词开
发和优化、解码方法设
置等）



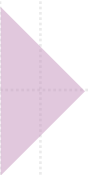
示例

- Instructions
- Please refer to the Chinese text and its accompanying English translations below (Version 1: NMT; Version 2: LLM). Answer the following questions.
- (a) Comment on the translation issues present in the two English translations provided.
- (b) Provide a revised English translation that improves accuracy, fluency, and clarity.
- (c) Explain the changes made.

更多内容

- ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation (Siu 2023a)
- Deep Learning and Translation Technology - Chapter 47, *Routledge Encyclopedia of Translation Technology* (2nd Edition) (Siu 2023b)





谢谢！