

Revolution of Technology for Translation and Localisation in the AI Era

Sai-cheong SIU
Associate Professor
Programme Director, Master of
Arts in Translation (Computer (-
Aided) Translation)
The Hang Seng University of Hong
Kong



Acknowledgements

- Funding Support: The work described in this talk is partially supported by the following:
 - a grant from the Innovation and Technology Commission (Project No. ITF/238/17);
 - a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. UGC/FDS14/H16/18);
 - a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. UGC/IDS(R)14/23).

Artificial intelligence (AI) is everywhere!

-  **Chatbots:** Facilitating human-like digital conversations
-  **Generative AI:** Creating new content, including images and music
-  **Game-playing AI:** Mastering and competing in complex games
-  **Image Recognition:** Classifying and categorizing images
-  **Object Detection:** Recognizing specific items within images
-  **Trend Analysis and Prediction:** Predicting future trends, such as stock prices and weather patterns
-  **Robotic Control:** Directing robotic operations with AI



AI can also enhance
translation technology
and professional
translation.

Outline

- AI for Translation: An Overview
- AI in Specialized Translation: Case Studies
 - Translation of Financial Documents
 - Platforms 1-4
 - Translation of Government Documents
 - Platforms 5-8

AI-driven innovations in financial translation: Examples



Platform 1: Term extractor



Platform 2: Text explainer



Platform 3: Sentence finder



Platform 4: IPOTranslate

AI-driven innovations in translation of government documents: Examples



Platform 5: GovPR Translation



Platform 6: Useful Expression Finder

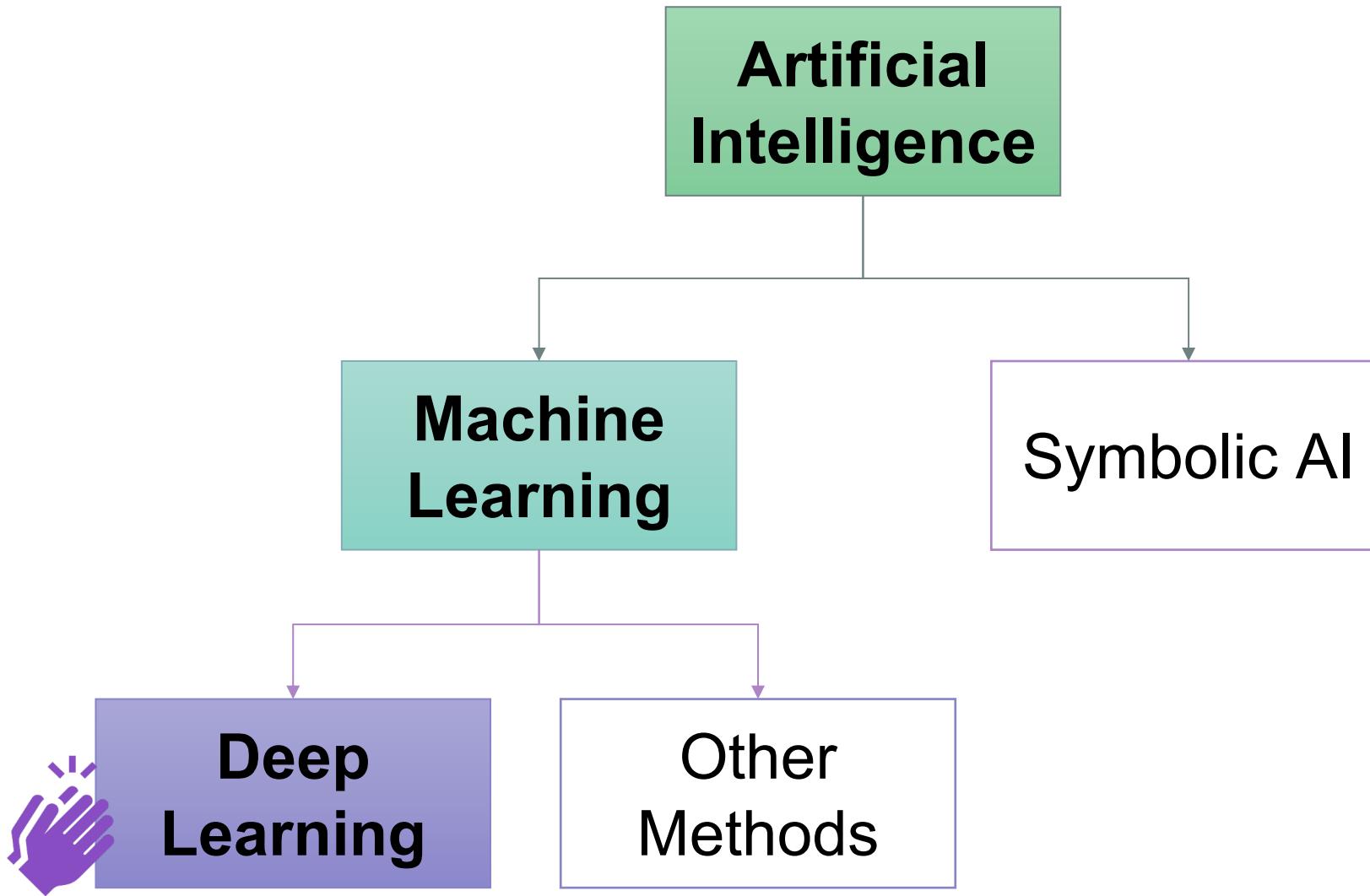


Platform 7: BlendMT



Platform 8: MTCheck

AI for Translation: An Overview



Features of deep learning



1. Multilayered artificial neurons



2. Feature extraction and representation

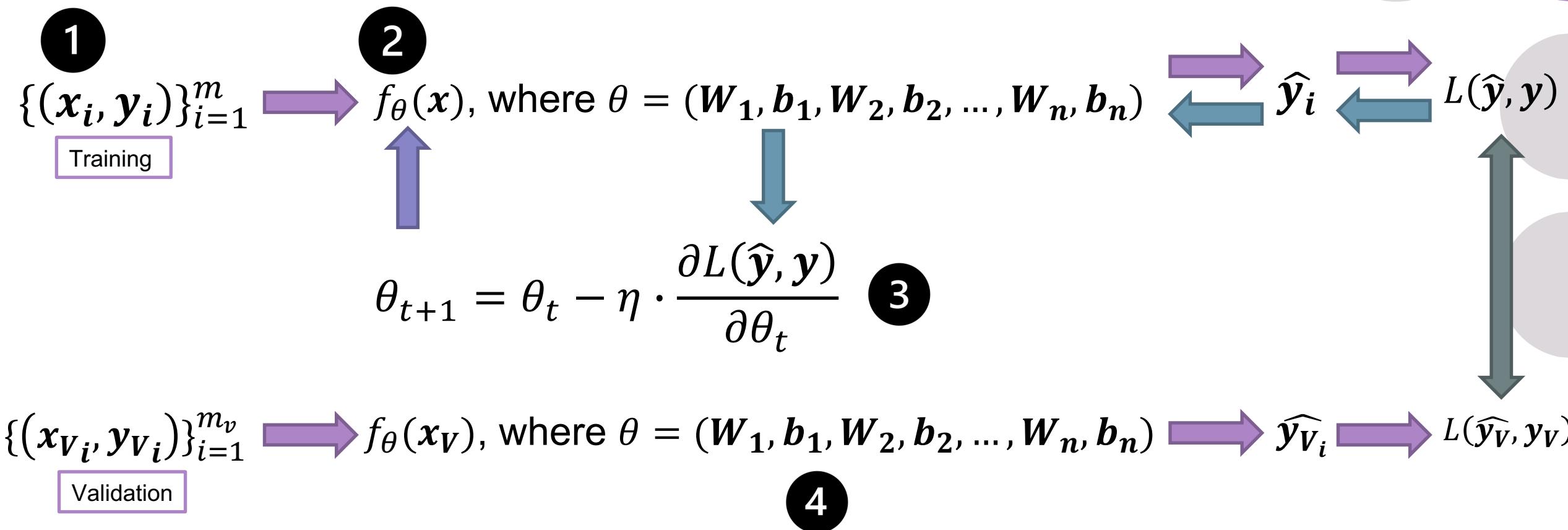


3. Versatility across domains



4. Resurgence in popularity

Formally, we have the following:



AI-driven tools applicable to translation: Examples

Neural machine
translation

Large language
models (e.g.,
ChatGPT)

Conventional methods

- I go to school by train.
- 我／去／學校／坐／火車／。
- 我坐火車去學校。



Rule-based
Machine
Translation

- I go to school by train.
- I go to school by taxi. 我
坐計程車去學校。
- 我坐火車去學校。



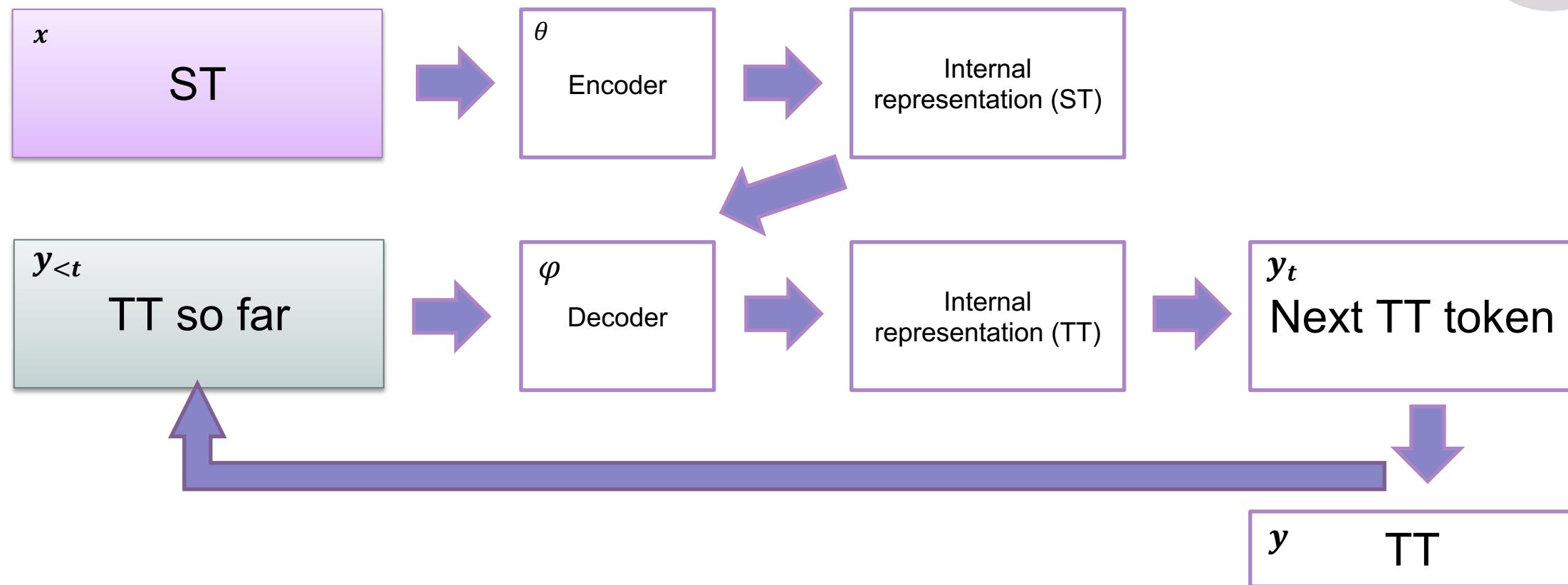
Example-based
Machine Translation

- Translation Model
 $P(\text{"I"}|\text{"我"}) \times P(\text{"go"}|\text{"去"}) \times \dots$
- Language Model
 $P(\text{"我坐火車去學校"})$



Statistical
Machine
Translation

Neural Machine Translation



Common types of networks

1. Recurrent Neural Networks (RNN)

Bidirectional RNN with Attention

Example:

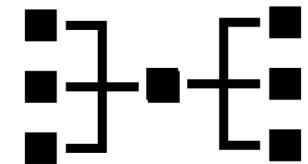
$$\begin{aligned}\vec{h}_t &= \vec{f}(x_i, \vec{h}_{t-1}) & \vec{h}_t &= \vec{f}(x_i, \vec{h}_{t+1}) \\ h_t &= [\vec{h}_t^\top; \vec{h}_t^\top]^\top \\ a(s_{t-1}, h_m) &= \frac{\exp(score(s_{t-1}, h_m))}{\sum_{k=1}^i \exp(score(s_{t-1}, h_k))} \\ \mathbf{z} &= q(\{h_1, \dots, h_i\}) = \sum_{m=1}^i a(s_{t-1}, h_m) h_m \\ p(y_t | y_{<t}, \mathbf{z}) &= g(s_{t-1}, y_{t-1}, \mathbf{z})\end{aligned}$$

2. Convolutional Neural Networks (CNN)

one-dimensional convolutional layer

Example:

$$y_{d,n} = \sum_{m=1}^M \sum_{k=0}^{K-1} w_{kM+m,n} x_{d+k,m}$$



ChatGPT

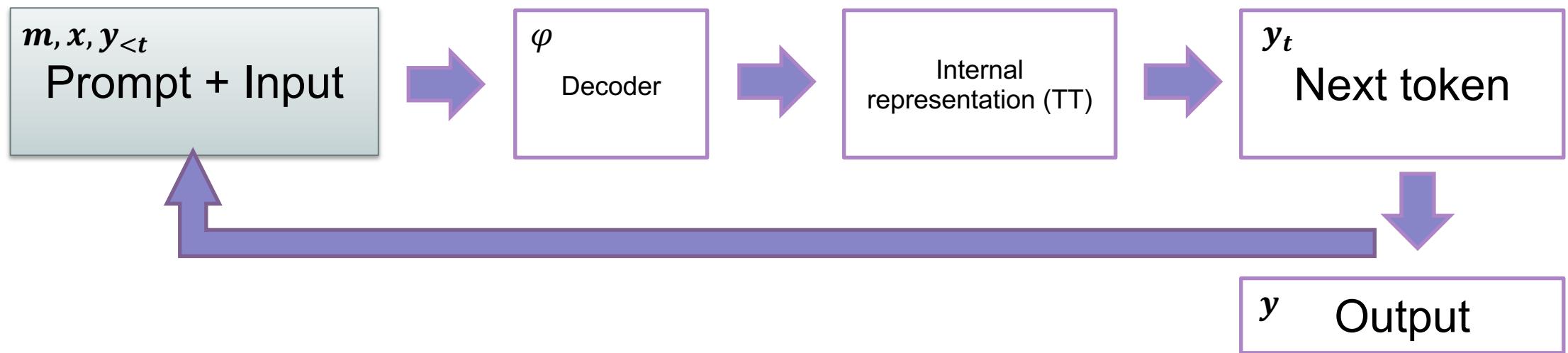
3. Self-attention Neural Networks (e.g., Transformer)

Multi-head Attention with Queries, Keys and Values

Example:

$$\begin{aligned}scoremat(Q, K) &= softmax(QK^\top \cdot scale) \\ attention(Q, K, V) &= (scoremat(Q, K))V \\ head_h &= attention(W_Q Q, W_K K, W_V V) = attention(Q_h, K_h, V_h) \\ multihead(Q, K, V) &= W_{multihead}[head_1; head_2; \dots; head_H]\end{aligned}$$

Decoder-based LLMs like ChatGPT



Formally, we have

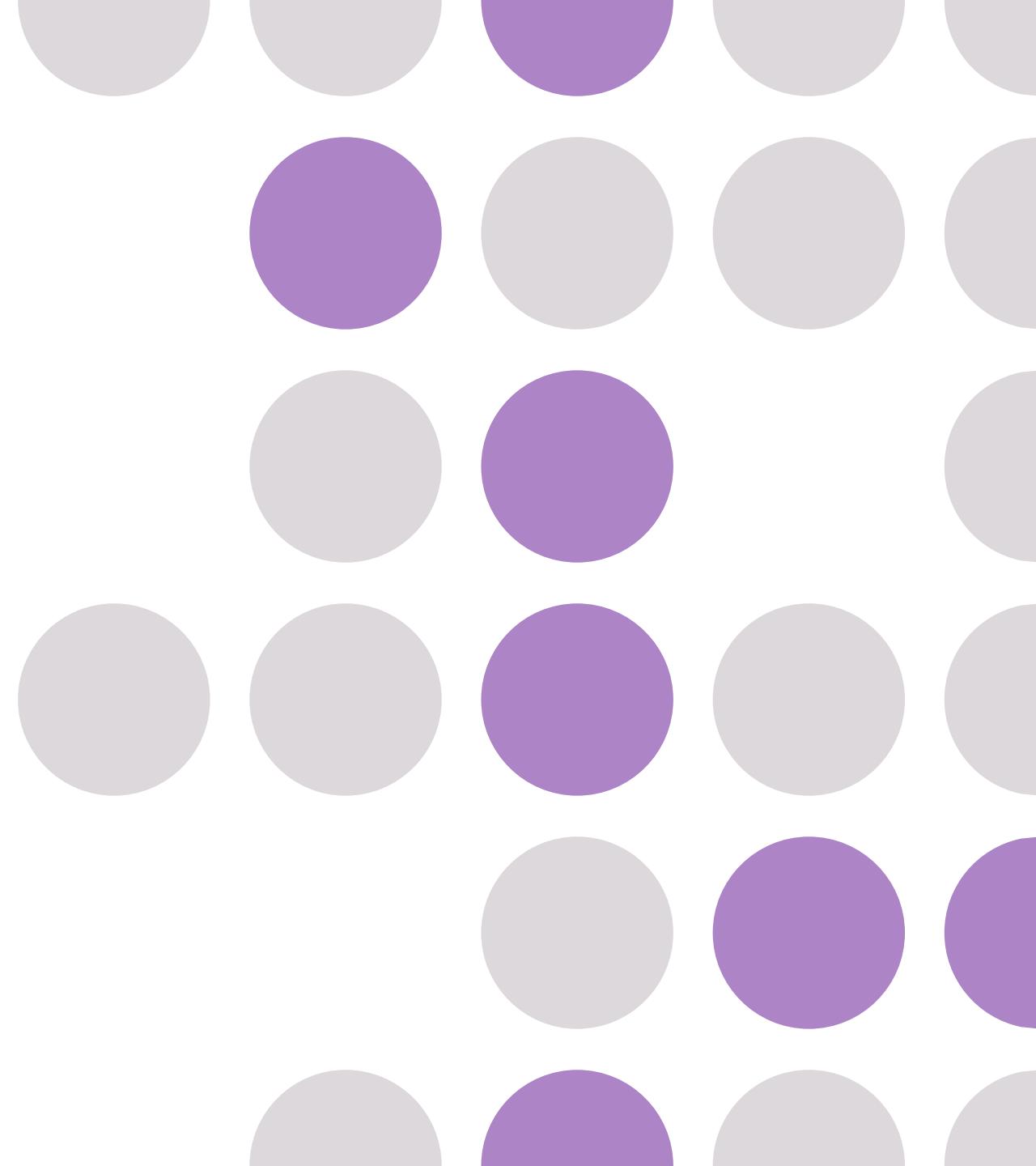
Neural Machine Translation

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

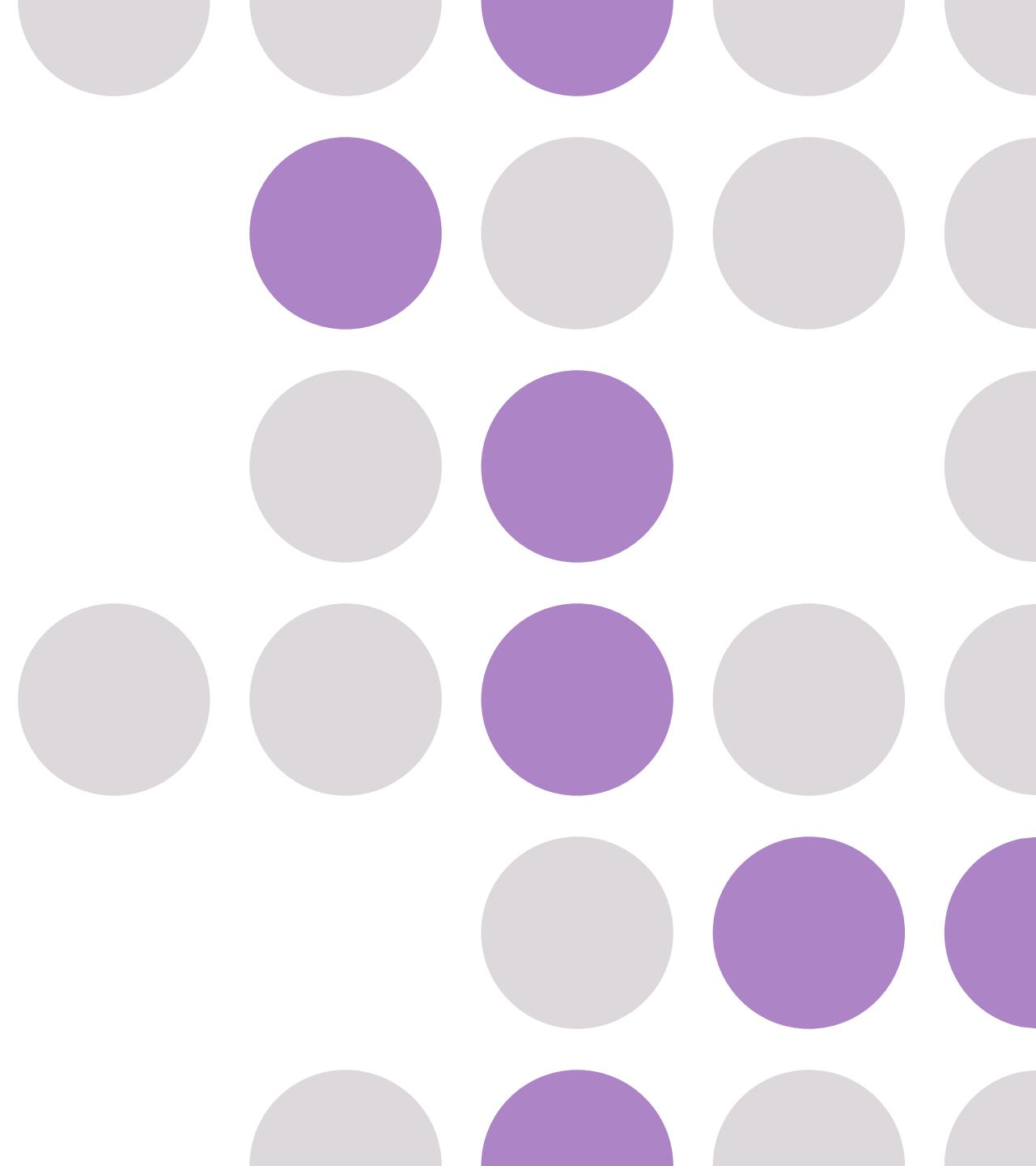
Decoder-based Large Language Models

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

AI in Specialized Translation



Case Study 1: Translation of Financial Documents



Financial translation is integral to multiple sectors in HK's vibrant economy.



Banking

Investment
management

Insurance



Real estate

Accounting
and auditing

Legal
services

Role of financial translation in HK's economy



Significance of financial translation in HK



Enhanced Communication

Facilitates cross-language interactions among investors, clients, and stakeholders



Transparency

Ensures accurate understanding of financial reports and contracts, fostering informed investment and trust in international transactions



Legal Compliance

Aids in maintaining compliance with local and international financial laws, reducing legal risk



Global Market Reach

Enables effective engagement with international clients, increasing potential for foreign investment

Key financial documents



1. Regulatory Documents

Mandatory filings for regulatory bodies (e.g., IPO prospectuses)



2. Company Annual Reports

Detailed annual overviews of company's performance and activities



3. Corporate Announcements and Circulars

Important updates and news for stakeholders (e.g., major corporate decisions)



4. Market Insights & Financial News

Investment analyses and financial news articles



5. Financial Product Marketing Materials

Promotional content for financial products (e.g., product brochures, newsletters, websites)



6. Legal and Contractual Documents

Legal agreements, contracts, terms and conditions, and disclaimers in financial operations

Challenges of financial translation



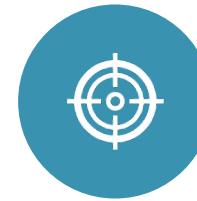
Terminology

Ensuring accurate translation of industry jargon



Domain Knowledge

Mastering diverse financial domains



Compliance & Accuracy

Familiarizing with regulations to ensure precision



Confidentiality

Protecting sensitive information



Time Constraints

Managing speed and accuracy



Dynamic Source Text

Translating updated or non-finalized text while maintaining format

Current tools for financial translators

Specialized
dictionaries

Electronic
databases

Machine
translation

Document
comparison
tools

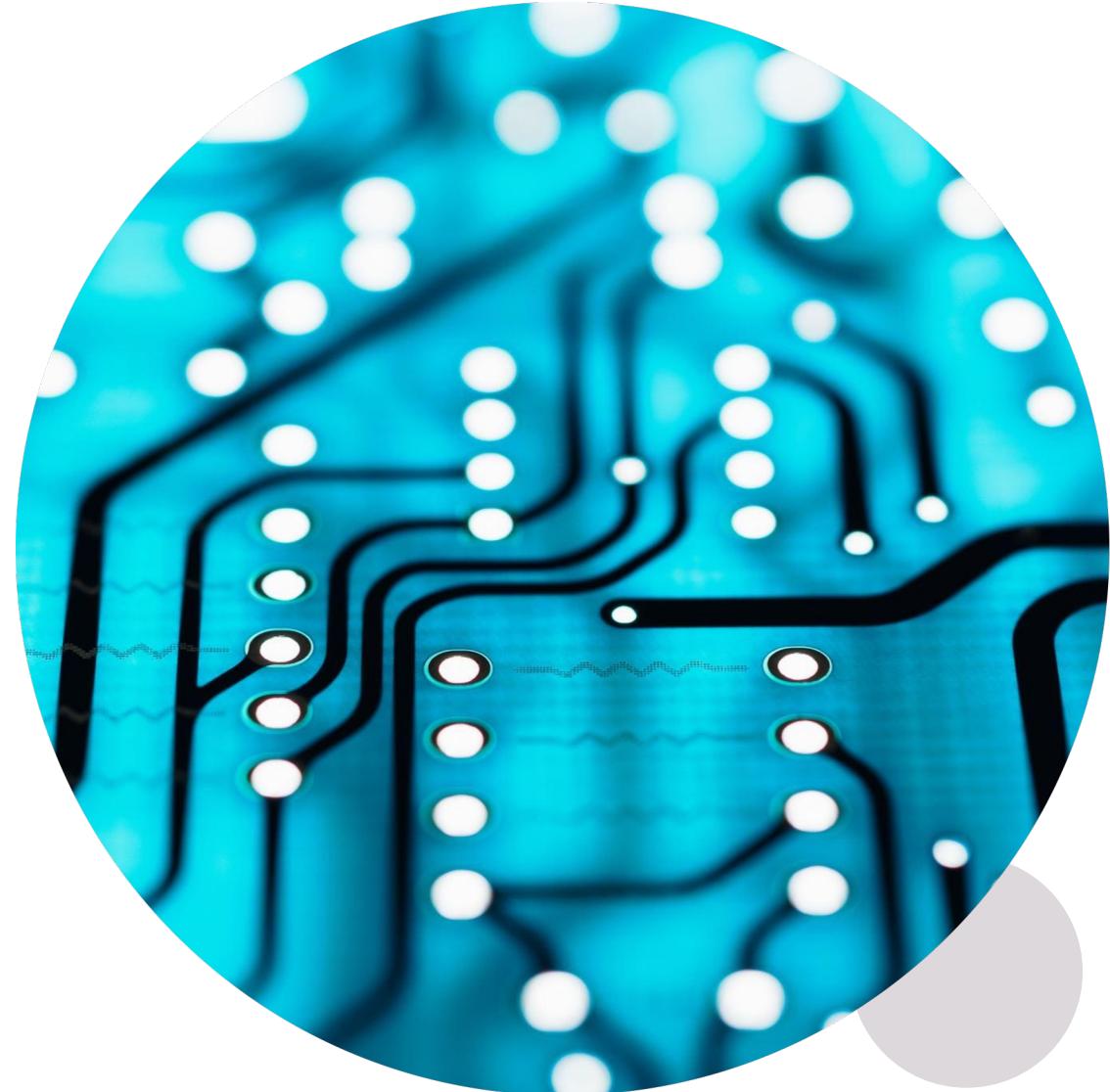
Translation
memories

Terminology
databases

Quality
assurance
tools

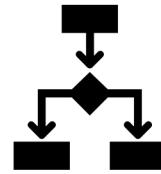
Platform 1: Term Extractor

- **Objective:** Our primary goal is to accurately and efficiently extract relevant terms from bilingual documents, thereby achieving **AI-assisted term extraction and termbase building**.
- **Translation Enhancement:** Our system aims to enhance overall translation quality by improving the consistency, accuracy, and efficiency of term translation.

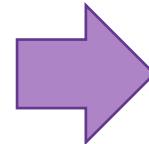


Conventional methods

Linguistic methods



Grammar and syntax rules;
POS tagging



Statistical methods



Term frequency and document
frequency



Term Length:
Define minimum and maximum length of terms, typically in number of words (e.g., 1-4 words)



Frequency and Statistical Thresholds: Set a minimum occurrence count for a term to be considered relevant (e.g., at least 5 times) and establish thresholds for statistical measures like TF-IDF (e.g., TF-IDF score above a specific threshold)



Part-of-Speech Patterns: Specify patterns for linguistic methods (e.g., looking for noun phrases or verb phrases)



Dictionary or Stopword List: Provide a list of known terms to seek or common words to ignore

Current challenges (1)

Term Length

- Some relevant terms or phrases may exceed the user-specified length.

Frequency-Based Limitations

- Less frequent but highly relevant terms may be overlooked, while common but irrelevant words may be incorrectly identified.

Overlapping Terms

- Extraction may result in overlapping terms, requiring manual filtering (e.g., "heart", "heart disease", and "coronary heart disease").

Current challenges (2)

Linguistic Rules Complexity

- Constructing comprehensive and accurate linguistic rules can be challenging due to the complexity of natural language.

Statistical Method Limitations

- These methods may struggle with words that have multiple meanings, rare words, and may also require substantial computational resources.

Bi-Text Size

- If the text is too short, the accuracy of term extraction may be compromised.

Term Extractor

By Sai Cheong SIU (siusaicheong@gmail.com)

Provide your English and Chinese texts and then click **Go** to extract bilingual terms in English and Chinese.

Your Text in English

Hong Kong is a major international financial centre, comprising an integrated network of institutions and markets which provide a wide range of products and services to local and international customers and investors. Hong Kong's financial markets operate under effective and transparent regulations that are in line with international standards. Another feature is high liquidity in the markets. The Government of the Hong Kong Special Administrative Region (HKSAR) abides by the principle of keeping intervention into the way in which the market operates to a minimum and has endeavoured to provide a favourable environment in which business operates. Its policy of low and simple taxation allows maximum room for business initiatives and innovation. There is a strong emphasis on the rule of law and fair market. There are no barriers of access to the market by foreign businesses, with free flow of capital and information. There are also no exchange controls.

Your Text in Chinese

香港是主要的國際金融中心，金融機構和市場緊密聯繫，為本地和海外的客戶及投資者提供各類投資產品及服務。香港金融市場的在有效和具透明度的監管下運作，各項監管規例都與國際標準一致。另一特點是資金流動性高。香港特別行政區（香港特區）政府恪守盡量不干預金融市場運作的原則，並盡力提供一個有利營商的環境。政府實施低稅政策和推行簡單的稅制，使各類商業有更多主動權及創新空間。香港十分重視法治及維持市場的公平競爭，不會阻止外國公司參與本地的金融市場，資訊及資金自由流動。此外，本港亦無實施外匯管制。

Go

international financial centre

國際金融中心

financial institutions

金融機構

financial markets

金融市場

products and services

產品及服務

local and international customers

本地和海外的客戶

customers and investors

客戶及投資者

effective and transparent regulations

有效和透明的監管

Example

- Hong Kong is a major international financial centre, comprising an integrated network of institutions and markets which provide a wide range of products and services to local and international customers and investors. Hong Kong's financial markets operate under effective and transparent regulations that are in line with international standards. Another feature is high liquidity in the markets. The Government of the Hong Kong Special Administrative Region (HKSAR) abides by the principle of keeping intervention into the way in which the market operates to a minimum and has endeavoured to provide a favourable environment in which business operates. Its policy of low and simple taxation allows maximum room for business initiatives and innovation. There is a strong emphasis on the rule of law and fair market. There are no barriers of access to the market by foreign businesses, with free flow of capital and information. There are also no exchange controls.
- 香港是主要的國際金融中心，金融機構和市場緊密聯繫，為本地和海外的客戶及投資者提供各類投資產品及服務。香港金融市場的在有效和具透明度的監管下運作，各項監管規例都與國際標準一致。另一特點是資金流動性高。香港特別行政區（香港特區）政府恪守盡量不干预金融市場運作的原則，並盡力提供一個有利營商的環境。政府實施低稅政策和推行簡單的稅制，使各類商業有多主動權及創新空間。香港十分重視法治及維持市場的公平競爭，不會阻止外國公司參與本地的金融市場，資訊及資金自由流動。此外，本港亦無實施外匯管制。

An online tool

Configurations: Min. frequency: 1; length range: 1-8

English	Chinese
Hong Kong	香港
high liquidity	資金流動性
business	營商
foreign businesses	外國公司參與本地
free flow	自由流動
products	產品
financial markets	金融市
markets	市場
the principle	原則
Hong Kong's	香港金融市場
simple taxation	簡單的稅制
business initiatives	商業
institutions	機構

Our tool

17 terms were extracted.

No extra configuration is required from the user's side.

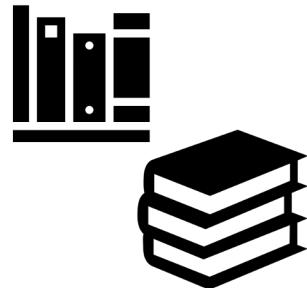
More accurate and useful expressions of varying lengths were extracted

international financial centre	國際金融中心
financial institutions	金融機構
financial markets	金融市場
products and services	產品及服務
local and international customers	本地和海外的客戶
customers and investors	客戶及投資者
effective and transparent regulations	有效和透明的監管
international standards	國際標準
high liquidity	資金流動性高
Government of the Hong Kong Special Administrative Region	香港特別行政區政府
principle of keeping intervention into the way in which the market operates to a minimum	盡量不干預金融市場運作的原則
policy of low and simple taxation	低稅政策
maximum room for business initiatives and innovation	更多主動權及創新空間
rule of law	法治
fair market	公平競爭
free flow of capital and information	資訊及資金自由流動
exchange controls	外匯管制

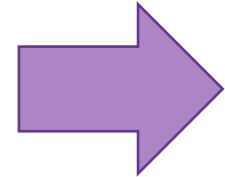
Platform 2: Text Explainer

- This tool helps explain financial text, both in terms of terminology and the overall meaning or ideas within the text.
- It complements dictionaries and glossaries.
- It is particularly useful when dealing with complex financial documents.

Current challenges



First point of reference:
Dictionaries and glossaries



Focus on Words and Short Phrases

- They provide limited help to users struggling with understanding larger units, long phrases, and multiple expressions.

Lack of Context-Awareness

- For words with multiple meanings, they only provide a list of definitions. Users must manually select the appropriate meaning by going through a long list and considering the context.

Solution offered by our tool



Contextual Explanation: Our tool provides explanations of expressions with reference to their context, automatically providing the definition most relevant to the context.



Support for Larger Units: Our tool offers support beyond individual words, including explanations of longer phrases and even sentences.

Text Explainer

By Sai Cheong SIU (siusaicheong@gmail.com)

It's a smart dictionary, featuring
(1) context awareness and (2)
support for longer expressions

Provide your text below and click Explain to get explanation.

Expression to be explained (leave blank for full-text explanation)

tokenised green bonds

Your Text

It has been reported that the Hong Kong Monetary Authority (HKMA) plans to issue, in this quarter, an inaugural batch of tokenised green bonds which, if issued smoothly, will be the first batch of government-issued tokenised green bonds in the world. HKMA has indicated that tokenised green bonds can reduce the costs and time in the issuance process, and can let green bond holders know clearly information such as the progress and daily reduction in carbon emissions of the projects financed.

Check

In the given financial text, "tokenised green bonds" refers to a new type of financial instrument that is designed to support environmentally friendly projects. These bonds are issued in the form of tokens, which are digital representations of value that can be traded on a blockchain platform. The use of tokenised green bonds is intended to make the issuance process more efficient and cost-effective. By using blockchain technology, the HKMA can reduce the time and costs associated with traditional bond issuance. Additionally, tokenised green bonds allow bondholders to track the progress and daily reduction in carbon emissions of the projects that are financed by the bonds. Overall, the term "tokenised green bonds" refers to a new type of financial instrument that is designed to support environmentally friendly projects and make the issuance process more efficient and cost-effective.

Examples

- **Example 1:** Explaining a common word within a given sentence
- **Example 2:** Explaining a multi-word expression containing a technical term, using a given sentence for context
- **Example 3:** Explaining a given sentence within the context of a paragraph

Example 1

Task: Explain a common word within a given sentence.

Our word here: “**sound**”

Consider this ...

Which one?

Noun

- S: (n) sound (the particular auditory effect produced by a given cause) "the sound of rain on the roof"; "the beautiful sound of music"
- S: (n) sound, auditory sensation (the subjective sensation of hearing something) "he strained to hear the faint sounds"
- S: (n) sound (mechanical vibrations transmitted by an elastic medium) "falling trees make a sound in the forest even when no one is there to hear them"
- S: (n) sound (the sudden occurrence of an audible event) "the sound awakened them"
- S: (n) audio, sound (the audible part of a transmitted signal) "they always raise the audio for commercials"
- S: (n) phone, speech sound, sound ((phonetics) an individual sound unit of speech without concern as to whether or not it is a phoneme of some language)
- S: (n) strait, sound (a narrow channel of the sea joining two larger bodies of water)
- S: (n) sound (a large ocean inlet or deep bay) "the main body of the sound ran parallel to the coast"

Verb

- S: (v) sound (appear in a certain way) "This sounds interesting"
- S: (v) sound, go (make a certain noise or sound) "She went 'Mmmmm"'; "The gun went 'bang'"
- S: (v) sound (give off a certain sound or sounds) "This record sounds scratchy"
- S: (v) sound (announce by means of a sound) "sound the alarm"
- S: (v) voice, sound, vocalize, vocalise (utter with vibrating vocal chords)
- S: (v) sound (cause to sound) "sound the bell"; "sound a certain note"
- S: (v) fathom, sound (measure the depth of (a body of water) with a sounding line)

Adjective

- S: (adj) sound (financially secure and safe) "sound investments"; "a sound economy"
- S: (adj) healthy, intelligent, levelheaded, level-headed, sound (exercising or showing good judgment) "healthy scepticism"; "a healthy fear of rattlesnakes"; "the healthy attitude of French laws"; "healthy relations between labor and management"; "an intelligent solution"; "a sound approach to the problem"; "sound advice"; "no sound explanation for his decision"
- S: (adj) sound (in good condition; free from defect or damage or decay) "a sound timber"; "the wall is sound"; "a sound foundation"
- S: (adj) good, sound (in excellent physical condition) "good teeth"; "I still have one good leg"; "a sound mind in a sound body"
- S: (adj) reasoned, sound, well-grounded (logically valid) "a sound argument"
- S: (adj) legal, sound, effectual (having legal efficacy or force) "a sound title to the property"
- S: (adj) sound (vigorous or severe) "a sound thrashing"
- S: (adj) sound (free from moral defect) "a man of sound character"
- S: (adj) heavy, profound, sound, wakeless ((of sleep) deep and complete) "a heavy sleep"; "fell into a profound sleep"; "a sound sleeper"; "deep wakeless sleep"
- S: (adj) sound (complete; thorough) "politicians have no sound grasp of technology issues"; "he has a sound understanding of the law"

**Now, we
have an
alternative ...**

Sound (1)

Expression to be explained (leave blank for full-text explanation)

sound

Context: In addition, the LCSD performance venues are procuring by phases 4K digital light processing projector, 4K pan–tilt–zoom camera, camera remote controller and related digital equipment for live streaming, immersive effect **sound** system, digital sound console and LED moving lights to support visual art creations of arts groups that incorporate arts tech as well as live streaming or video recording.

Your Text

In addition, the LCSD performance venues are procuring by phases 4K digital light processing projector, 4K pan–tilt–zoom camera, camera remote controller and related digital equipment for live streaming, immersive effect sound system, digital sound console and LED moving lights to support visual art creations of arts groups that incorporate arts tech as well as live streaming or video recording.

Check

In the given financial text, the term "sound" refers to the audio equipment that will be procured for live streaming and immersive effect sound system. The sound system will be used to support visual art creations of arts groups that incorporate arts tech as well as live streaming or video recording. Therefore, the term "sound" in this context refers to the audio equipment that will be used to produce and manage the sound for these artistic performances.

Explanation: In the given financial text, the term "sound" refers to the audio equipment that will be procured for live streaming and immersive effect sound system. The sound system will be used to support visual art creations of arts groups that incorporate arts tech as well as live streaming or video recording. Therefore, the term "sound" in this context refers to **the audio equipment** that will be used to produce and manage the sound for these artistic performances.

Sound (2)

Expression to be explained (leave blank for full-text explanation)

sound

Your Text

Context: Hong Kong is among the most advanced information and communication technology hubs in the world. Our well-established business support services and **sound** system to protect intellectual property (IP) rights provide favourable conditions for the growth of start-ups. Many public and private organisations offer support, including provision of co-working space, incubation programmes, business advisory services, and assistance in building partner networks.

Hong Kong is among the most advanced information and communication technology hubs in the world. Our well-established business support services and sound system to protect intellectual property (IP) rights provide favourable conditions for the growth of start-ups. Many public and private organisations offer support, including provision of co-working space, incubation programmes, business advisory services, and assistance in building partner networks.

Check

In the given financial text, the term "sound" is used to describe the system for protecting intellectual property (IP) rights in Hong Kong. The text states that the city has a well-established business support system and a "sound system" to protect IP rights, which provides favorable conditions for the growth of start-ups. This implies that the system for protecting IP rights in Hong Kong is effective and reliable, providing a stable and supportive environment for businesses to thrive.

Explanation: In the given financial text, the term "sound" is used to describe the system for protecting intellectual property (IP) rights in Hong Kong. The text states that the city has a well-established business support system and a "sound system" to protect IP rights, which provides favorable conditions for the growth of start-ups. This implies that the system for protecting IP rights in Hong Kong is **effective and reliable**, providing a stable and supportive environment for businesses to thrive.

Example 2

Task: Explain a multi-word expression containing technical terms, using a given sentence for context.

Expression to be explained (leave blank for full-text explanation)

pre-revenue/pre-profit biotechnology companies

Your Text

Input: In terms of financing, Hong Kong has implemented a new listing regime since end-April 2018 to facilitate the listing of **pre-revenue/pre-profit biotechnology companies** in Hong Kong. A total of 31 pre-revenue / pre-profit biotechnology companies have listed in Hong Kong under the new regime so far. The Corporate Venture Fund of HKSTPC has also invested in biotechnological start-ups engaging in drug delivery, stem cell technology and cancer treatment research, etc.

In terms of financing, Hong Kong has implemented a new listing regime since end-April 2018 to facilitate the listing of pre-revenue/pre-profit biotechnology companies in Hong Kong. A total of 31 pre-revenue / pre-profit biotechnology companies have listed in Hong Kong under the new regime so far. The Corporate Venture Fund of HKSTPC has also invested in biotechnological start-ups engaging in drug delivery, stem cell technology and cancer treatment research, etc.

Check

The expression "pre-revenue/pre-profit biotechnology companies" refers to companies in the biotechnology industry that have not yet generated revenue or have not yet turned a profit. These companies are typically in the early stages of development and are focused on researching and developing new products or technologies. The new listing regime implemented in Hong Kong is designed to make it easier for these companies to list

types of companies

Explanation: The expression "pre-revenue/pre-profit biotechnology companies" refers to companies in the biotechnology industry that have not yet generated revenue or have not yet turned a profit. **These companies are typically in the early stages of development and are focused on researching and developing new products or technologies. The new listing regime implemented in Hong Kong is designed to make it easier for these companies to list on the stock market, as they may not have the financial statements or track record required for traditional listings ...**

Example 3

Task: Explain a given sentence within the context of a paragraph.

Input:

"Hong Kong is a leading player in green finance. **In 2022, green and sustainable debts arranged or issued in Hong Kong amounted to over US\$80.5 billion**," Mr Hui said, adding that the US\$101 million tokenised green bond issued by the Hong Kong Special Administrative Region Government in February this year, the first of its kind released in the world by a government, could further entrench the awareness of sustainability issues in combination with technology.

Expression to be explained (leave blank for full-text explanation)

In 2022, green and sustainable debts arranged or issued in Hong Kong amounted to over US\$80.5 billion

Your Text

"Hong Kong is a leading player in green finance. In 2022, green and sustainable debts arranged or issued in Hong Kong amounted to over US\$80.5 billion," Mr Hui said, adding that the US\$101 million tokenised green bond issued by the Hong Kong Special Administrative Region Government in February this year, the first of its kind released in the world by a government, could further entrench the awareness of sustainability issues in combination with technology.

Check

The expression "In 2022, green and sustainable debts arranged or issued in Hong Kong amounted to over US\$80.5 billion" is a statement of fact that provides information about the amount of green and sustainable debt that was arranged or issued in Hong Kong in the year 2022. The context in which this expression is used is in the context of green finance, which refers to financial activities that support environmentally and socially responsible investment. The statement highlights Hong Kong's role as a leading player in green finance and provides an example of the growth of green and sustainable debt in the city. The reference to the US\$101 million tokenised green bond issued by the Hong Kong Special Administrative Region Government is also used to illustrate the use of technology in green finance.

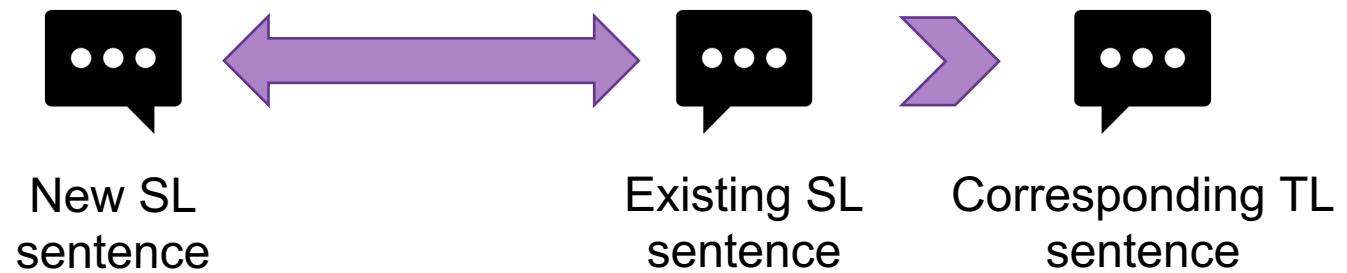
1. The context in which the given sentence is used.
2. Explanation of "green finance."
3. Connection of the given sentence with other parts of the text.

Platform 4: Sentence Finder

- This tool assists translators in identifying **previously translated or similar sentences or expressions** for reference and reuse. This may help reduce the need for retranslating every aspect of the content.
- It could also **facilitate the alignment of terminology and translation style** with those found in previously translated content.

Conventional form of TM

1. string similarity



2. Aligned bilingual sentences

Current Problems with TMs

- **Semantic vs. Syntactic Similarity**
- Conventional methods often depend on string similarity in terms of edit distance. However, **syntactically similar sentences may not be semantically relevant.**
- **Example:** "The company's stock price rose" and "The company's stock price fell" are syntactically similar but semantically opposite.

Current Problems with TMs

- On the other hand, semantically relevant expressions can be presented in different syntactic structures.
- **Example:** "The company's stock price plummeted" and "The company experienced a significant drop in its stock price" convey similar meanings but are syntactically different.

Current Problems with TMs

- **Data Requirements and Limitations**
- Conventional methods typically demand **bilingual aligned data** and often struggle to efficiently retrieve or use monolingual data.
- Monolingual data, which may be more readily available, offer examples originally written in the target language and serve as **valuable reference points** for translation.

Current Problems with TMs

- **Example:** A monolingual financial report could provide idiomatic language and industry-specific terminology that are authentic to the target language, offering valuable reference points for translation.

Features



Calculation of embedding similarity



Cross-lingual retrieval of monolingual and bilingual content without the need for alignment

Sentence Finder

By Sai Cheong SIU (siusaicheong@gmail.com)

Type your text below and then click **Explore!** to find useful sentences in Chinese.

Your Text

The company's stock price fell.

Explore!

Candidate 1

Source: 公司股價下跌。

Target: 公司股價下跌。

Score: 96.9

Candidate 2

Source: The company's stock price plummeted.

Target: 公司股價暴跌。

Score: 96.3

Candidate 3

Source: The company experienced a significant drop in its stock price.

Target: 該公司的股價大幅下跌。

Score: 79.7

Candidate 4

Source: The company's stock price rose.

Target: 公司股價上漲。

Score: 77.8

Example

- **Input:** The company's stock price fell.
- **Conventional method based on string similarity:**

Candidate	Score
The company's stock price rose.*	83.9%
The company's stock price plummeted. *	75.7%
The company experienced a significant drop in its stock price.*	32.3%

*Note: The corresponding Chinese translation is omitted here.

Example

- **Input:** The company's stock price fell.
- **Our Output:**

Candidate	Score	
The company's stock price rose.*	77.8%	Lower! TM: 83.9%
The company's stock price plummeted. *	96.3%	Higher! TM: 75.7%
The company experienced a significant drop in its stock price.*	79.7%	Higher! TM: 32.3%
公司股價下跌。	96.9%	New! TM: N/A

*Note: The corresponding Chinese translation is omitted here.

1. More reasonable scores!

2. Cross-lingual retrieval of content

Platform 4: IPOTranslate

“Machine Translation of IPO Documents”

The work described here was fully supported by the Innovation Technology Fund (ITF/238/17).



The need for bilingual IPO prospectuses

- An IPO prospectus is a comprehensive document that outlines the details of a company's financial health, operations, and plans.
- In Hong Kong, an IPO prospectus must be available in both English and Chinese to cater to the diverse investor base and comply with regulatory requirements.
- This has led to a high demand for professional translation services to ensure accurate and timely translations.

Limitations of current MT systems

- Machine translation systems may provide a solution, but many of them are designed for the **translation of general documents** rather than listing documents.
- The **terms and writing style** in automatic translation results often do not conform to those adopted in IPO prospectuses, requiring substantial editing.

Our solution: IPOTranslate

- We propose a specialized NMT system designed for the translation of IPO prospectuses, with the following two features:
- **Feature 1:** Domain-specific engines trained on both general and specialized documents.
- **Feature 2:** Dual-engine output based on two models with different network architectures, providing alternative translations for reference.



IPOTranslate

Technology for Translation of IPO Prospectuses

What's IPOTranslate?

IPOTranslate is a web-based platform for computer (-aided) translation of prospectuses for Initial Public Offerings (IPOs) from English into Chinese. The system is the product of a machine translation project supported by the Innovation Technology Fund (Project title: Machine Translation of IPO Documents (ITF/238/17)), which was completed in May 2019 and aimed to explore the possibilities of supporting the translation of IPO documents by applying deep learning technology.

IPO Translate (English-Chinese Translation)

Source Text

Hong Kong Exchanges and Clearing Limited, The Stock Exchange of Hong Kong Limited and Hong Kong Securities Clearing Company Limited take no responsibility for the contents of this Prospectus, make no representation as to its accuracy or completeness, and expressly disclaim any liability whatsoever for any loss howsoever arising from or in reliance upon the whole or any part of the contents of this Prospectus.

A copy of this Prospectus, having attached thereto the documents specified in 'Appendix V — Documents Delivered to the Registrar of Companies and Available for Inspection' to this Prospectus, has been registered by the Registrar of Companies in Hong Kong as required by section 342C of the Companies (Winding Up and Miscellaneous Provisions) Ordinance (Chapter 32 of the Laws of Hong Kong). The Securities and Futures Commission and the Registrar of Companies in Hong Kong take no responsibility for the contents of this Prospectus or any other document referred to above.

[Load Sample](#) [Translate](#)

Hong Kong Exchanges and Clearing Limited, The Stock Exchange of Hong Kong Limited and Hong Kong Securities Clearing Company Limited take no responsibility for the contents of this Prospectus, make no representation as to its accuracy or completeness, and expressly disclaim any liability whatsoever for any loss howsoever arising from or in reliance upon the whole or any part of the contents of this Prospectus.



IPO Translate (English-Chinese Translation)

Source Text

Hong Kong Exchanges and Clearing Limited, The Stock Exchange of Hong Kong Limited and Hong Kong Securities Clearing Company Limited take no responsibility for the contents of this Prospectus, make no representation as to its accuracy or completeness, and expressly disclaim any liability whatsoever for any loss howsoever arising from or in reliance upon the whole or any part of the contents of this Prospectus.

A copy of this Prospectus, having attached thereto the documents specified in "Appendix V — Documents Delivered to the Registrar of Companies and Available for Inspection" to this Prospectus, has been registered by the Registrar of Companies in Hong Kong as required by section 342C of the Companies (Winding Up and Miscellaneous Provisions) Ordinance (Chapter 32 of the Laws of Hong Kong). The Securities and Futures Commission and the Registrar of Companies in Hong Kong take no responsibility for the contents of this Prospectus or any other document referred to above.

Load Sample

Translate

Hong Kong Exchanges and Clearing Limited, The Stock Exchange of Hong Kong Limited and Hong Kong Securities Clearing Company Limited take no responsibility for the contents of this Prospectus, make no representation as to its accuracy or completeness, and expressly disclaim any liability whatsoever for any loss howsoever arising from or in reliance upon the whole or any part of the contents of this Prospectus.

Translation 1

香港交易及結算有限公司、香港聯合交易所有限公司及香港證券結算有限公司就本招股章程的內容概不承擔任何責任，概不就其準確性或完整性表示任何聲明，並明確概不就本招股章程所載的全部或任何部分產生的任何虧損而產生的任何損失承擔責任。

Translation 2

香港交易及結算有限公司、香港聯合交易所有限公司及香港中央結算有限公司對本招股章程的內容概不負責，對其準確性或完整性概不發表任何聲明，並明確表示概不就因本招股章程全部或任何部分內容而產生或因依賴該等內容而產生的任何損失承擔任何責任。

A copy of this Prospectus, having attached thereto the documents specified in "Appendix V — Documents Delivered to the Registrar of Companies and Available for Inspection" to this Prospectus, has been registered by the Registrar of Companies in Hong Kong as required by section 342C of the Companies (Winding Up and Miscellaneous Provisions) Ordinance (Chapter 32 of the Laws of Hong Kong).

Translation 1

本招股章程的副本連同本招股章程「附錄五一送呈公司註冊處處長及備查文件」所載文件所附的文件，已根據公司（清盤及雜項條文）條例第342C 條（香港法例第32章）的規定於香港公司登記。

Translation 2

本招股章程副本連同本招股章程「附錄五一送呈公司註冊處處長及備查文件」所述文件，已根據公司（清盤及雜項條文）條例（香港法例第32章）第342C 條的規定由香港公司註冊處處長登記。

The Securities and Futures Commission and the Registrar of Companies in Hong Kong take no responsibility for the contents of this Prospectus or any other document referred to above.

Translation 1

證券及期貨事務監察委員會及香港公司註冊處處長概不就本招股章程或上文所述任何其他文件的內容負責。

Translation 2

證券及期貨事務監察委員會及香港公司註冊處處長對本招股章程或上述任何其他文件的內容概不負責。

Example 1

Source Text

You should rely only on the information contained in this prospectus and the related Application Forms to make your investment decision.

Human Translation

閣下應僅依賴本招股章程及相關申請表格所載的資料作出投資決定。

Example 1

A famous online MT engine

您應僅依賴本**招股說明書**中包含的**信息**和相關的申請表來做出**您的**投資決定。

Our Translations

1. 閣下**僅應**依賴本招股章程及相關申請表格所載資料作出投資決定。
2. 閣下應**僅依賴**本招股章程及相關申請表格所載資料作出投資決定。

Example 2

Source Text

Hong Kong Exchanges and Clearing Limited, The Stock Exchange of Hong Kong Limited and Hong Kong Securities Clearing Company Limited take no responsibility for the contents of this prospectus, make no representation as to its accuracy or completeness and expressly disclaim any liability whatsoever for any loss howsoever arising from or in reliance upon the whole or any part of the contents of this prospectus.

Human Translation

香港交易及結算所有限公司、香港聯合交易所有限公司及香港中央結算有限公司對本招股章程的內容概不負責，對其準確性或完整性亦不發表任何聲明，並明確表示概不會就因本招股章程全部或任何部分內容而產生或因依賴該等內容而引致的任何損失承擔任何責任。

A famous online MT engine

香港交易及結算所有限公司，香港聯合交易所有限公司及香港中央結算有限公司對本招股章程的內容概不負責，亦不對其準確性或完整性作出任何陳述，並明確表示不對任何損失承擔任何責任。**因本招股章程的全部或部分內容而產生或依賴該等招股章程的全部或部分內容。**

Our Translations

1. 香港交易及結算有限公司、香港聯合交易所有限公司及**香港證券結算有限公司**就本招股章程的內容概不承擔任何責任，概不就其準確性或完整性作出任何聲明，並明確不會就本招股章程所載的全部或任何部分內容而產生的任何損失產生任何任何責任。**閣下應僅依賴本招股章程及相關申請表格所載資料作出投資決定。**
2. 香港交易及結算所有限公司、香港聯合交易所有限公司及香港中央結算有限公司對本招股章程的內容概不負責，對其準確性或完整性概不發表任何聲明，並明確表示概不就因本招股章程全部或任何部分內容而產生或因依賴該等內容而產生的任何損失承擔任何責任。

Experiment: results

	Google	IPOTranslate Model 1	IPOTranslate Model 2
BLEU Score	36.89	44.48	53.73

$$BLEU = 100 \cdot \min(1, e^{1 - \frac{r}{c}}) \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right)$$

Case Study 2: Translation of Government Documents

Translation of Government Documents in HK

- Government documents in HK are generally issued in both English and Chinese (Traditional Chinese and Simplified Chinese), which are the two official languages of the city under Article 9 of the Basic Law and the Official Languages Ordinance (Cap. 5).

Platform 5: GovPR Translation

- An AI-driven platform for the translation of government press releases

The work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. UGC/FDS14/H16/18).

Our system



The screenshot shows the GovPR Translation Demo (English-Chinese) interface. At the top, there's a logo consisting of two overlapping hexagons forming a diamond shape, followed by the text "GovPR Translation". Below this is a dark teal header section containing the title "GovPR Translation Demo (English-Chinese)" and the text "Version 0.2 (2021)". The main content area has a light gray background. It starts with a paragraph about the translation engine's features, mentioning neural networks trained on general and government documents, and the use of a translation memory. It also notes its usefulness for translators and language professionals. Below this is another paragraph about the demo version, stating it's from 2021 and supports English-Chinese translation, with a note about GPU usage. At the bottom, there's a text input field with placeholder text "Type your English text here:", a "Clear" button, a "Go" button, and a small informational box stating "This platform is designed for the translation of government press releases."

GovPR Translation is an automatic translation engine for the rendition of HKSAR government press releases. It features (1) the development of neural translation networks trained on general documents, partially in-domain documents (government documents), and in-domain documents (press releases between 2016 and 2018) and (2) the use of a translation memory prior to automatic translation to facilitate the processing of expressions with existing translations.

The present work is also useful for [translators and language professionals](#) as it provides useful translation drafts that take into consideration of the writing style of government press releases and common expressions, and this may help reduce post-editing efforts and thus enhance the speed and quality of translation.

The demo provided here is Version 0.2 released in 2021 with support for English-Traditional Chinese translation. As we use a computer without a GPU, it may take longer time to translate your text. Please be patient.

Type your English text here:

This platform is designed for the
translation of government press
releases.

Hong Kong Government Press Releases

- An important communication channel between the government and the public
- Information provided by different bureaus and departments
- Example:
<https://www.info.gov.hk/gia/general/today.htm>
- Three versions: English, Traditional Chinese, and Simplified Chinese

Examples



Announcements



Speeches



Statements



Event promotional
materials

Themes

Immigration

Labour

Urban
Planning

Finance

Law and
Order

Housing
and Land

Recreation

Public
Health

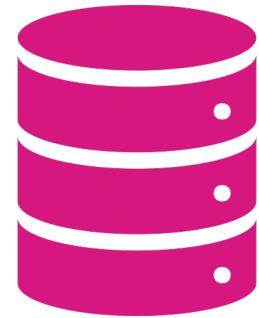


General Machine Translation?

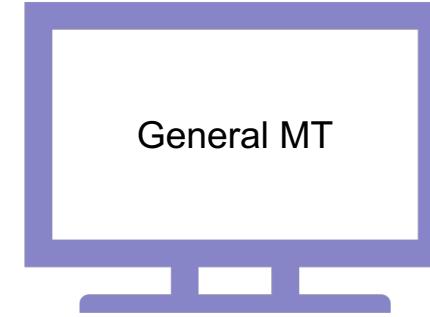
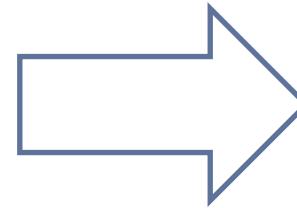
- Word Choices
- Sentence Structure
- Terminology
- Proper Nouns
- Formatting
- Writing Style

Our Approach

- TM + MT



General data



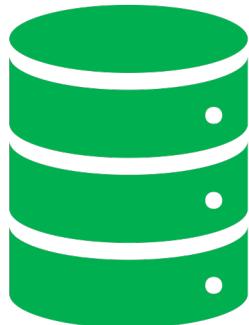
General MT

How?



Domain-specific data

How can we make use of this?



Domain-specific data

A Common Approach

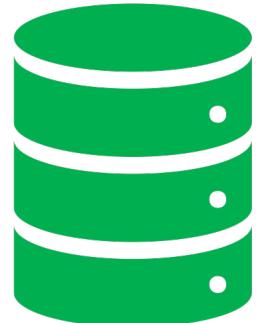
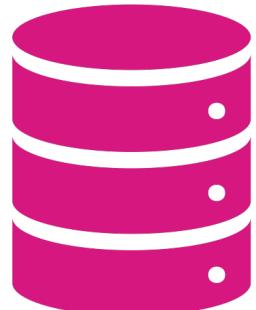


We can use domain-specific data.



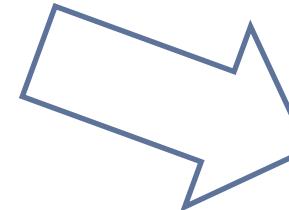
However, this often requires re-training or additional training of the model.

1

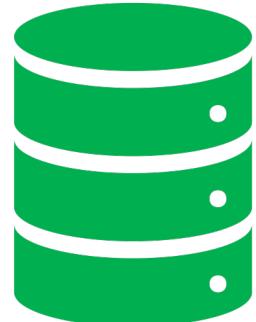


General data

Training from scratch

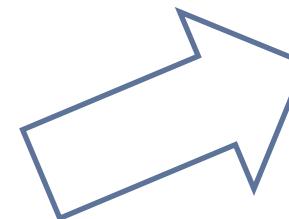


2

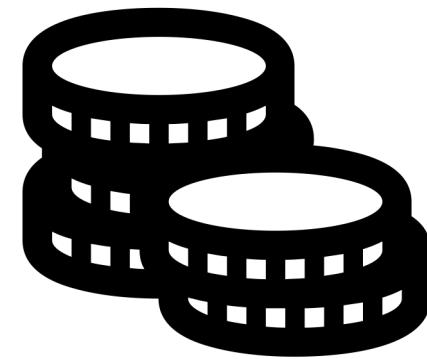
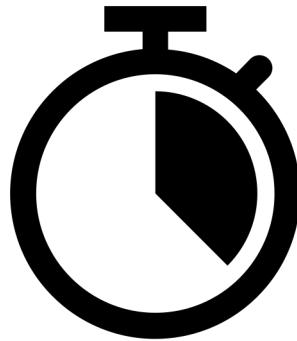
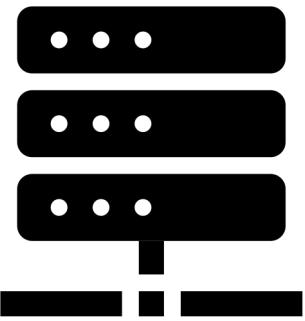


Domain-specific data

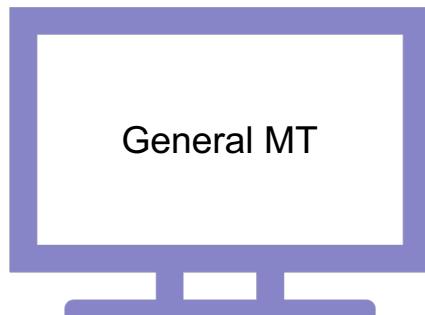
Further Training



(Re-)raining needs

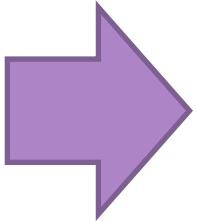
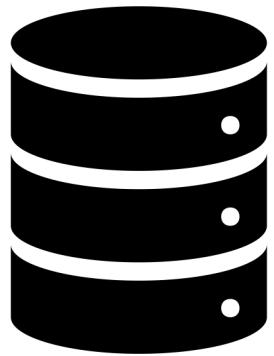


Is it possible to get rid of model (re-)training?



Domain-specific data

Our Approach



Translation Memory

- Designed for MT (rather than HT)
- Containing phrases and sentences
- Facilitating the translation of domain-specific content

Machine Translation

- No re-training or additional training of the model

Our Sample TM

- Bilingual government press releases (2016-2018)
- 408,000 sentence pairs (after automatic alignment)
- 270,000 translation units (after removal of segments with low interlingual similarity)

TM Examples

Sentences	<p>The HKSAR Government attaches great importance to upholding academic freedom and institutional autonomy. 香港特區政府一直致力維護學術自由及院校自主。</p> <p>The Government is determined and committed to enhancing the well-being of elderly people. 此外，政府有決心和承擔為長者謀福祉。</p>
Common expressions	<p>LegCo committee meeting and public hearing 立法會委員會會議及公開聆訊</p> <p>Public urged to stay alert against avian influenza 市民應提高警惕預防禽流感</p> <p>The membership of the Advisory Committee is as follows: 諮詢委員會的成員如下：</p>
Job titles	<p>Secretary for Innovation and Technology 創新及科技局局長</p> <p>Under Secretary for Home Affairs 民政事務局副局長</p> <p>Director-General of Trade and Industry 工業貿易署署長</p>

TM Examples

Names of units	Architectural Services Department 建築署 Business Environment Council 商界環保協會 Sha Tau Kok District Rural Committee 沙頭角區鄉事委員會
Names of policies or initiatives	Qualifying Debt Instrument Scheme 合資格債務票據計劃 Redevelopment of Tai Hang Sai Estate 大坑西邨重建計劃 SME Financing Guarantee Scheme 中小企融資擔保計劃
Other technical terms	Effective Exchange Rate Index 港匯指數 Free Trade Agreements (FTAs) 自貿協定 Green Bond and Green Finance 綠色債券和金融

Experiments

- Experiment 1: A Popular Online Engine
- Experiments 2 and 3: Our Own NMT Models

Experiment 1



Our Test Dataset

- Press releases in January 2019
- 5,000 bilingual sentences
- 156,473 English tokens and 130,646 Chinese tokens

Experiment 1: Results

	Google Translate without pre- translation	Google Translate with pre- translation	BLEU difference after pre- translation
BLEU score based on the test set (BLEU4)	23.73	35.51	+11.78

$$BLEU = 100 \cdot \min(1, e^{1-\frac{r}{c}}) \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

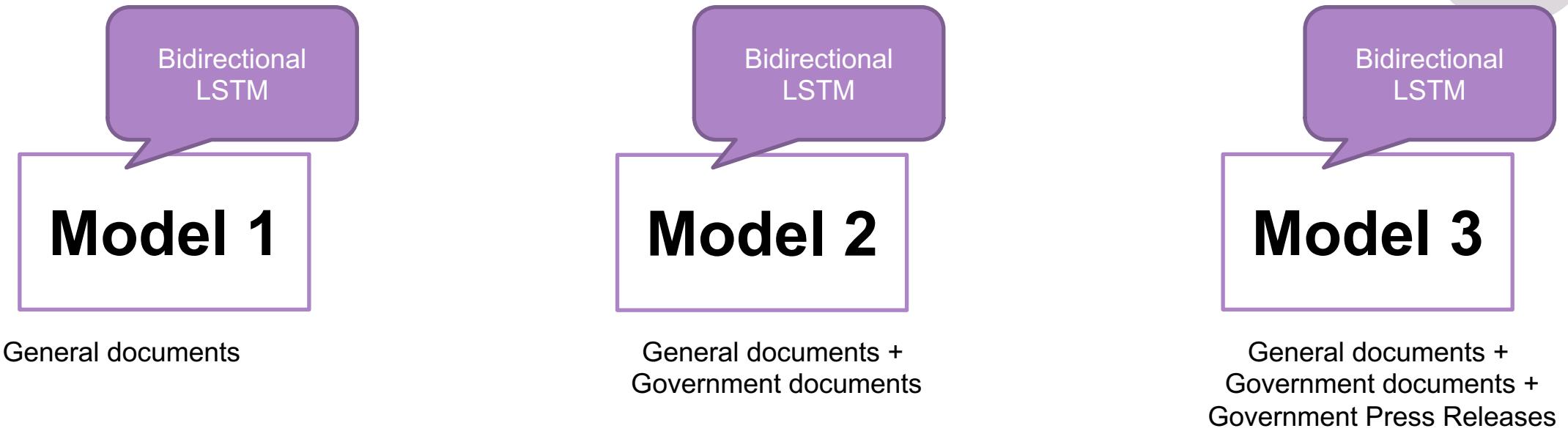
Example 1

- **Source Text:** Investigation by the Special Investigation Team of Traffic, Kowloon East is underway.
- **Reference Translation:** 東九龍總區交通部特別調查隊正跟進調查該宗意外。
- **Output 1 (without pre-translation):** 九龍東交通特別調查組正進行調查。
- **Output 2 (with pre-translation):** 東九龍總區交通部特別調查隊正跟進調查案件。

Example 2

- **Source Text:** Admission to the ward has been suspended and restricted visiting has been imposed.
- **Reference Translation:** 該病房已暫停接收新症，並實施有限度探訪安排。
- **Output 1 (without pre-translation):** 該病房已暫停入場，並已實施限制探視。
- **Output 2 (with pre-translation):** 有關病房已暫停接收新症，並實施有限度探訪安排。

Experiment 2



With/out TM

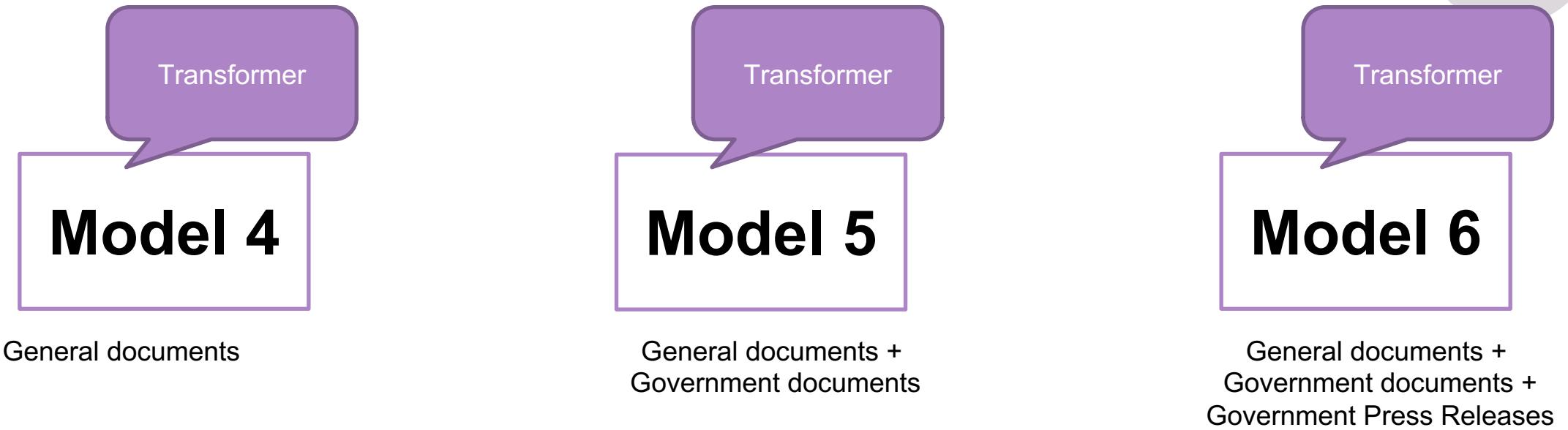
Experiment 2: Results

	BLEU score of model without pre-translation	BLEU score of model with pre-translation	BLEU difference after pre- translation
Model 1	15.56	29.28	+13.72
Model 2	18.09	31.23	+13.14
Model 3	38.75	44.71	+5.96

Experiment 3

- The recent popularity of using the Transformer model (Vaswani et al., 2017) (or its variants) for not only machine translation but also other natural language processing applications (e.g., Devlin et al., 2019) and even image processing (Dosovitskiy et al., 2020) suggests that it would be worthwhile to assess the applicability of our method to NMT engines adopting the self-attention architecture.

Experiment 3



With/out TM

Experiment 3: Results

	BLEU score for model without pre-translation	BLEU score for model with pre-translation	BLEU difference after pre- translation
Model 4	14.40	28.28	+13.88
Model 5	18.56	30.97	+12.41
Model 6	42.86	46.16	+3.3

Platform 6: Useful Expression Finder

- The TM built for GovPR Translation can also be utilized by translators
- With semantic similarity calculation, users can find useful bilingual expressions from TM, including not only sentences, as conventional TMs do, but also smaller units like words and phrases.

Useful Expression Finder (Chinese-English)

Type a Chinese expression below and find bilingual sentences that help you translate the expression.

Input

(Length: 4 char(s))

民心相通

Go

Relevant Bilingual Sentences that may help you translate the expression above

Chinese: 民心相通

English: People-to-people Bonds

Score: 15.629993187340167

Chinese: 促進民心相通

English: Strengthening People-to-people Bond

Score: 15.094909622649421

Chinese: 是次活動能促進滬港連繫、民心相通、文明互鑑和文化交流

English: She pointed out that the Festival can better connect Shanghai and Hong Kong and foster people-to-people bonds, as well as enhance mutual learning and cultural exchange between the two cities.

Score: 12.194874053716761

Chinese: 東西文化交流的平台，在促進民心相通方面擔當重要的角色

English: Hong Kong also offers a platform for cultural exchanges between East and West, and can play a key role in enhancing people-to-people bonds.

Platform 7: BlendMT

- **Diversity of Approaches:** Different engines use distinct algorithms and training data. Merging captures a broader range of linguistic nuances missed by any single model and mitigate individual weaknesses.
- **Enhanced Fluency and Readability:** Blending diverse sentence structures and phrasing often results in a more natural-sounding, human-like final translation.

Platform 7: BlendMT

- **Improved Accuracy for Complex Texts:**
Technical or domain-specific language benefits from cross-referencing multiple interpretations to arrive at the most accurate rendition.
- **Boosting Efficiency & Cost-Effectiveness:**
Instead of extensive manual post-editing, combination tools streamline the process, saving time and resources.

BlendMT: Smarter Translations

Provide Your Machine Translations

MT1 Output:

MT2 Output:

Blend & View

Blend Translations

Example: ST

Text input: 我們會持續優化「互聯互通」機制，強化全球最大離岸人民幣業務樞紐地位，助力人民幣國際化。主要措施包括持續提升基礎建設，升級債務工具中央結算系統以便利國際投資者以不同幣種結算各類資產；提升固定收益市場的基建，例如為人民幣計價的回購交易建立中央清算系統，使在港發行的國債於離岸市場成為更受歡迎的抵押品；及繼續研究優化「跨境理財通」。

Example: MT1 and MT2

MT 1: We will continue to optimize the "interconnection" mechanism, strengthen our position as the world's largest offshore RMB business hub, and promote the internationalization of RMB. The main measures include continuing to improve infrastructure and upgrading the central settlement system for debt instruments to facilitate international investors to settle various assets in different currencies; improving the infrastructure of the fixed income market, such as establishing a central clearing system for RMB-denominated repurchase transactions, so that Government bonds issued in Hong Kong have become more popular collateral in the offshore market; and we will continue to study and optimize the "Cross-border Wealth Management Connect".

MT 2: We will continue to enhance the "interconnection and connectivity" mechanism to strengthen our position as the world's largest offshore RMB business hub and facilitate the internationalization of RMB. Key measures include continuous enhancement of infrastructure, upgrading the Central Moneymarkets Unit to facilitate international investors to settle various assets in different currencies; upgrading the infrastructure of the fixed income market, for example, setting up a central clearing system for RMB-denominated repurchase transactions, so as to make treasury bonds issued in Hong Kong a more popular form of collateral in the offshore market; and continuing to study the enhancement of the "Cross-border Wealth Management System".

Sentence 1:

Text Input	MT1	MT2	Combined Translation	Explanation
我們會持續優化「互聯互通」機制，強化全球最大離岸人民幣業務樞紐地位，助力人民幣國際化。	We will continue to optimize the “interconnection” mechanism, strengthen our position as the world's largest offshore RMB business hub, and promote the internationalization of RMB.	We will continue to enhance the “interconnection and connectivity” mechanism to strengthen our position as the world's largest offshore RMB business hub and facilitate the internationalization of RMB.	We will continue to enhance the “interconnection and connectivity” mechanism to strengthen our position as the world's largest offshore RMB business hub and promote the internationalization of RMB.	MT2 uses “interconnection and connectivity” which is slightly more comprehensive and flows better.

Sentence 2:

Text Input	MT1	MT2	Combined Translation	Explanation
主要措施包括持續提升基礎建設，升級債務工具中央結算系統以便利國際投資者以不同幣種結算各類資產；	The main measures include continuing to improve infrastructure and upgrading the central settlement system for debt instruments to facilitate international investors to settle various assets in different currencies;	Key measures include continuous enhancement of infrastructure, upgrading the Central Moneymarkets Unit to facilitate international investors to settle various assets in different currencies;	Key measures include continuous enhancement of infrastructure, upgrading the central settlement system for debt instruments to facilitate international investors to settle various assets in different currencies;	MT1's wording is clearer and more direct regarding the purpose of the upgrade.

Platform 8: MTCheck

- This is a tool designed to enhance the quality assurance of machine translation output.
- Existing CAT tool QA checkers tend to focus on surface-level errors (grammar, spelling, punctuation).
- This tool offers actionable revisions & insights. It doesn't just *flag* errors, but *provides corrected translations* and concise explanations for suggested changes. This empowers translators to learn and improve.

MTCheck

Advanced Machine Translation Evaluation

1. Your Text

Original Text:

Enter your source content here...

Machine Translation:

Paste the MT output here...

2. Import Terminology (Optional)

Term (Source) - Term (Target) (One pair per line):

Enter terms here...

Analyze MT

Example (ST)

完善新型工業發展策略和體制建設

76. 我們將制訂香港新型工業中長期發展方案，並推動成立「香港新型工業發展聯盟」，促進「政、產、學、研、投」緊密協作，構建香港新型工業合作平台，包括提供融資機會、促進新來港上市企業與本地大學開展創科合作等。

推進第三個「InnoHK創新香港研發平台」

77. InnoHK現匯聚本地和全球各地約2 500名科研人員。政府已開展籌備建設第三個InnoHK研發平台，聚焦先進製造、材料、能源及可持續發展，吸引世界級的科研團隊與本地院校合作，推動科研，匯聚人才。

Example (Sample MT)

Improving the strategy and institutional framework for the development of new industries

76. We will formulate a medium-to-long-term development plan for Hong Kong's new industries and promote the establishment of the Hong Kong Alliance for the Development of New Industries to foster close collaboration among the Government, industry, academia, research and investment, and to build a platform for cooperation among Hong Kong's new industries, including the provision of financing opportunities, and the facilitation of collaboration in innovation and technology between newly-listed enterprises and local universities, etc. We will also promote the establishment of a third InnoHK R&D platform to provide a platform for the development of new industries.

Taking forward the third InnoHK R&D platform

77. InnoHK is currently home to some 250 researchers from local and worldwide. The Government has commenced preparations for the construction of the third InnoHK R&D platform, which focuses on advanced manufacturing, materials, energy and sustainable development, and attracts world-class research teams to collaborate with local institutions to promote R&D and pool talents.

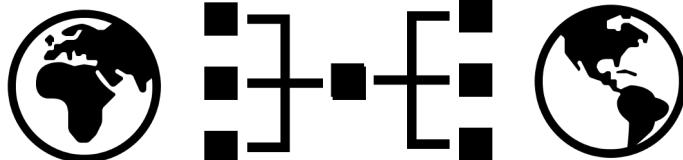
ST Sentence	MT Output	Error Type(s)	Revised Translation	Explanation
完善新型工業發展策略和體制建設	Improving the strategy and institutional framework for the development of new industries	Accuracy (Minor Oversimplification)	Revised: Enhancing and streamlining the strategies and institutional frameworks to foster the development of new industries.	The original MT slightly simplifies the nuanced meaning of “完善” (perfection/improvement). The revised version captures a more comprehensive approach to development.
76. ... (Full ST Sentence from previous response)	We will formulate a medium-to-long-term development plan for Hong Kong's new industries and promote the establishment of the Hong Kong Alliance for the Development of New Industries... (rest of MT)	Clarity Issues (Repetition & Flow), Term Usage	Revised: To cultivate Hong Kong's new industries, we will develop a comprehensive medium-to-long-term plan and establish the Hong Kong New Industrialisation Development Alliance . This alliance will foster collaboration among government, industry, academia, research, and investment, creating a platform that includes financing opportunities and encouraging innovation partnerships between newly listed companies and local universities.	* Combined sentences for smoother flow. * Emphasized the alliance's name using the provided bilingual term. * Rephrased for improved clarity and emphasis on the alliance's role.
77. InnoHK現匯聚本地和全球各地約2 500名科研人員。政府已開展籌備建設第三個InnoHK研發平台，聚焦先進製造、材料、能源及可持續發展，吸引世界級的科研團隊與本地院校合作，推動科研，匯聚人才。	InnoHK is currently home to some 250 researchers from local and worldwide. The Government has commenced preparations for the construction of the third InnoHK R&D platform, which focuses on advanced manufacturing, materials, energy and sustainable development, and attracts world-class research teams to collaborate with local institutions to promote R&D and pool talents.	Number Error , Term Usage	Revised: InnoHK currently hosts approximately 2,500 researchers from Hong Kong and worldwide. Recognizing its success, the government is preparing the third InnoHK Research Cluster . This new platform will concentrate on advanced manufacturing, materials, energy, and sustainable development, attracting top-tier research teams to collaborate with local universities, driving innovation and fostering a talent pool.	* Corrected the researcher count from 250 to the accurate 2,500. * Used the provided bilingual term “InnoHK Research Cluster” for consistency.

The Way Forward

Revolution of Technology for Translation and Localisation
in the AI Era (Sai-cheong Siu, 7 November 2024)

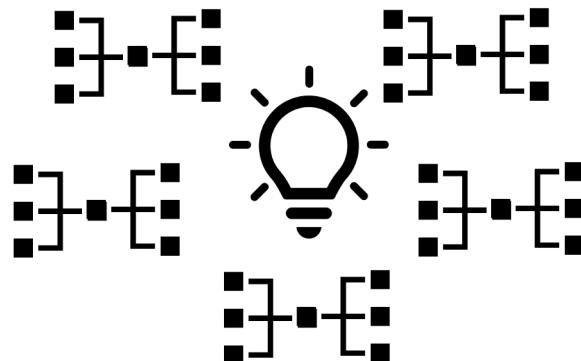
The Way Forward (see Siu 2023b)

Multilingual NMT

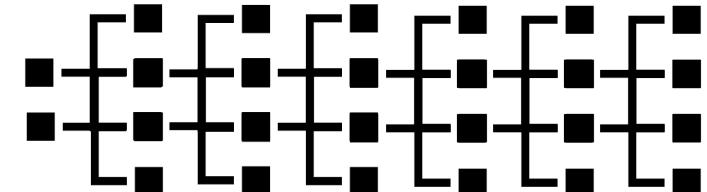


Multiple language pairs

Multiple Models

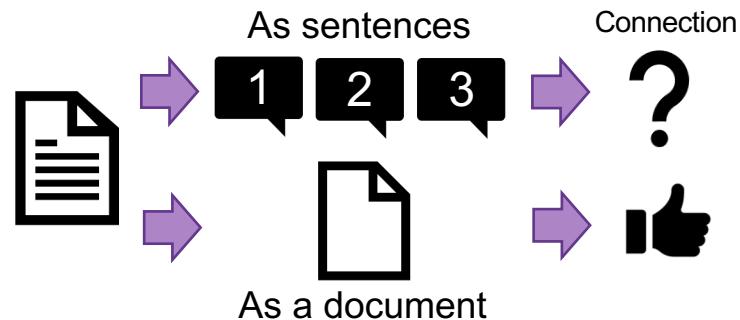


Deeper Models

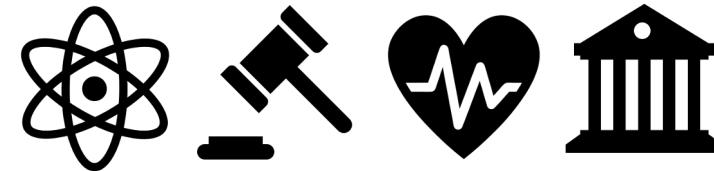


More Layers

Document-level NMT

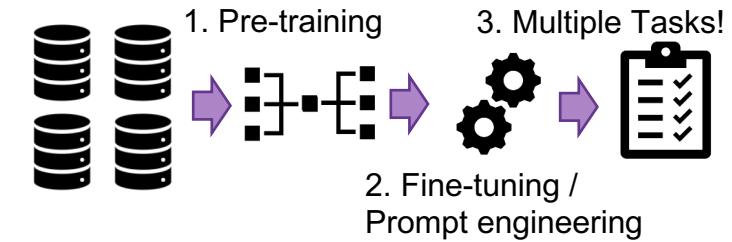


Domain Adaptation



Specialized translation models

Pre-trained language models (e.g., ChatGPT)



Read More

“ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation” (Siu 2023a) [[Link](#)]

- An Overview of ChatGPT and its Predecessors
- ChatGPT as a Tool for Translators: Exploring Use Cases
- Strengths and Limitations of ChatGPT in the Context of Translation
- Impacts of ChatGPT on Translators and Language Professionals
- The Future of ChatGPT as a Translation Tool



Read More

“Deep Learning and Translation Technology” - see Chapter 47, *Routledge Encyclopedia of Translation Technology* (2nd Edition) (Siu 2023b)

- An Introduction to Deep Learning
- Main Tasks of Deep Learning
- An Overview of Neural Machine Translation
- Recurrent Neural Networks for Translation
- Convolutional Neural Networks for Translation
- Self-attention Neural Networks for Translation
- Deep Learning for Other Translation-related Applications
- Other Noteworthy Trends



Read More

“Revolutionising Translation with AI: Unravelling Neural Machine Translation and Generative Pre-trained Large Language Models” (Siu 2023c) [[Link](#)]

- An Overview of Deep Learning
- The Transformer Model Explained
- AI for Translation: NMT and Generative Pre-trained
- Strengths and Limitations of NMT and LLMs
- Recommendations for Translation Educators
- Recommendations for Developers



Thank you!

scsiu@hsu.edu.hk

siusaicheong@gmail.com