



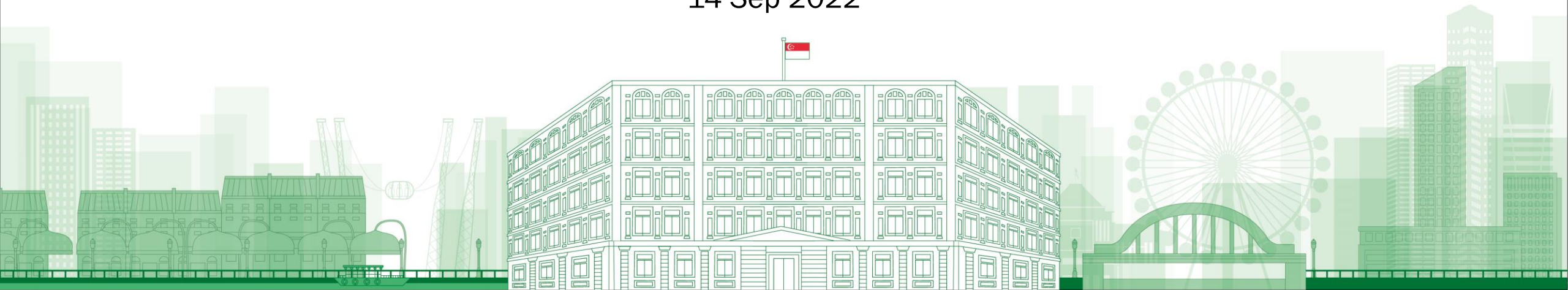
Ministry of Communications
and Information



Uplifting Singapore's Translation Standards with the Community through Technology

Adeline Sim, Ministry of Communications and Information, Singapore
Zheng Weihua, Institute for Infocomm Research (I²R), A*STAR, Singapore

14 Sep 2022



Content



INTRODUCTION



SG TRANSLATE NEURAL MACHINE TRANSLATION ENGINE



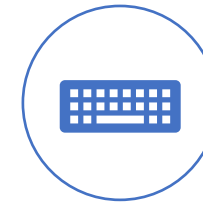
METHODOLOGY

- Data Selection and Augmentation
- Terminology Translation with Fuzzy Match
- Multi-stage Domain Adaptation
- Data Tagging



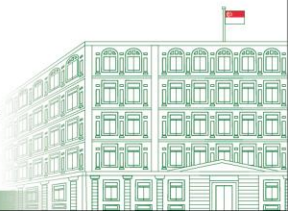
ILLUSTRATIVE PERFORMANCE

- Local Culture
- Daily Life in Singapore
- Government



SG TRANSLATE TOGETHER WEB PORTAL

- Sustainable Approach
- Shaping Singapore's Translation Landscape and Nurturing the Next Generation of Translators



Introduction



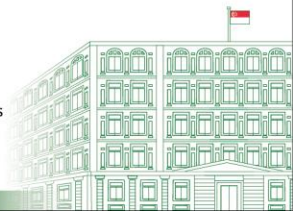
Ministry of Communications
and Information

The Ministry of Communications and Information (MCI) oversees the development of the infocomm technology, cyber security and media sectors in Singapore; the national library, national archives and public libraries; as well as the Singapore Government's information and public communication policies.



Institute for
Infocomm Research
I²R

The Agency for Science, Technology and Research (A*STAR) is Singapore's lead public sector agency that spearheads economic oriented research to advance scientific discovery and develop innovative technology.



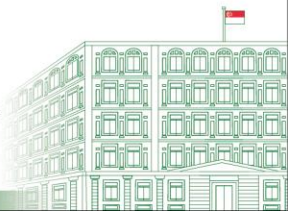
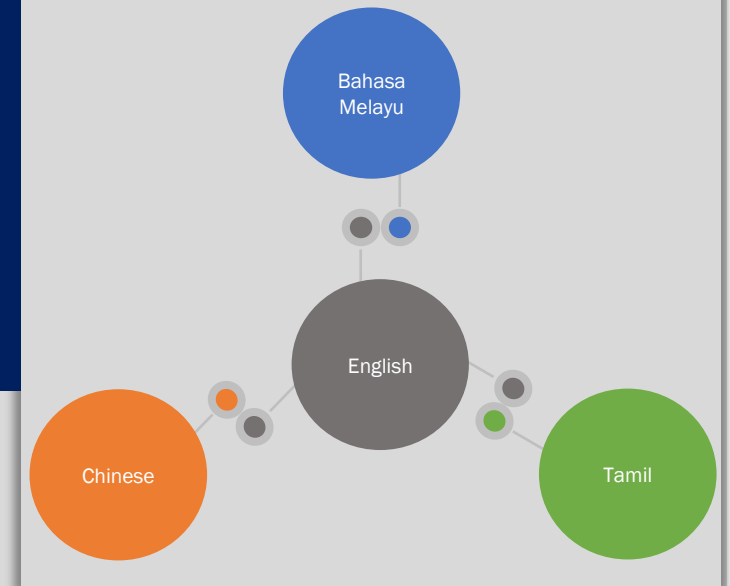
Introduction



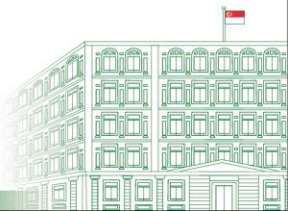
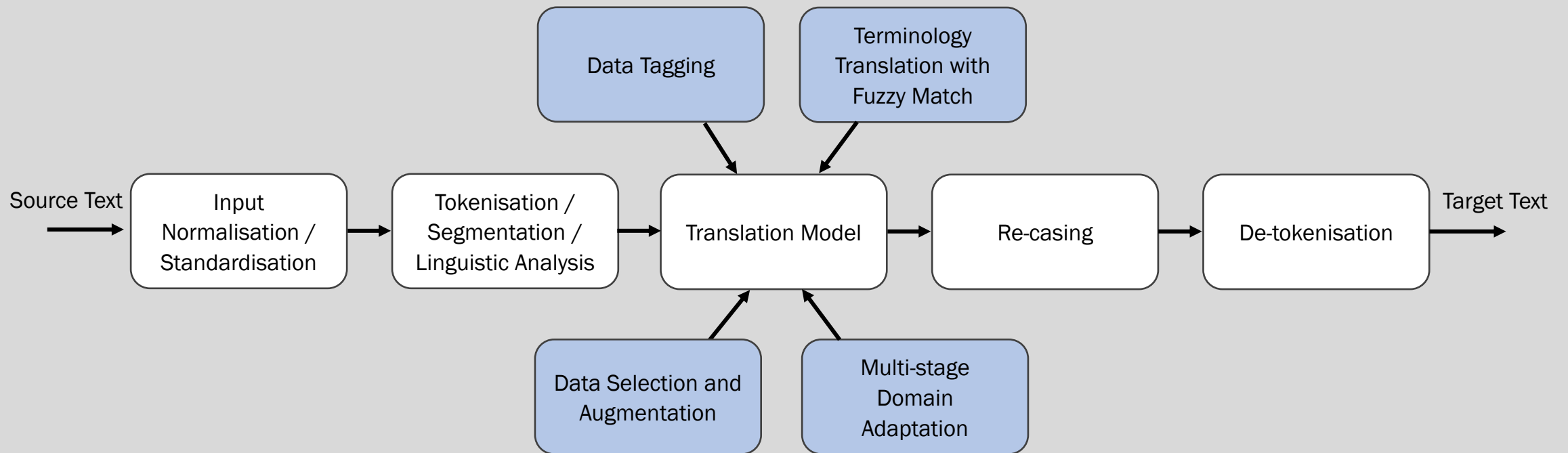
Ministry of Communications
and Information

CUSTOMISED NEURAL MACHINE TRANSLATION (NMT)

- Trained with localised data
- Launched in the Singapore Government Intranet
- Supports Singapore's 4 official languages



SG Translate Neural Machine Translation Engine



Methodology

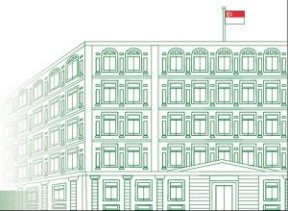
Data Selection
and Augmentation

Multi-stage
Domain Adaptation

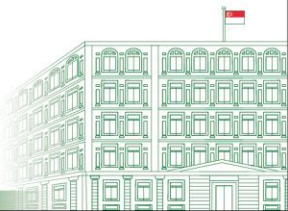
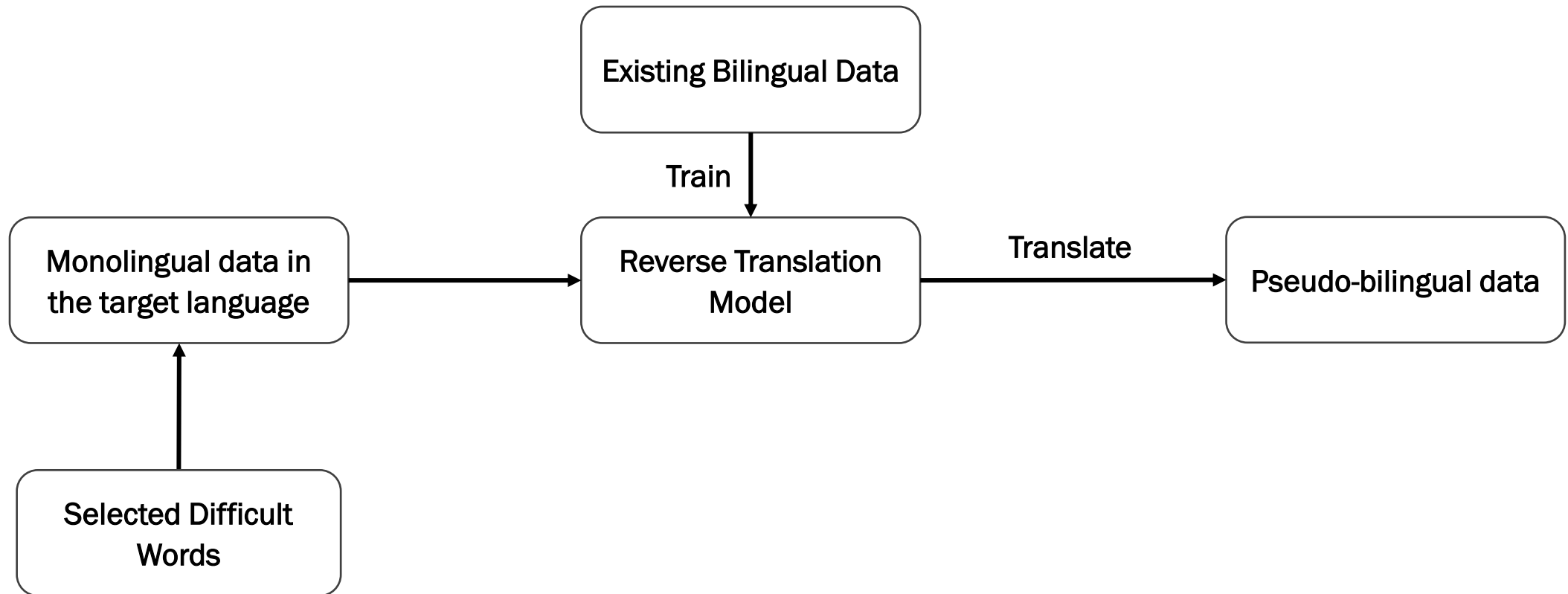


Terminology Translation
with Fuzzy Match

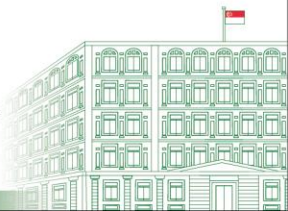
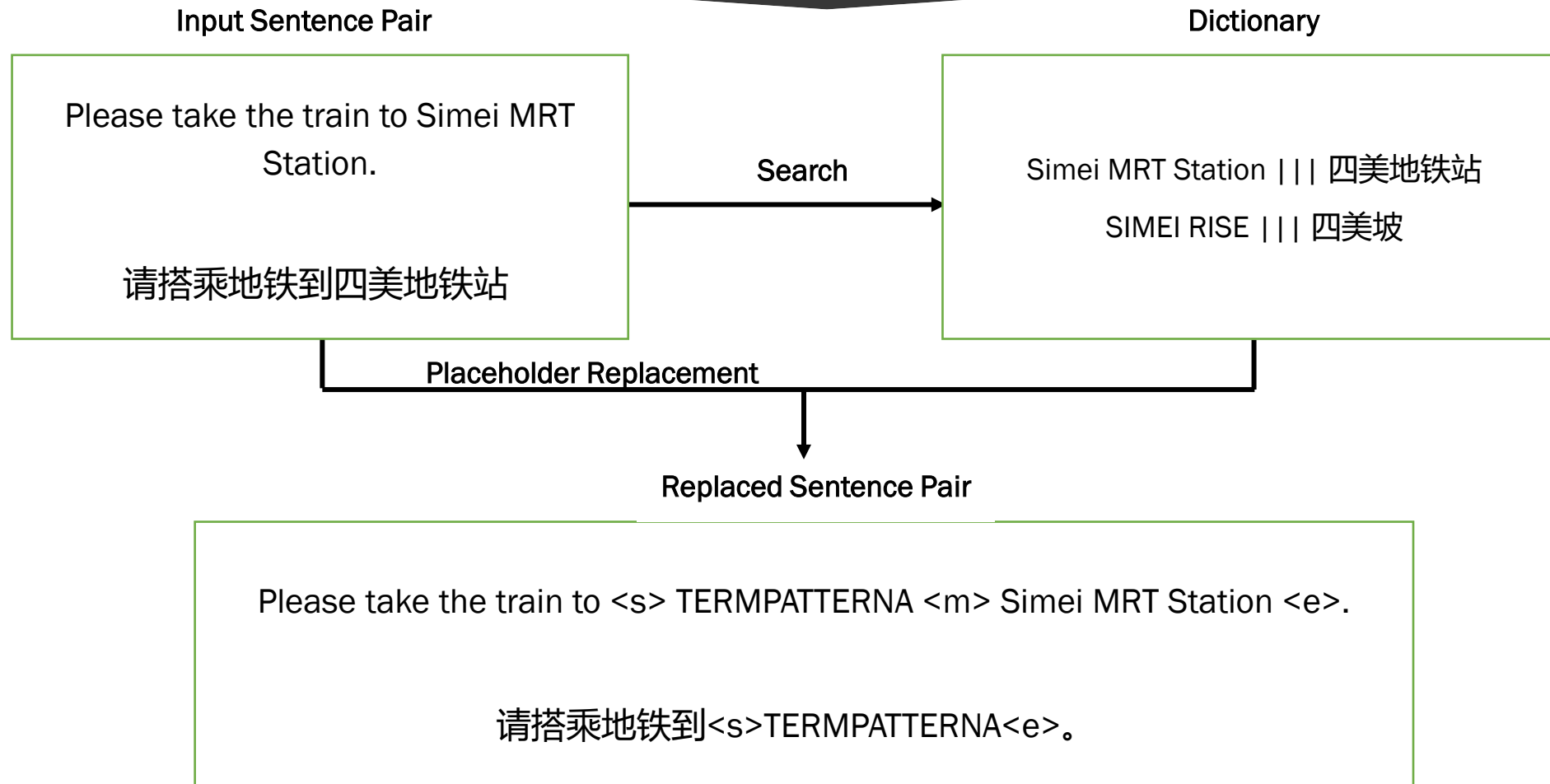
Data Tagging



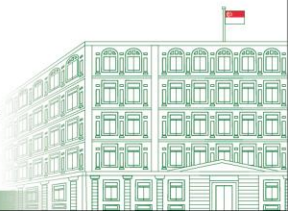
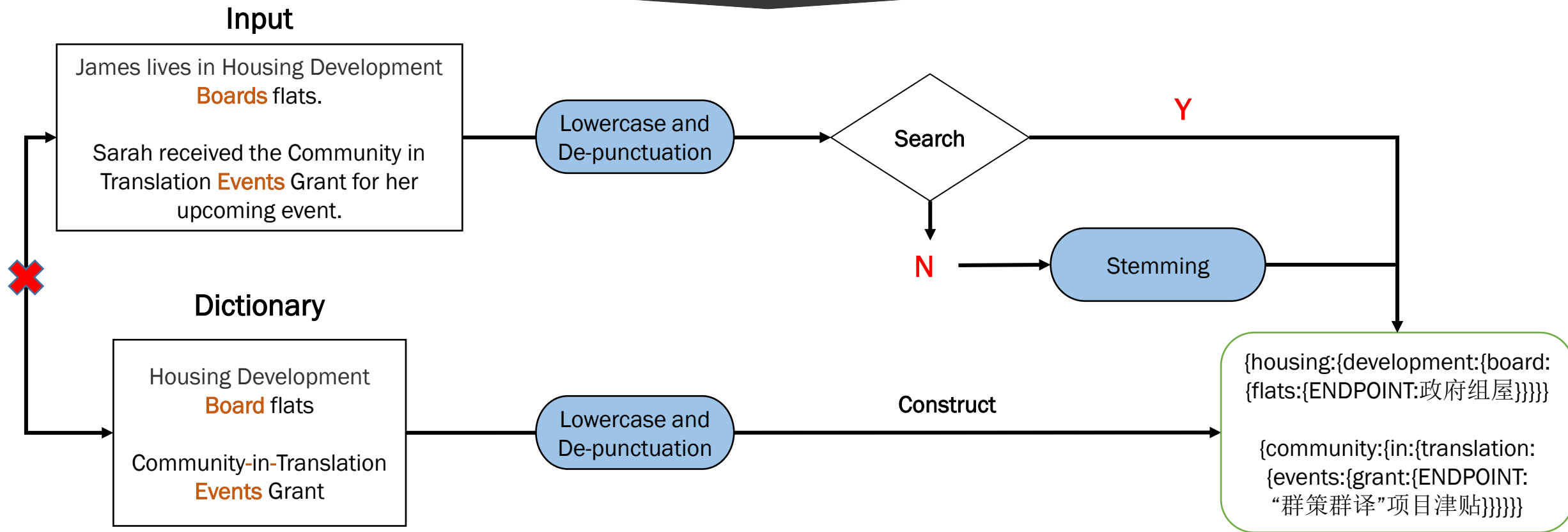
Data Selection and Augmentation



Terminology Translation with Fuzzy Match



Terminology Translation with Fuzzy Match



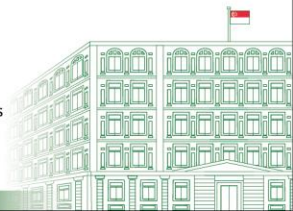
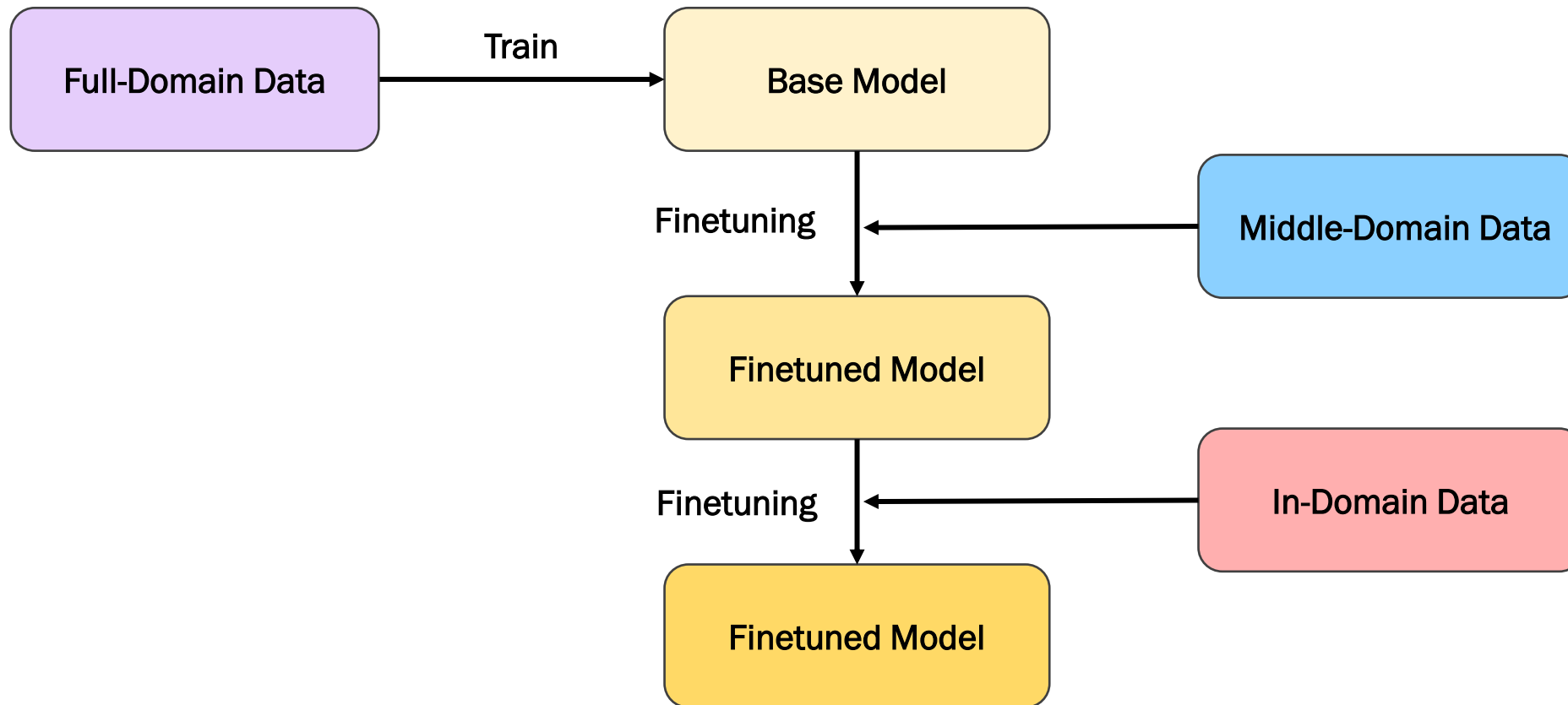
Terminology Translation with Fuzzy Match

Comparison of model performance with and without fuzzy match

Language Pair	BLEU	
	With fuzzy match	Without fuzzy match
English to Tamil (EN2TA)	15.76	13.12
English to Chinese (EN2ZH)	16.55	13.7
English to Malay (EN2MS)	16.93	14.57
Tamil to English (TA2EN)	17.31	15.75
Chinese to English (ZH2EN)	19.77	18.82
Malay to English (MS2EN)	18.03	17.14



Multi-stage Domain Adaptation



Data Tagging

Labels marking the quality of the data are added at the beginning of sentences at the source end of the training data.

Before Data Tagging

请携带你的**新加坡国民身份证**到指定地点登记。

Please bring your NRIC to the designated location for registration.

After Data Tagging

<PA> 请携带你的**新加坡国民身份证**到指定地点登记。

Please bring your NRIC to the designated location for registration.

Before Data Tagging

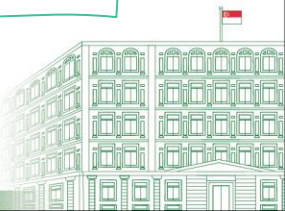
请携带你的**新加坡证件**到指定地点登记。

Please bring your NRIC to the designated location for registration.

After Data Tagging

<BT> 请携带你的**新加坡证件**到指定地点登记。

Please bring your NRIC to the designated location for registration.



Examples of Localised Translation

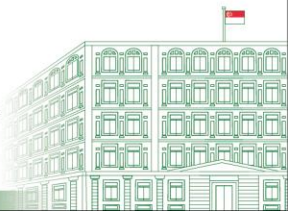
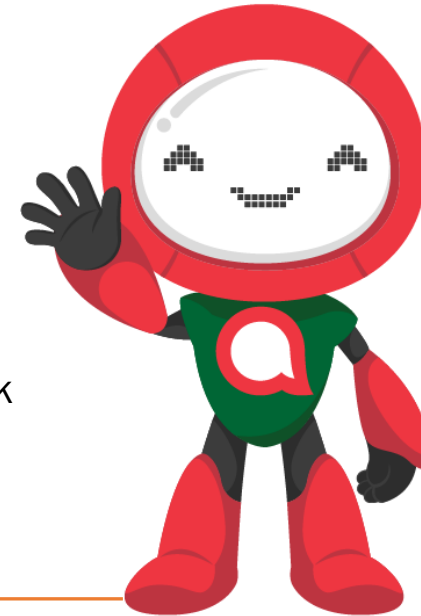
Translation related to local culture

Source (Chinese)

一提到中元节，人们一般上会想到歌台。

Target (English)

When it comes to Hungry Ghost Festival, people generally think of getai.



Examples of Localised Translation

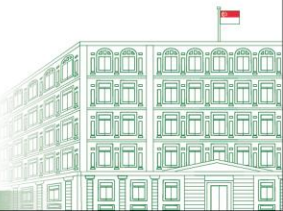
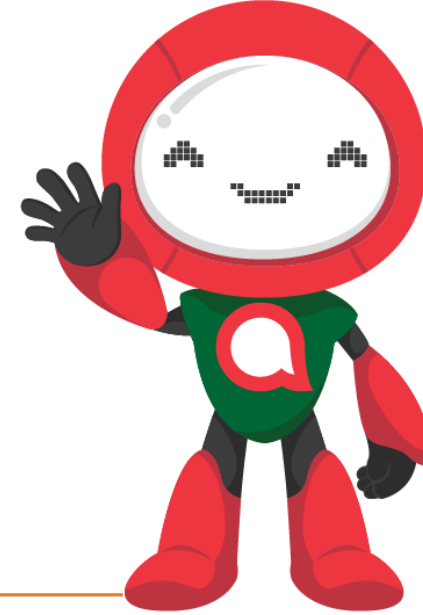
Translation related to everyday life

Source (English)

Bring your NRIC or valid passport, and poll card.

Target (Tamil)

உங்கள் அடையாள அட்டை அல்லது
செல்லுபடியாகும் கடவுச்சீட்டு மற்றும் வாக்கு
அட்டை ஆகியவற்றைக் கொண்டு வாருங்கள்.



Examples of Localised Translation

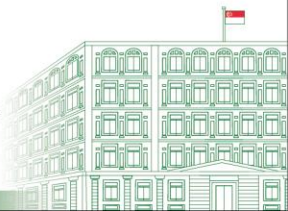
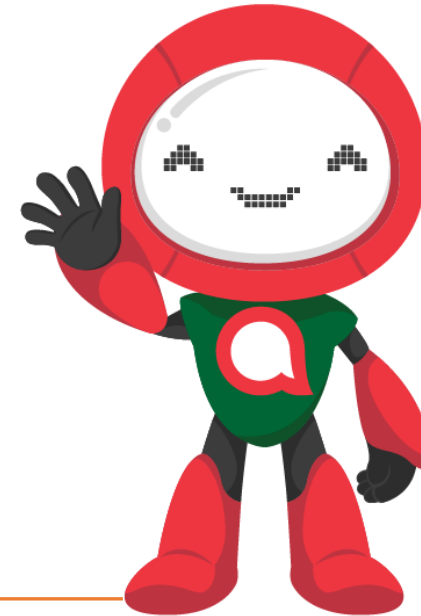
Translation related to Government

Source (Malay)

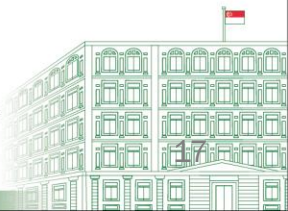
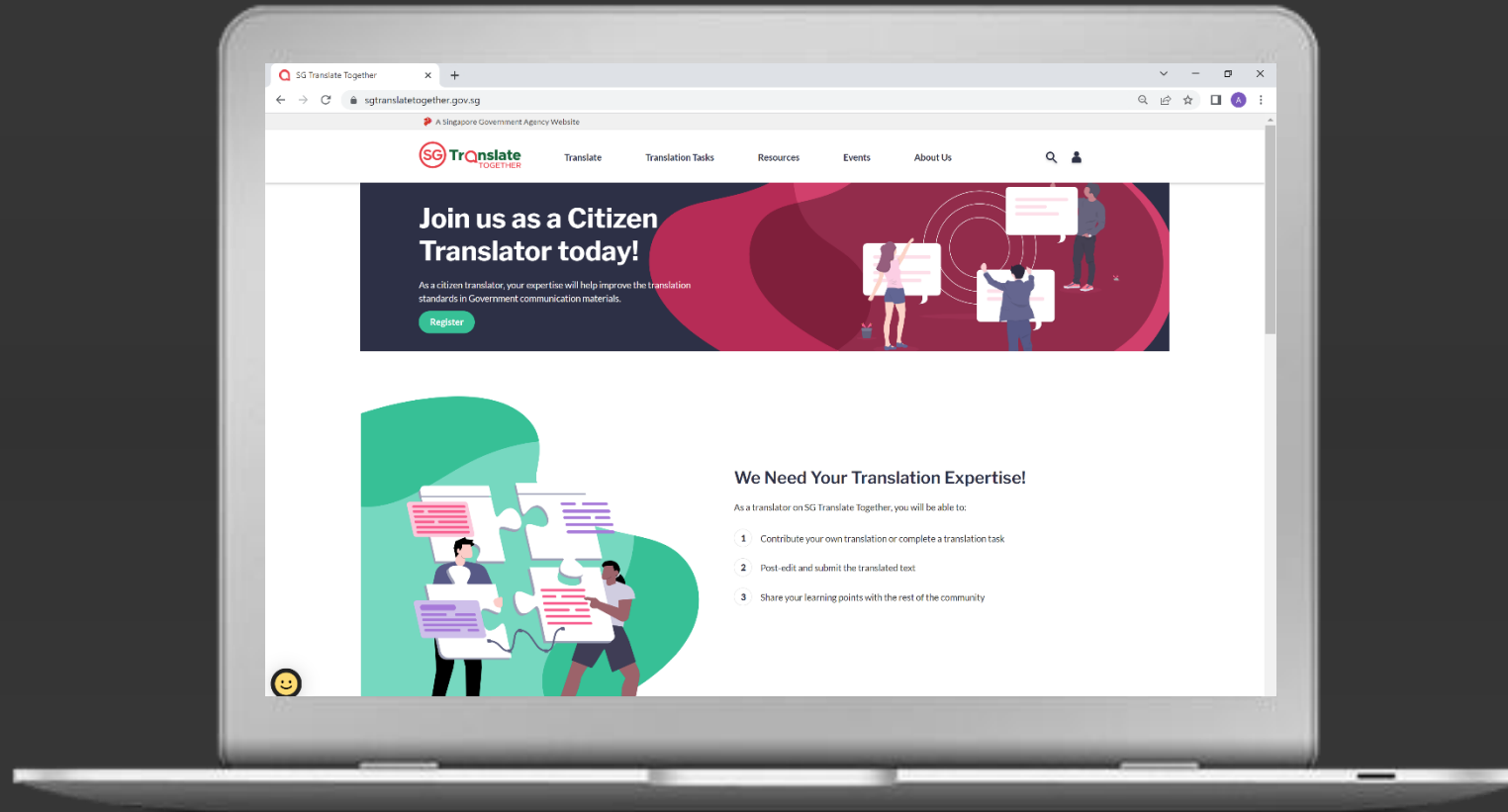
Awak tak tahu ke yang kita semua ada MediShield Hayat?

Target (English)

Don't you know we all have MediShield Life?







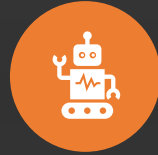
Admin Module

User management
Task management
Terminology Update
Incremental Learning



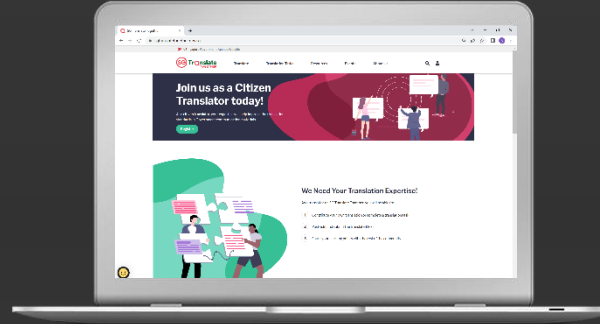
SG Translate

Customised MT engine that is accessible by the public for text-to-text translation



Translation Tasks

Bite-sized translation tasks posted by the Government administrators



Events

Past and upcoming events related to translation



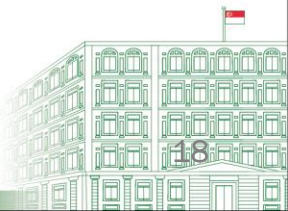
Resources

Useful resources such as glossaries and databases



Community Forum

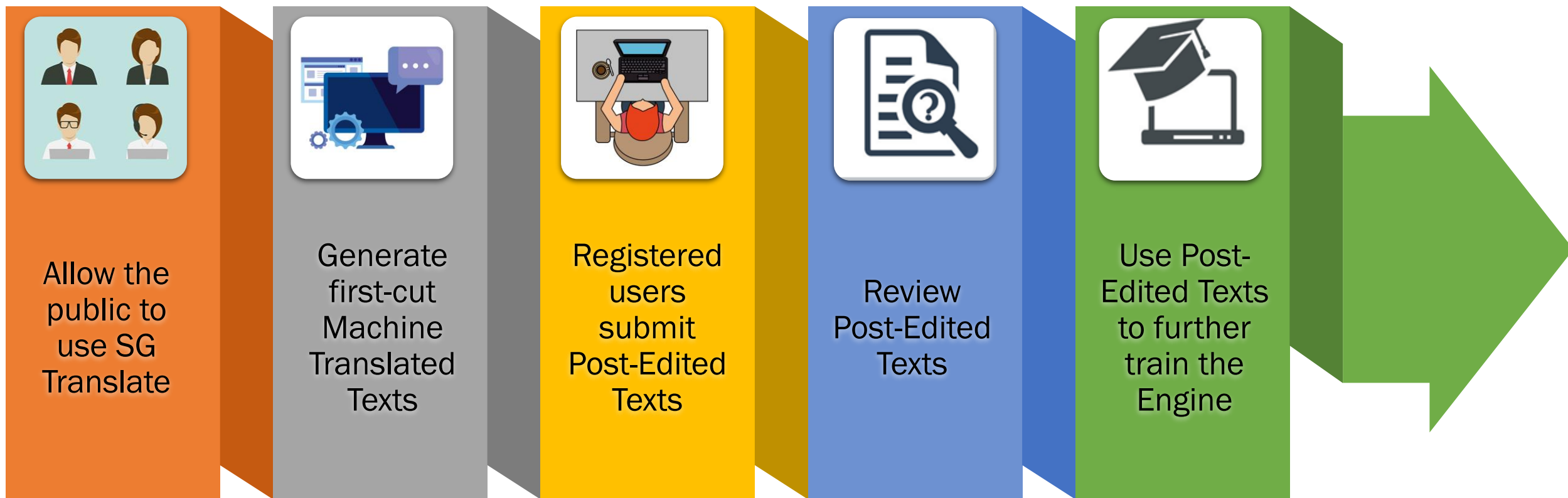
Coming soon!
Registered users can interact and exchange views





SG **TrQnslate**
TOGETHER

Overview of SG Translate Together



Sustainable Approach

⚠ continual training of the developed MT engine

Data from Government Agencies

Bilingual data is collected on a regular basis



Data from Partners

Bilingual data is collected from partners based on agreed timeline



Data from SGTT

Translators/Proofreaders can submit their edited translations



Clean the data collated and match the sentence pairs

Data Processing



Proofreaders review the translations by editing, rating and providing feedback

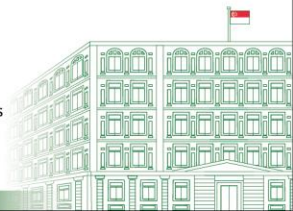
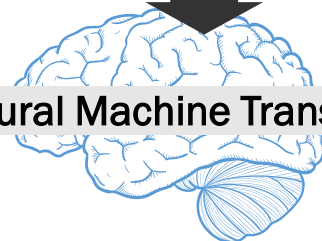
Learn new knowledge and update the translation model

Incremental Learning

To overcome catastrophic forgetting

Model Ensemble Technique

Neural Machine Translation



Shaping Singapore's Translation Landscape and Nurturing the Next Generation of Translators

Forming a **COMMUNITY**
Fostering a **SPIRIT** of sharing and learning
Nurturing the **NEXT GENERATION** of translators

Bridging communication and strengthening mutual understanding

Translation is important in Singapore's multiracial and multilingual society.



Joint ownership

Lower the barriers in putting out translated content and improve accessibility to information.



However, we emphasise the importance for human post-editing to ensure that the correct message is being conveyed to the target audience.



Harnessing technology

Harness technology to improve work processes and collaborate with the community.

Supporting and nurturing translators

There are initiatives to support translators (e.g. Translation Talent Development Scheme).

Registered users can receive feedback from SGTT Proofreaders on how to improve and through the process hone their translation skills even if they do not have prior experience.





Ministry of Communications
and Information



Thank You

 www.mci.gov.sg

 SingaporeMCI

 MCI singapore

 SgMCI

