**Data Visualization**

> 1929 or 1989?

= 1929年还是1989年?

> Trump has few political allies to help him out, while Clinton’s many surrogates include her husband, former President Bill Clinton; President Barack Obama and First Lady Michelle Obama, the breakout star of this election; and Vice President Joe Biden and Clinton’s vice-presidential running mate, Tim Kaine.

= 特朗普几乎没有政治盟友帮他指点迷津，而克林顿有许多代理人，包括她的丈夫、前总统比尔·克林顿； 奥巴马总统和在此次选举中冉冉升起的新星、第一夫人米歇尔·奥巴马；

> And every CCP leader since has carried forward his vision of a greater China, adjusting and expanding it as the country’s power grows.

= 之后的每一任共产党领袖无不以发扬毛的大中国目标为己任，随着中国国力不断增长大中国也不断地调整和扩张。

> Instead of pursuing the mirage of a two-state solution, the US should use its leverage in the region to clear the way for the emergence of a binational, democratic state that guarantees full political and civil equality for all of its inhabitants.

= 美国不应该追求两国方案这一海市蜃楼，而应该动用其在该地区的资源扫清障碍、建立一个双民族民主国家，并确保所有居民都获得完全的政治和民事平等。

> In Latin America, the committed intellectual is alive and well.

= 积极参政的知识分子正在拉美表现得生龙活虎。

> African girls are also disproportionately vulnerable to contracting HIV.

= 非洲女孩还非常容易感染艾滋病。

> When a woman gives birth to a child with a congenital syndrome deriving from the Zika virus, the response should be grounded in the dignity, value, and rights of each individual.

= 当妇女生下因为寨卡病毒而存在先天性疾病的儿童时，其应对手段应该基于尊严、价值观和个体权利。

**Tokenization**

Why train new tokenization?

1. The most commonly used tokenizer (ex: SentencePiece and Tiktoken) is trained on multiply languages, but my translation task only focuses on 2 languages. Therefore the default tokenizer contains lots of “useless” tokens for my specific task.
2. Training data used in SentencePiece and Tiktoken is mostly English. So they contains mostly English tokens. I want my tokenizer to contain roughly equal number of tokens for my languages.

Starting BPE algorithm

100%|██████████| 10000/10000 [13:55:41<00:00, 5.01s/it, Vocab=10255, IdsLen=1.43e+7, Ratio=4.64]

**Data Processing**

1. Tokenize
2. Create language vocab
3. Train/Test/Valid split
4. Create data loader

**Training Curves**

|  |  |
| --- | --- |
|  |  |

**Test dataset Performance**

100%|██████████| 133/133 [00:42<00:00, 3.14it/s, Loss=1.62, BLUE=0.69]

> if the exit cannot be navigated successfully , a dovish fed is more likely to blow bubbles .

= 如果不能成功实施退出，那么一个鸽派的美联储更有可能制造泡沫。

< 如果美联储不成功地退出，那么就无法避免一场泡沫的泡沫。

> erdoğan responded by removing many gülenists from their positions in the police and judiciary .

= 埃尔多安的应对措施是清除了一大批警方和司法系统中的葛兰主义者。

< 埃尔多安的众多对手来自埃尔多安的警察和他们的葛兰主义者。

> the technology that allows us to unravel cellular and subcellular processes and mechanisms , identify the causes of diseases and develop more specific and effective treatments , and determine who is biologically related to us and to what degree combines knowledge from biology , computer sciences , information technology , and material sciences .

= 让我们明白细胞和亚细胞过程和机制、认识疾病的原因并开发出更具体有效的治疗方法、确定谁与我们在生物上有亲缘关系的技术是生物学、计算机科学、信息技术和材料科学知识的结合。

< 这一发现的技术可以通过多种方式来评估和分析和分析信息，以及更有效的知识和信息，以及更有效的信息和信息，以及信息和信息和信息的知识和信息。

> some participants thought they had agreed to a sort of semi - fixing of exchange rates in the form of target zones , but the powerful german bundesbank never shared that interpretation .

= 有些与会国赞成某种以目标区间为形式的半固定汇率，但说话很有分量的德国联邦银行却并不支持那种思路。

< 一些参与者认为，他们的观点是一个强大的观点，但在德国央行的汇率中，德国央行从来没有共同的汇率。

> the idea that underlies all of these initiatives , and the notion of inclusive capitalism itself , is that companies must be managed for the long term .

= 这些动作的背后思想，以及包容性资本主义本身的概念，在于公司管理必须着眼长期。 遵循这一方法的公司会关心未来劳动力的技能；

< 这些观点认为，资本主义的思想是资本主义的概念，必须是建立长期的资本主义。

> russia is by no means seeking to isolate itself .

= 俄国根本不希望自我孤立。

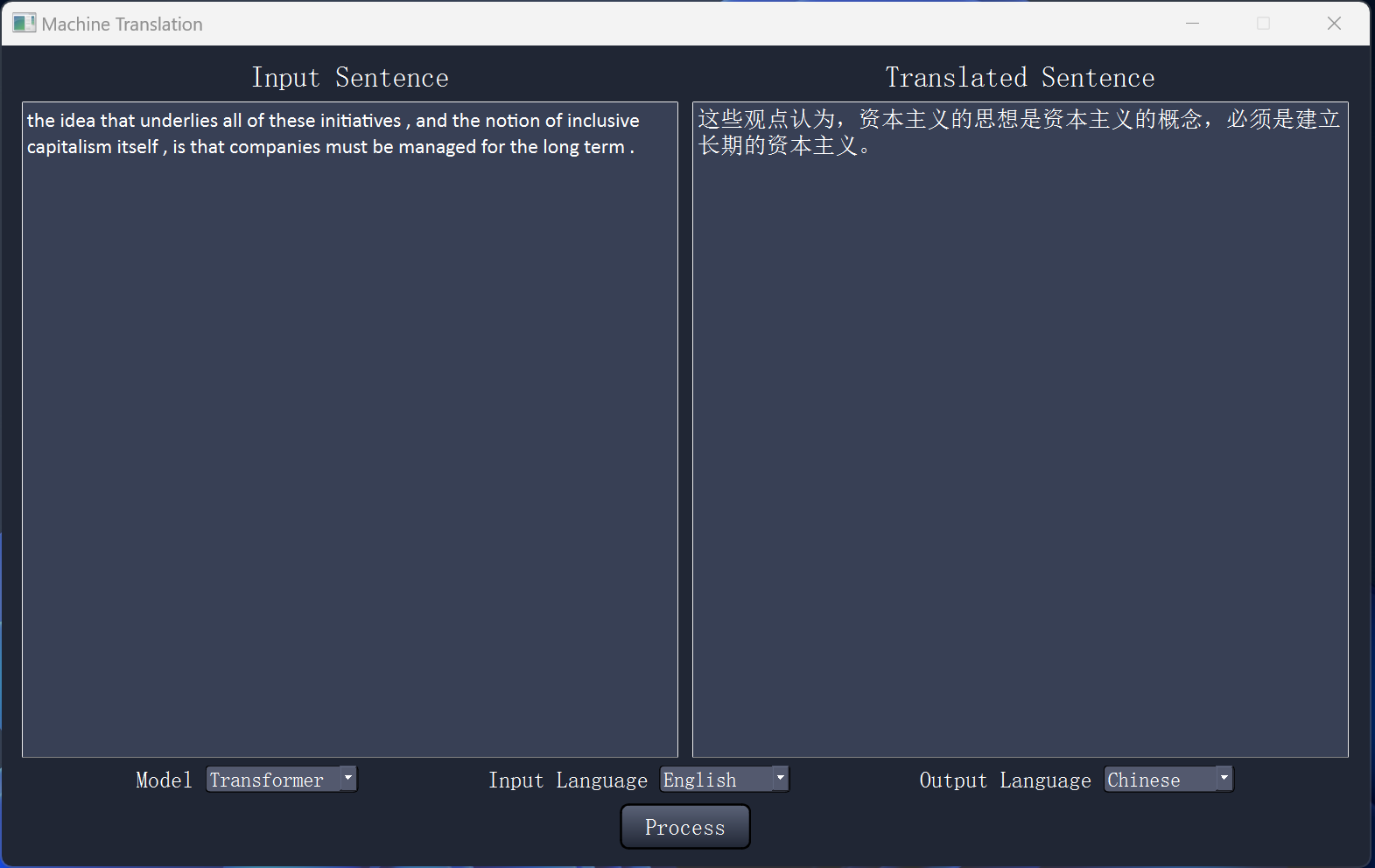
< 俄罗斯也没有理由认为自己是不让步的。

> is the financial world a safer place today ?

= 如今，三年过去了，金融业变得安全了吗？

< 当今世界的金融安全吗？

**GUI**



User Manual:

1. The Dropdown Model menu lets you pick which model to use.
2. The Dropdown Language let you pick the input and output language (currently only English to Chinese available)
3. Then press “Process: button to generate a segmentation map.