

测试说明

我们的project主题是用大模型做音乐剧译配。给定英文歌词，AI模型会自动生成对应的中文译配。我们使用不同模型和方法生成了一些结果，请你根据我们制定的规则为这些结果打分。

测试共分为两个部分。第一部分对单句在翻译质量和可唱性等方面分别打分。此部分共30x5=150道题。第二部分对整段打分，需要综合考虑歌词文本及其与音乐的配合。此部分共12x5=60道题。涉及到歌词与音乐配合时，我们会提供参考音频。

一. 单句打分

你将拿到：一句英文歌词，一句中文翻译，一个包含此英文歌词的段落；一个原始歌曲片段，一个把歌词唱出来的参考音频（注：作者唱功有限，如果你对那首歌比较熟悉，可以自己唱出来感受）。

你需要做的事：先**仅根据文本**，在成句性、翻译准确性、歌词性三方面分别打分；再听原始歌曲片段和译配片段，为译配歌词与音乐的配合情况打分。评分标准见下。

成句性（仅需考虑中文文本是否成句、流畅）

- 1分：不是人话--内容荒唐、不合逻辑或看一两眼看不懂

爱，纯且贞，远远地

- 2分：部分成句，但有硬伤（**无法接受**），如出现一些非常不合适的词汇，或者缺少必要成分、有严重歧义、语序混乱等

第一次长久以来的（语序混乱，应为“长久以来的第一次”）

- 3分：大部分成句，没有硬伤（**勉强可以接受**），但有尴尬的用词、或不太要紧的歧义，与正常的汉语句式相比差点意思

命运点亮希望光（“希望光”用词略显尴尬）

- 4分：很通顺，能轻易地理解含义

当你感觉站不稳的时候

准确性（结合段落，评判歌词的翻译是否准确）

- 1分：大于50%部分翻译错误，或少量关键部分（如主被动、动词）翻译不对或缺失，**无法接受**

Fellows will fall in line

兄弟长相厮守（完全不对，应为“男人们会排队等待”）

- 2分：小于50%部分翻译不精准，**勉强可以接受**（允许意译，允许忽略或改变少量不重要信息）

Don't know if I'm elated or gassy

不知我是欢喜还是气胀（gassy在这里译为气胀不准确）

- 3分：基本准确，但还有进步空间，如英文俗语直译而没翻译出来引申义，或加上一些小细节更好等

What is it like in his shoes?

穿他鞋，感觉如何？（俗语in sb's shoes，翻译成“如果我是他”更好）

Sven, the pressure is all on you

- 史文，压力都在肩头（小瑕疵，应当是“压力都在你肩头”）
- 4分：意思非常准确（允许意译，允许忽略或改变少量不重要信息）

To run where the brave dare not go

跋涉，无人敢行的路

歌词性（仅需考虑中文文本是否适合当歌词）

- 1分：不像真正的歌词

那一个人被伤疤抹掉

- 2分：可以当做歌词，但平平无奇，没什么亮点

并不表示你只需要笑着忍痛

当你破碎在原地

- 3分：适合当做歌词，且有一定文学性

因为好久没在生命里

那一人，受辱满身伤痕

- 4分：非常适合当做歌词，且有创意，有表现力，**令人眼前一亮**

跋涉，无人敢行的路

这命运召唤我启航！

音词配合单句评价

主要关注三方面——

歌词字数：是否需要在一个音符内塞进多个字，或一个字对应很多音符？一般来讲，一个音符对应一个字最合适。

停顿：旋律中的停顿是否会把完整的句子/词组拆开？理想情况下旋律和语义的停顿位置相同。

倒字：是否有非常严重的倒字（把一个词听成了另一个，如“归来吧”听成“鬼来吧”）。

不需要关注翻译的是否准确。

各个分数对应的音频示例在文件 [单句示例.mp3](#) 中。

- 1分：歌词字数不完美，听起来不够舒服，有改进空间。

For the first time in forever

在人生中第一次（字数偏少）

- 2分：歌词字数很合适，但停顿非常不合适或有非常严重的倒字。

is anybody waving back at me?

有没有人向我挥手回看（“挥手”之间有停顿）

- 3分：歌词字数很合适，停顿较为合适，倒字不太严重，但仍有听起来奇怪的地方。

To right the unrightable wrong.

解决对不对的事情（“对不对”听起来奇怪，有点倒字）

- 4分：歌词字数很合适，停顿合适，倒字不严重。

For the first time in forever

永远的第一次体验（虽然翻译不太好，但音词配合还不错）

二. 整段打分

你将拿到：一段英文歌词，一段中文翻译，和一个把翻译歌词唱出来的参考音频（注：作者唱功有限，如果你对那首歌比较熟悉，可以自己唱出来）。

你需要做的事：对于整段，为歌词质量及其可唱性各打一个分数。

整段综合评价

歌词质量：

- 1分：大部分歌词都不像人话，或大部分歌词脱离原始歌词含义。
- 2分：大部分歌词像人话，但仍存在少数尴尬的地方（不可接受），如用词不当或翻译不对。
- 3分：歌词勉强能接受，但仍有瑕疵。
- 4分：看不出来是译配，好像原本就是中文歌词。

音词配合：

- 1分：音词配合很差，如很多句字数不对、在需要押韵的段落中非常不押韵...
- 2分：音词配合总体还凑合，但存在一些很尴尬的问题，如不合理的停顿、严重的倒字...
- 3分：音词配合没有大问题，但仍有瑕疵。
- 4分：听不出来是译配，好像原本就是中文歌。