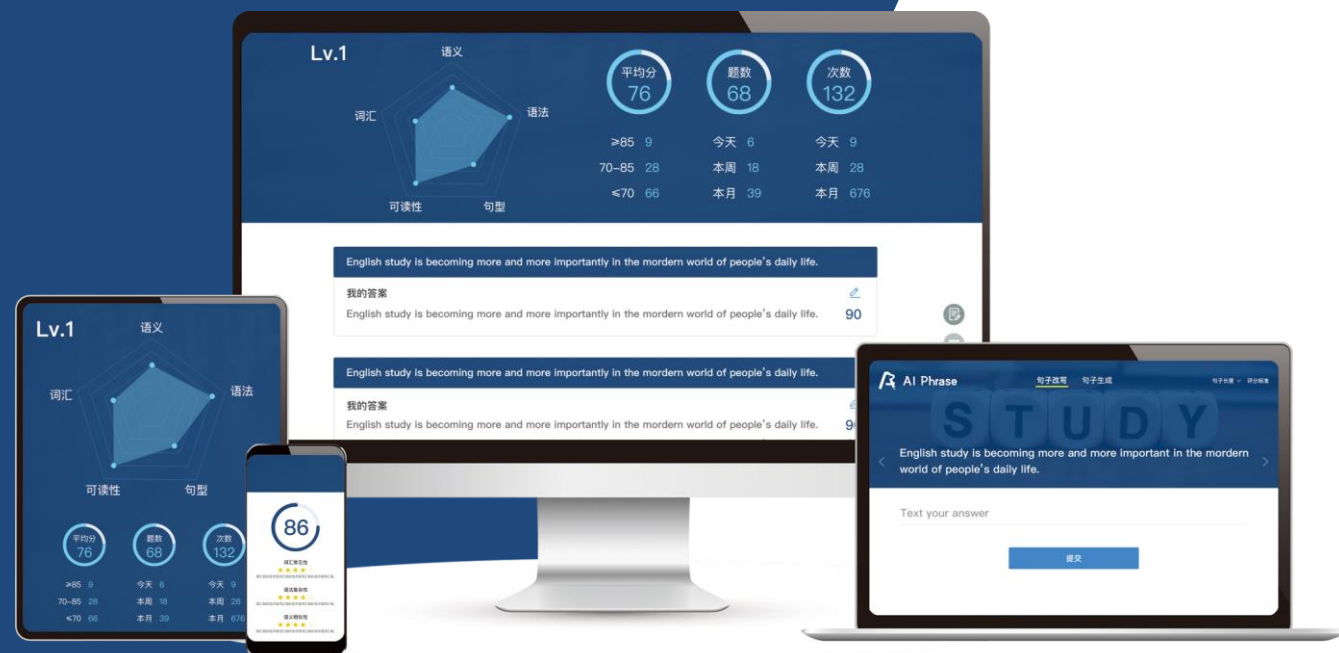


48 AI改作文

使用AI帮助学生改写英文语句

团队 妙比生花
企业 流利说



目录

CONTENTS

- ① 项目介绍
- ② 产品框架
- ③ 核心功能
- ④ 产品展示



项目介绍

需求分析



学生

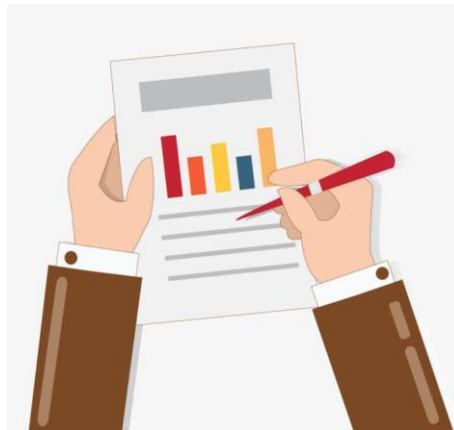
句子改写练习无法及时得到专业反馈

教师

句子练习批改任务重；
习题素材有限



解决方案



AI智能评估

从语义、语法、词汇、句型
等多个维度分析用户的改写
语句，给出评分和反馈

AI智能改写

针对输入语句，
生成多条同义改写语句，
丰富用户的语言素材

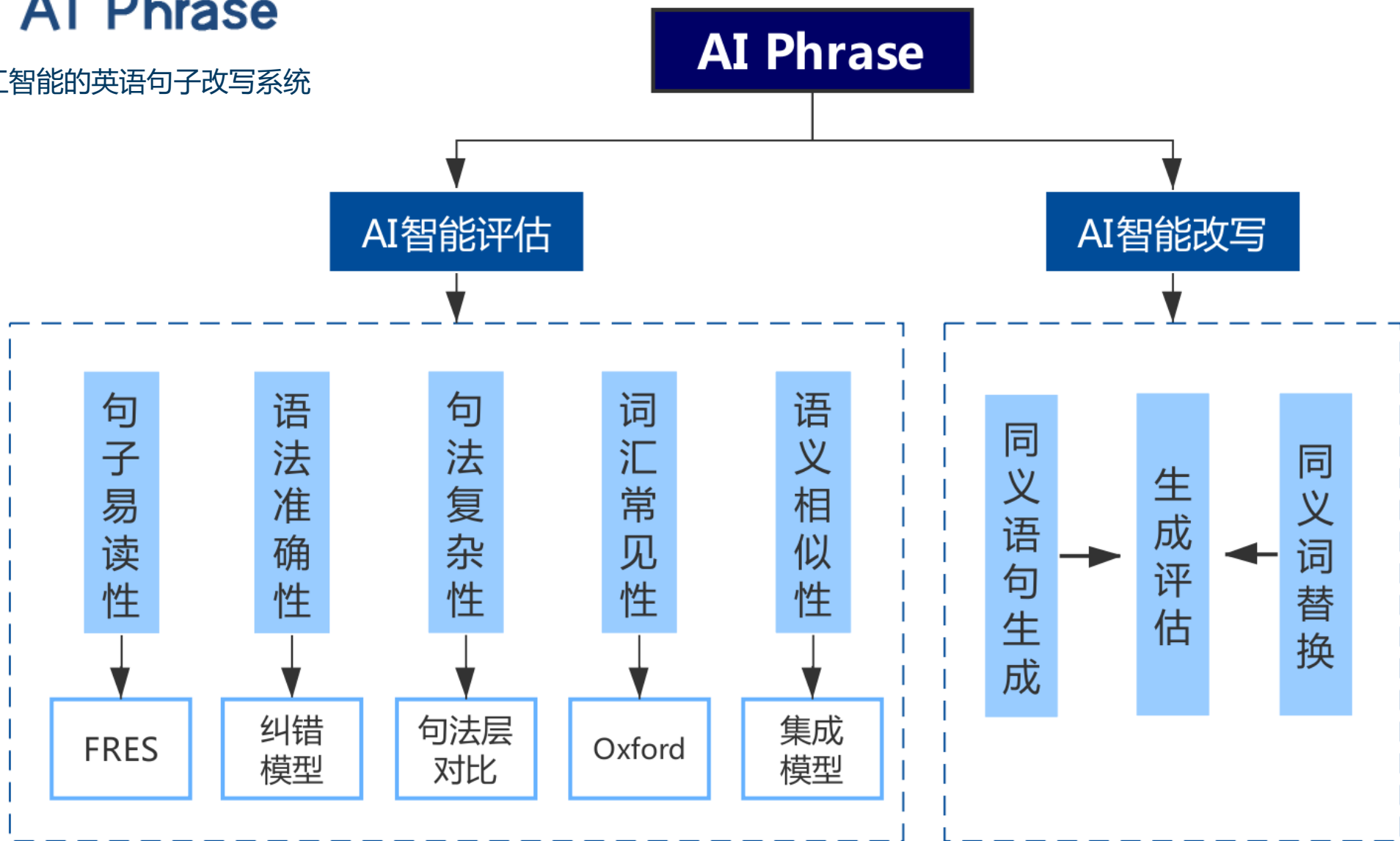




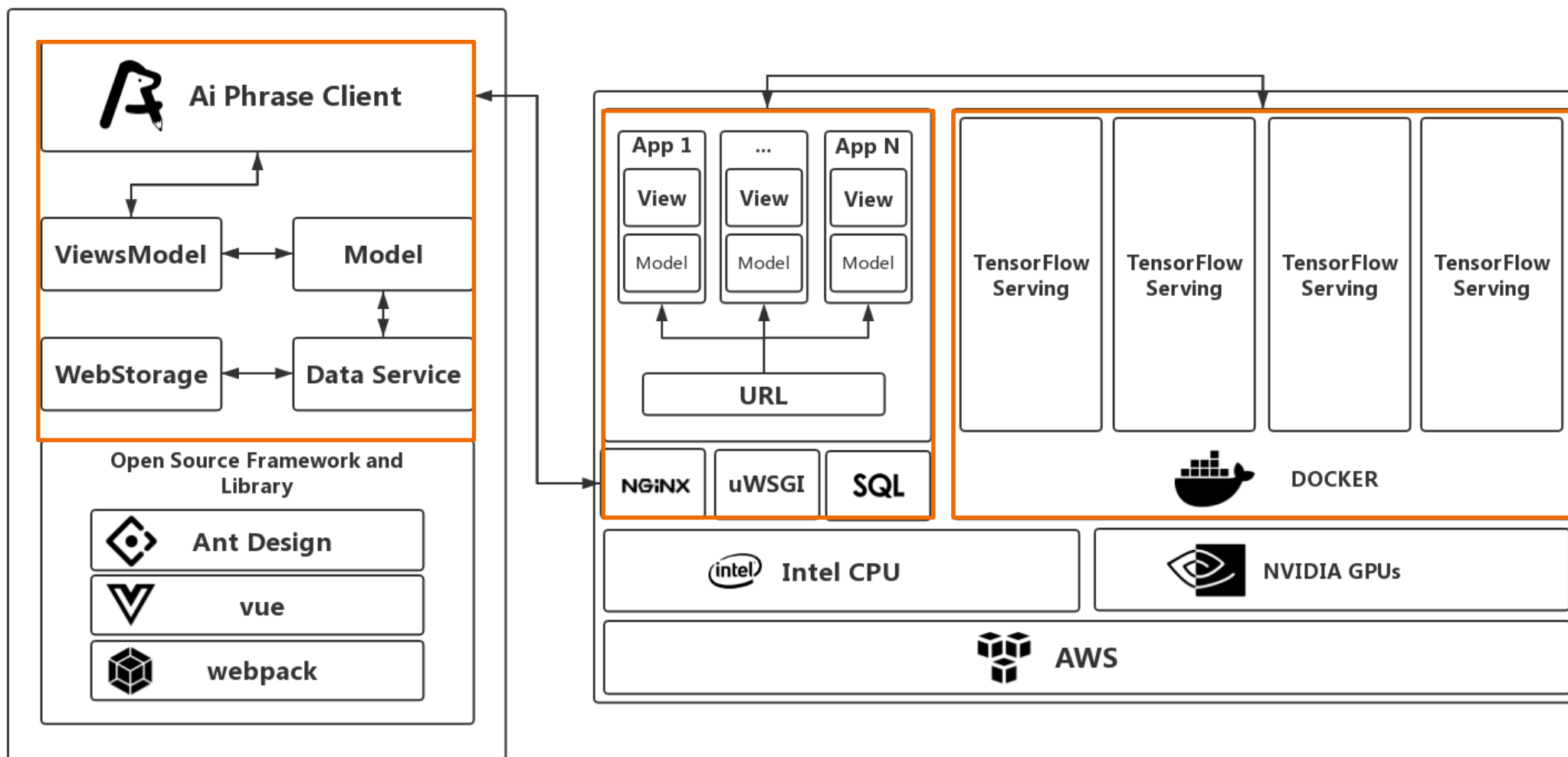
AI Phrase

基于人工智能的英语句子改写系统

产品框架



软件架构



微服务架构 模型松耦合、微模块化部署，提升可维护性和拓展性

高并发 引入 Nginx + UWSGI 提升系统高并发请求处理性能 为将来分布式机群部署提供可能

核心功能

CORE SERVICE



项目难点

语义理解、同义改写
人类语言复杂

To be or not to be
that's a question

生存还是毁灭
这是个问题

X_X??

人工智能

模型难以高效学习

数据

同义语句标注困难

Sentence Pair?

Sentence A

Sentence B

YES

YES

NO

可靠数据需要多人协作
可用优质数据少

方法提升

可用

有缺陷

难应用

Sentiment
Analysis

Information
Extraction

QA

NER

Summarization

POS
Tagging

Machine
Translation

Paraphrase

目前技术局限下，语义理解乃至
同义改写难以直接应用

自身方案

数据

需要海（湖）量的数据

语义相似数据集

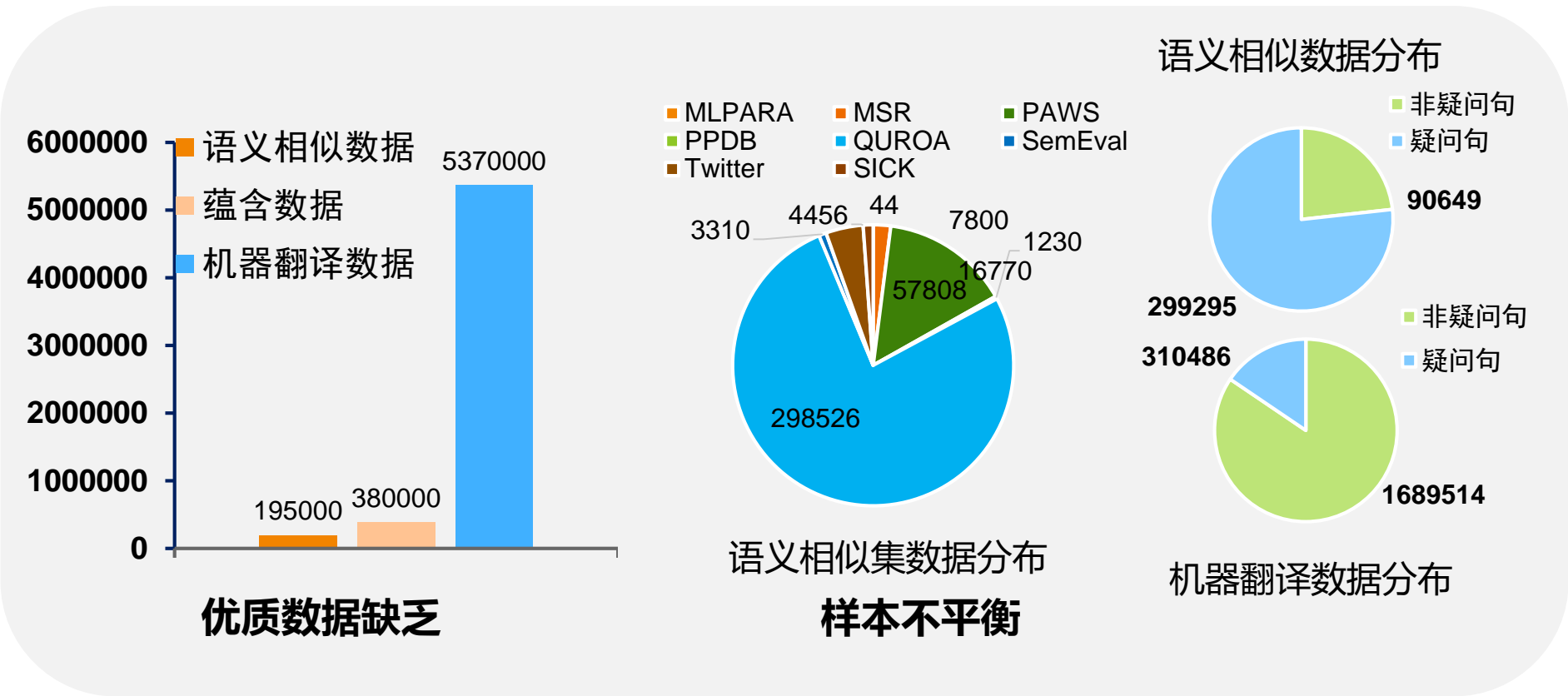
QuoraMSRPAWSTwitterPPDBMLPARASemEval 2012-2015SemEva 2016

蕴含推理数据集

SNLMNLIQNLI SICK

机器翻译语句对

Para-Nmt



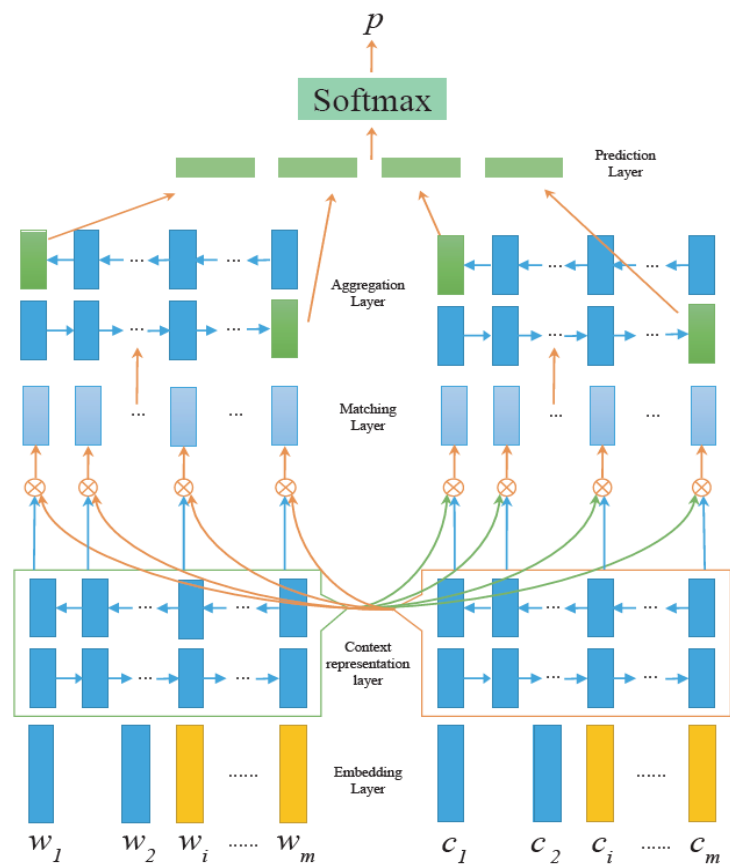
扩充数据集
150万同义语句对
数据集A

精简数据集
19万同义语句对
数据集B

独立验证测试集
4万同义语句对
验证集、测试集

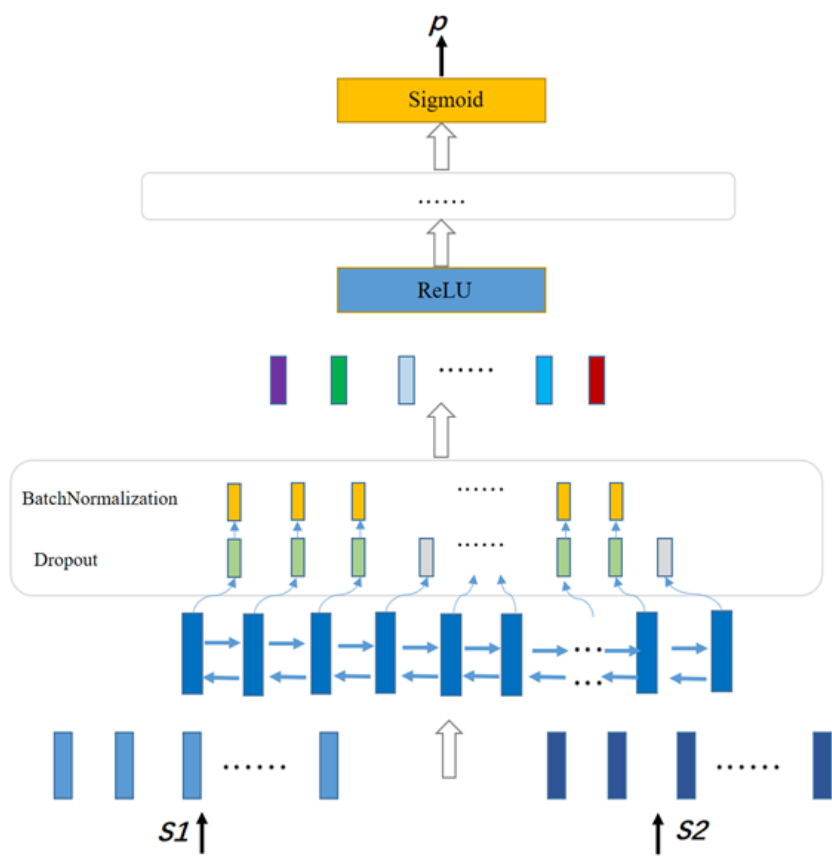
语义相似度判断模型

BiMPM交互匹配判别模型



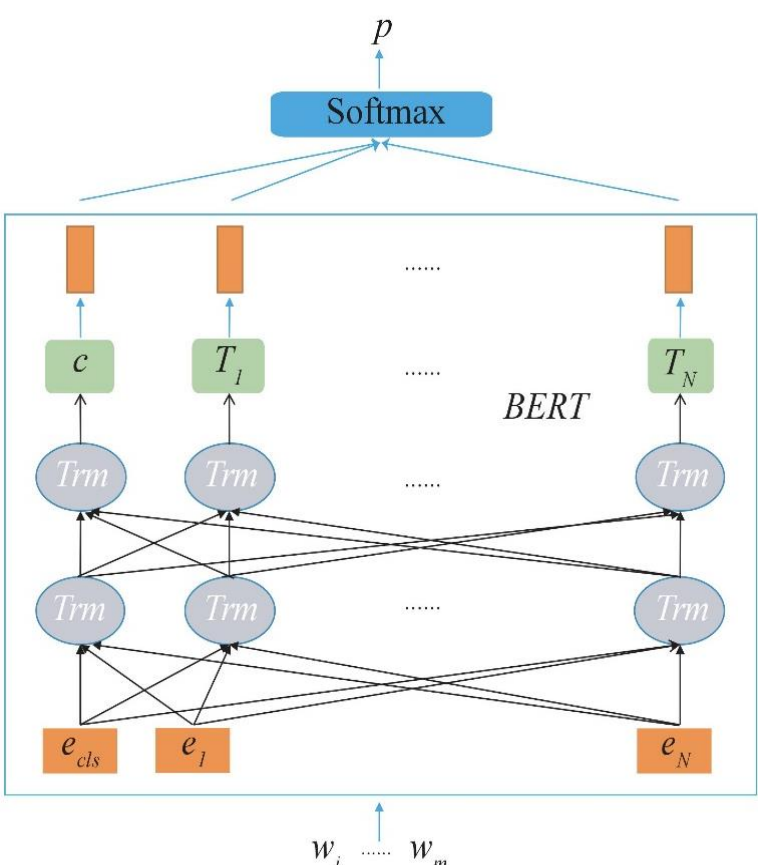
Model	train	dev	test
BiMPM	0.873	0.852	0.843

BiLSTM + Attention编码分类



Model	train	dev	test
BiLSTMA	0.962	0.934	0.903

BERT上下文词向量检测模型



Model	train	dev	test
BERT	0.971	0.946	0.935

模型集成

字符级信息表达
句子交互匹配
训练相对耗时较高

BIMPM

文本序列特征编码
注意力机制
模型训练快速
快速适应不同条件

BiLSTM+Att

基于transformer的
上下文向量表达
海量数据预训练
先验知识

BERT

多模型集
群训练



模型性能提升 泛化性增强

独立验证测试集 4万同义语句对

Model	simple_dev	simple_test
BiMPM	0.856	0.824
Att+BiLSTM	0.804	0.791
BERT	0.847	0.807
Ensemble	0.876	0.834

平均性能提升3%

AI 智能评估

词汇丰富度

使用Oxford 5K词汇评价标准

界定常见词与生僻词



语句易读性

使用FRES通顺等级指标

语句长度与音节数
评估阅读流畅性



语法准确度

Language-check

开源工具包

规则评估纠错



句法复杂度

句法树层级比对

判断改写句结构复杂度

Although putin is undoubtedly an important and influential figure, his domestic reach has its limits.

句子易读性



语句复杂难懂，适合本科生阅读

句法复杂性



相比原句，句子成分间的依存关系复杂

AI 智能改写

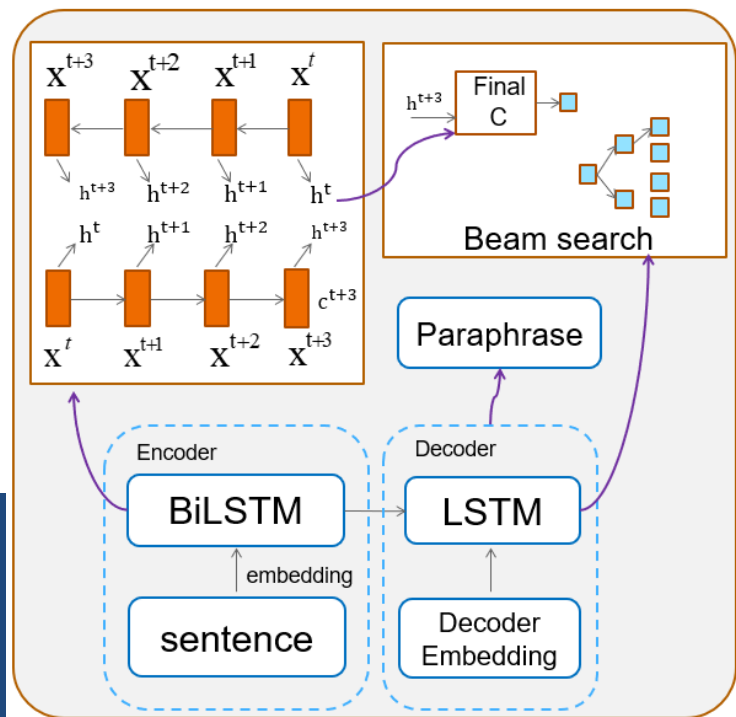
算法迭代尝试
构

状态传递方式

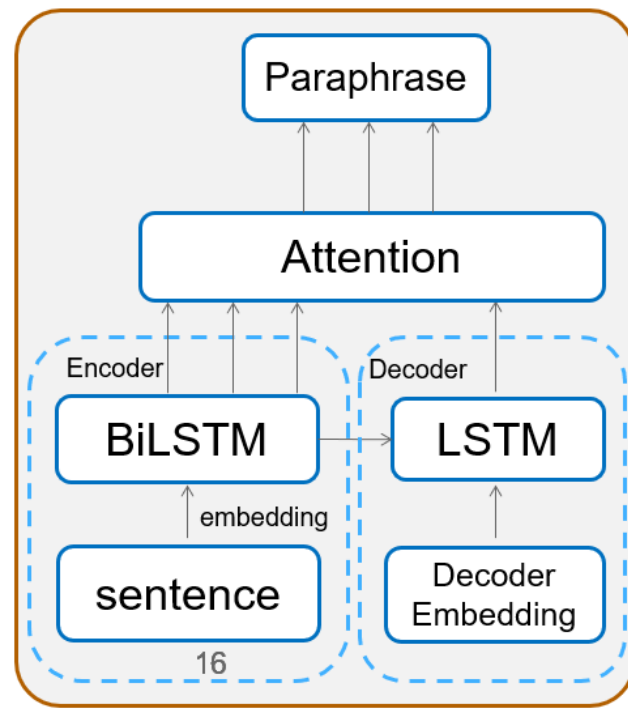
Cell

层级结

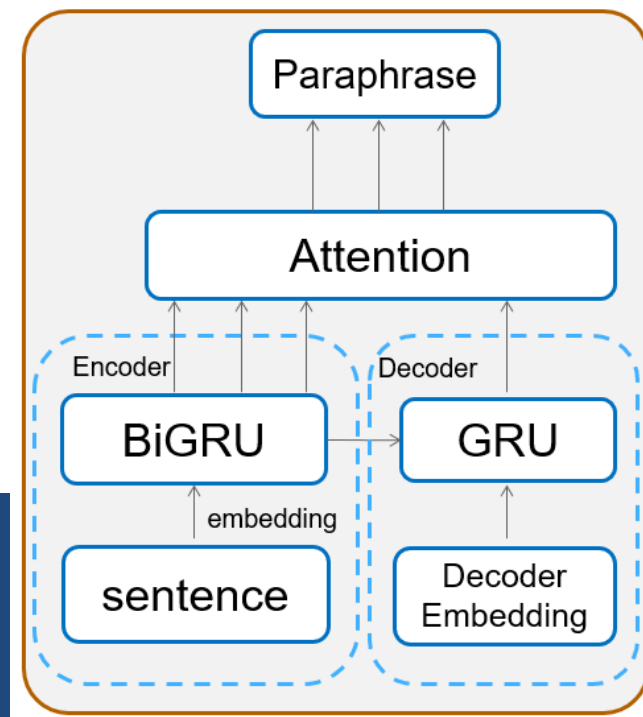
BiLSTM



BiLSTMA

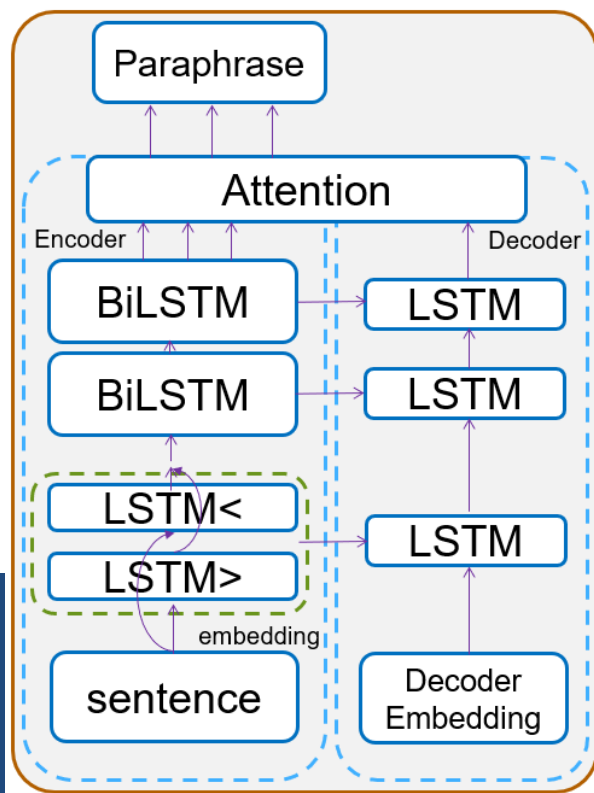


BiGRUA

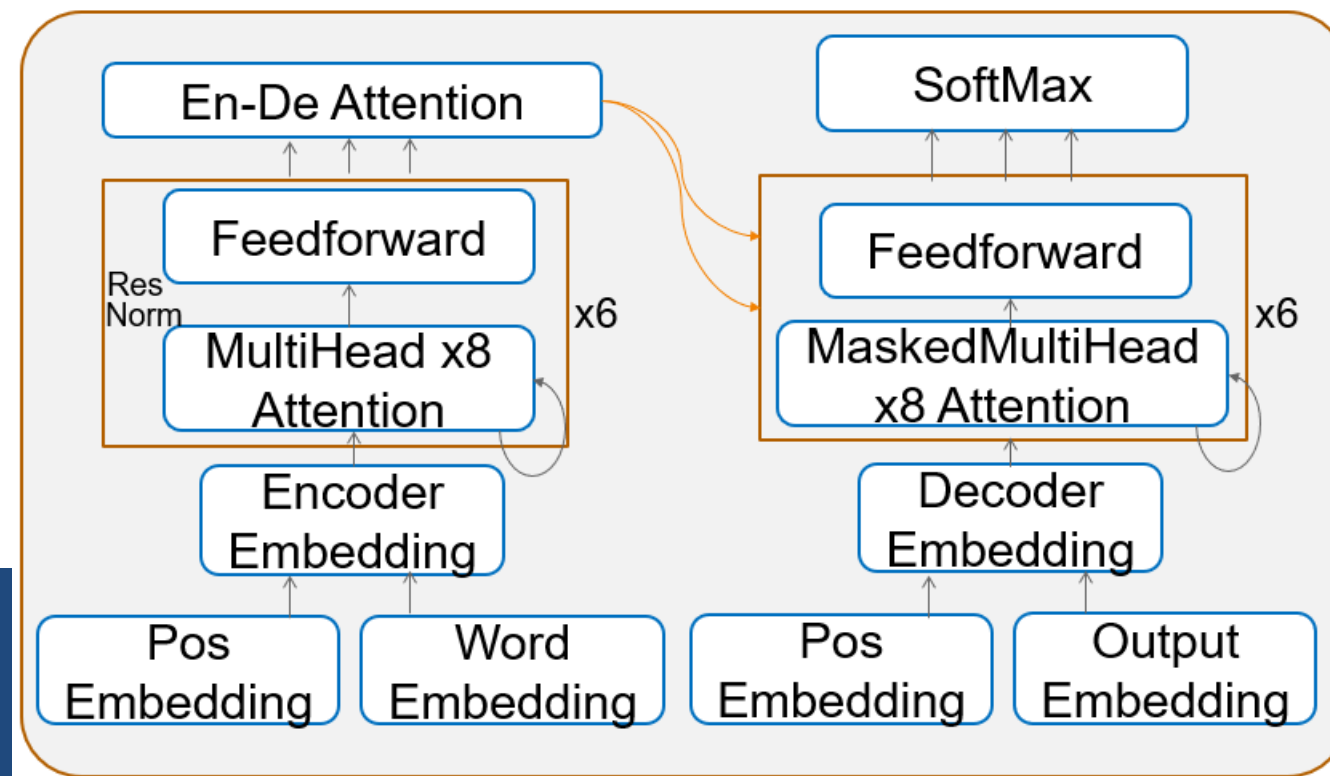


AI 智能改写

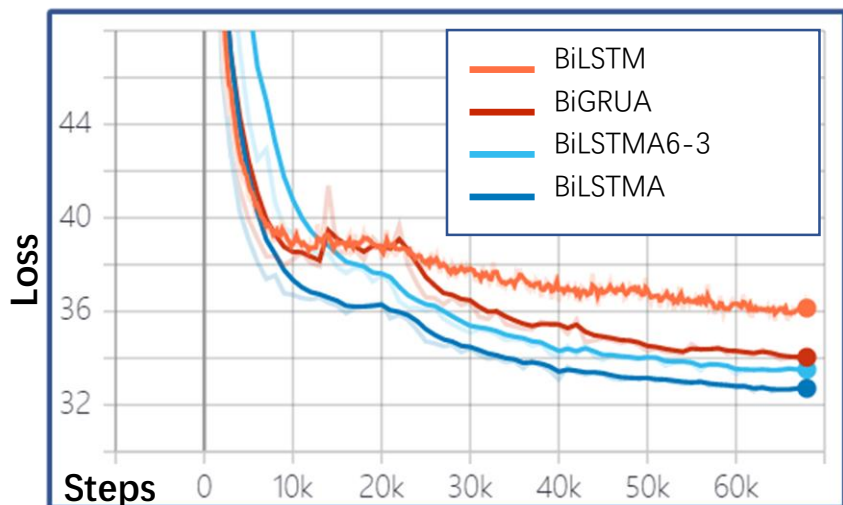
BiLSTMA 6+3



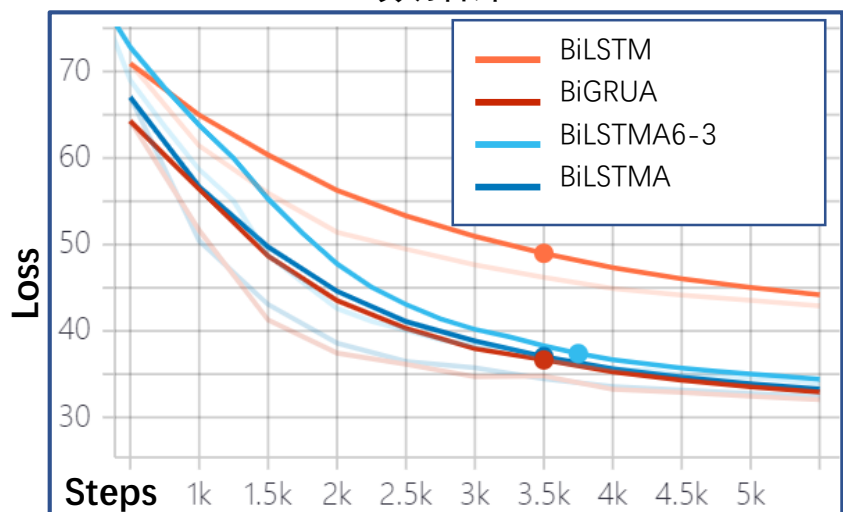
Transformer 8Multihead 6block



模型评估



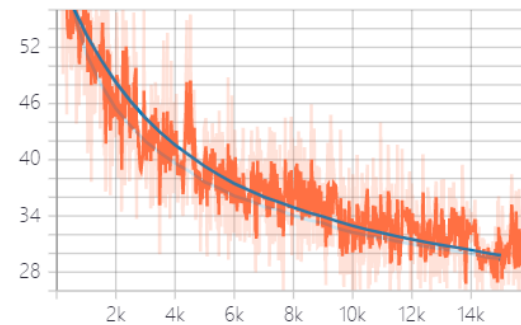
数据集A



数据集B

	Model	Bleu1	Bleu4
数据集 A	BiLSTM	0.6417	0.2769
	BiGRUA	0.6158	0.3118
	BiLSTMA	0.6405	0.3653
	BiLSTMA6-3	0.6929	0.4085
数据集 B	BiLSTM	0.5301	0.2213
	BiGRUA	0.5835	0.2868
	BiLSTMA	0.6167	0.3224
	BiLSTMA6-3	0.6079	0.3325
	Transformer	0.6069	0.3293

Bleu互译质量评估



Transformer

```
- source: why is spanish accepted as a second language in sanctuary cities , when all founding fathers spoke english only ?
- expected: why is spanish accepted as a second language in sanctuary cities , when the founding fathers spoke only english ?
- got: why is spanish accepted as a second language in sanctuary city , when all english spoke only in english language ?

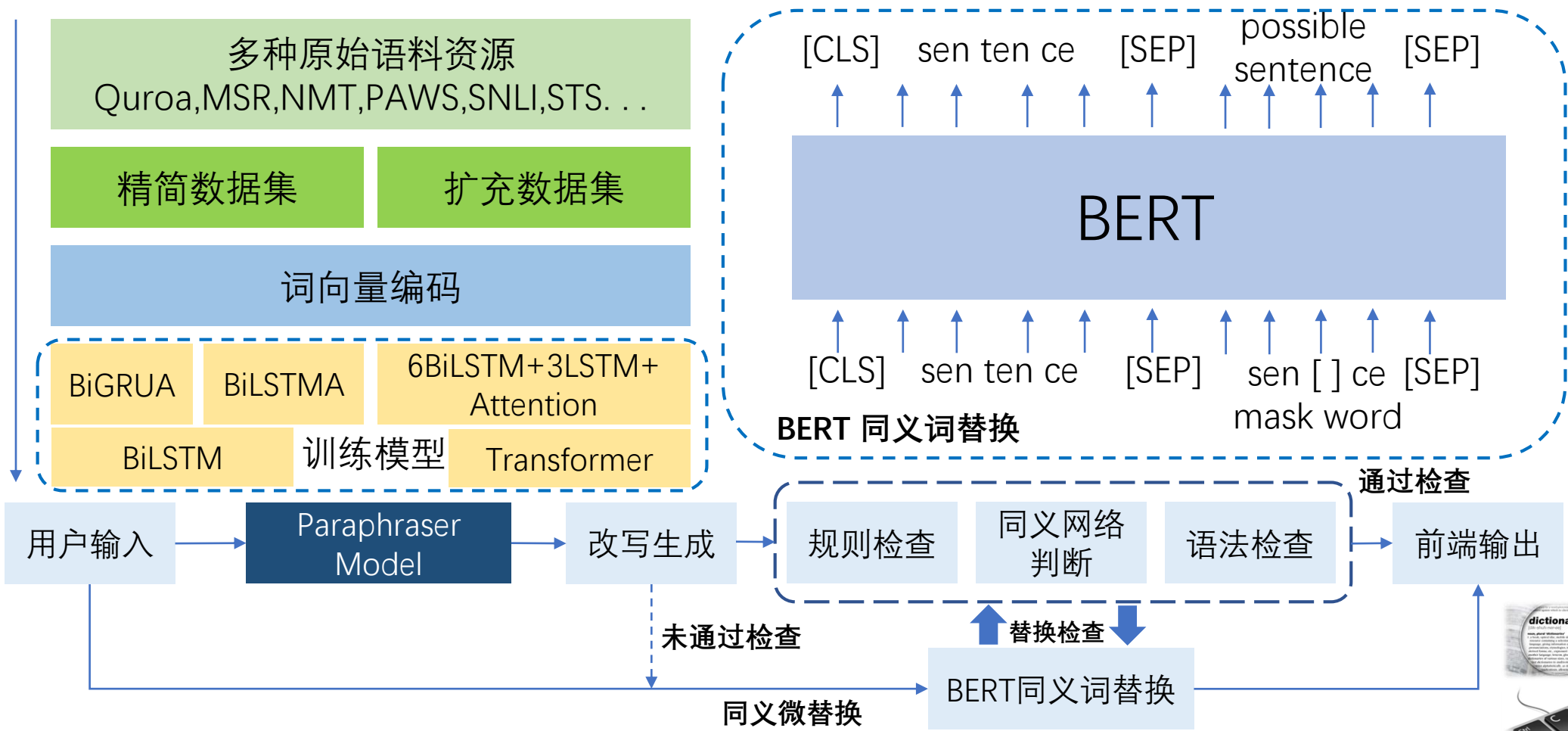
- source: how i hack facebook account ?
- expected: how can we hack fb ?
- got: how can i hack a facebook account ?

- source: love on the line is the name of a crazy penis album produced in 2008 , released only in australia .
- expected: love on the line is the name of a crazy penis album that was produced in 2008 and was released only in australia .
- got: love is the name of a penis album produced in the only australia , which was released in australia .
```

可用性三重保险

1

算法保险



2

人身保险

There are only a few programs on television that i like.	✓
There are programs on television that i like watching.	○
There are only a couple of good programs on television that i like watching.	○
There are only the few good programs on television that i like watching.	✗
There are only a few good programs from television that i like watching.	✓

改写参考

辅助检查

3

CtrlC保险

There are only a few good programs on television that I like watching.

Paraphrase it

✓ 已复制



潜力同义句对

存入数据库

产品展示

PRODUCT DEMO



网址: <http://52.80.106.20> 验证码: 3773



你身边的AI写作专家



THANKS

