

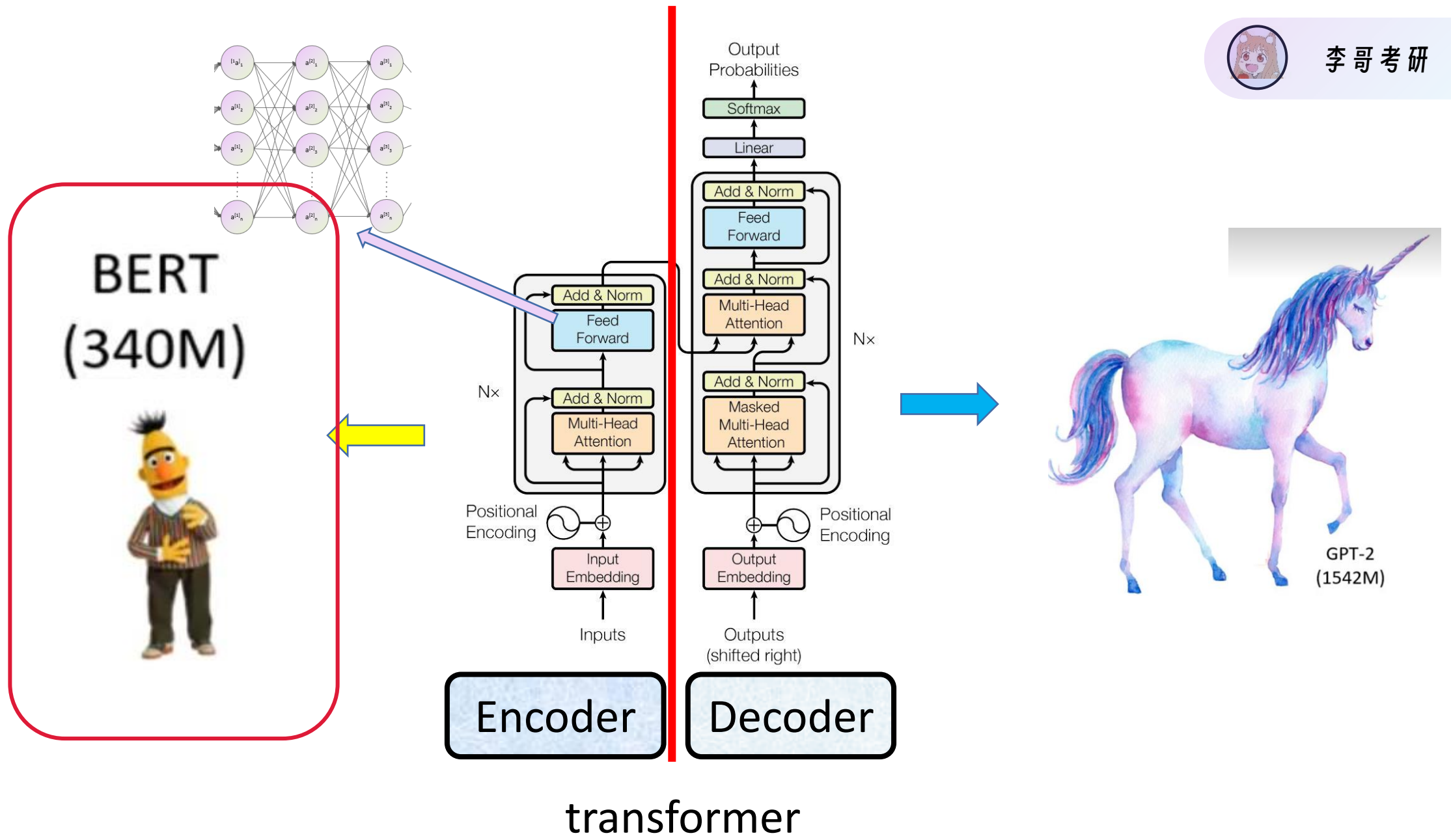


李哥 考研

# 第三部分： self-attention

## 与 Bert

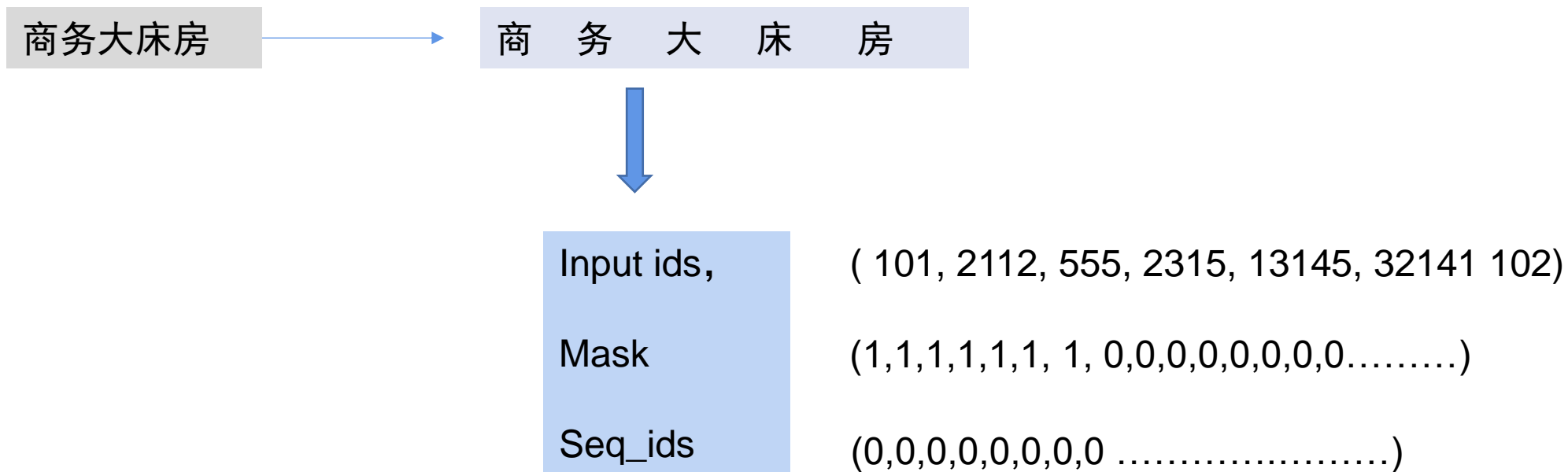




# 需要分词



李哥考研



# 需要分词



李哥考研

商务大床房 → 商 务 大 床 房 → CIs 商 务 大 床 房 sep pad pad

今天的风儿真是喧嚣呀 → 今 天 的 风 儿 真 是 喧 嚣 啊 → CIs 今 天 的 风 儿 真 是 sep



Input ids,	(2112, 555, 2315, 13145, 32141 )
Mask	(1,1,1,1,1, 0,0,0,0,0,0,0,0.....)
Seq_ids	(0,0,0,0,0,0,0,0 .....)

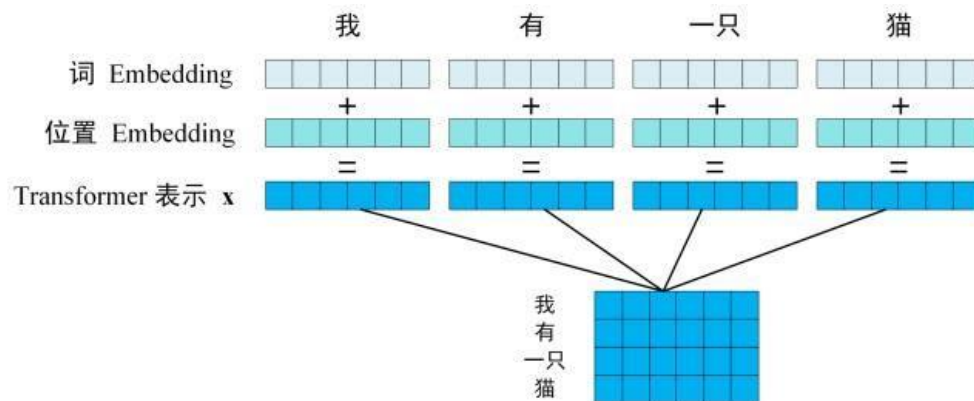


# Bert输入embedding



李哥 考研

Transformer:

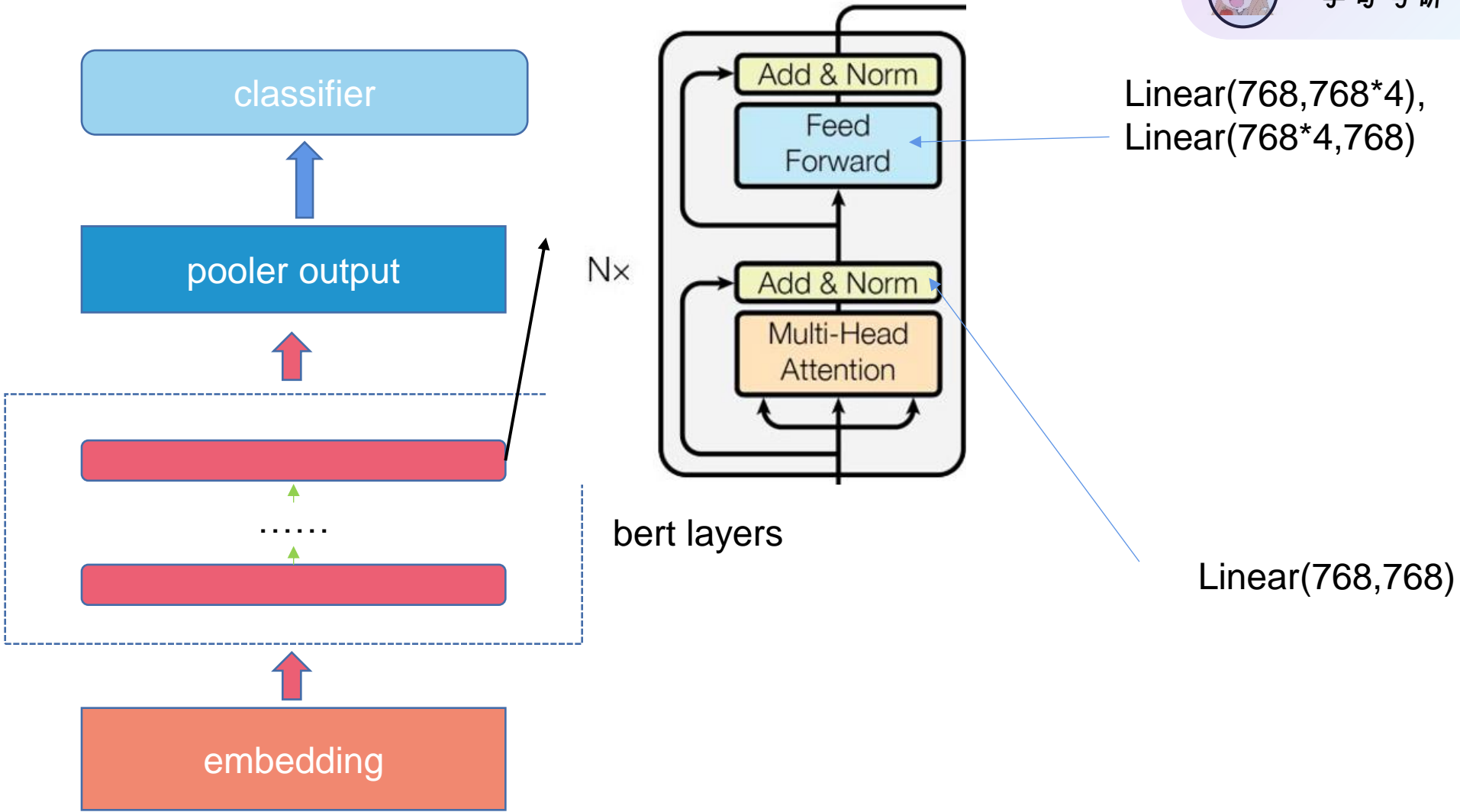


Bert:

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	$E_{my}$	$E_{dog}$	$E_{is}$	$E_{cute}$	$E_{[SEP]}$	$E_{he}$	$E_{likes}$	$E_{play}$	$E_{##ing}$	$E_{[SEP]}$
Segment Embeddings	$E_A$	$E_A$	$E_A$	$E_A$	$E_A$	$E_A$	$E_B$	$E_B$	$E_B$	$E_B$	$E_B$
Position Embeddings	$E_0$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$

可以看到一些特殊的token。

# 计算bert参数





李哥 考研

Bert就是一个特征提取器

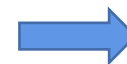
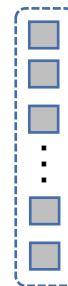
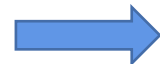


预训练任务

大佬帮我们训好了

微调

我今天很高兴



好/坏

# 数据集



李哥考研

- 1, "距离川沙公路较近,但是公交指示不对,如果是""蔡陆线""的话,会非常麻烦.建议用别的路线.房间较为简单."
- 1,商务大床房, 房间很大, 床有2M宽, 整体感觉经济实惠不错!
- 1,早餐太差, 无论去多少人, 那边也不加食品的。酒店应该重视一下这个问题了。房间本身很好。
- 1,宾馆在小街道上, 不大好找, 但还好北京热心同胞很多~宾馆设施跟介绍的差不多, 房间很小, 确实挺小, 但加上低价位因素, 还是无超所值的; 环境不错, 就在小胡同内, 安静整洁, 暖气好足-\_-||。。。呵还有一大优势就是从宾馆出发, 步行不到十分钟就可以到梅兰芳故居等等, 京味小胡同, 北海距离好近呢。总之, 不错。推荐给节约消费的自助游朋友~比较划算, 附近特色小吃很多~
- 1, "CBD中心,周围没什么店铺,说5星有点勉强.不知道为什么卫生间没有电吹风"





0,"房间太小,宾馆有租四轮自行车的,很破,骑不动,后来服务员说他们的车都是东方绿舟淘汰的.酒吧装修的还可以,就是蚊子太多,象轰炸机."

0,我去年的时候去过该酒店,那时候住的是无烟房,感觉还好,这回去住的是高级标准间,进去了房间就有一种怪味道要求前台换房,换过了以后还是多少有一点怪味。不过房间的墙壁特脏,后来想算了,就没在换,不过感觉今年的金鼎要比去年卫生环境差很多,装修也旧了!

0,尼斯酒店的几大特点:噪音大、环境差、配置低、服务效率低。如:1、隔壁歌厅的声音闹至午夜3时许、凌晨6时起又“享受”健身操场噪音;窗户破旧不隔音。2、电视无港、澳、凤凰、英文.....等台。3、前台服务人员少(2名),在早餐前已交房卡要求结算,待早餐后仍未算出来,无任何消费,而且只有我们在结帐,住店的客人极少。如此四星?少见。建议与“凯莱酒店”(尼斯附近)签约(全国连锁),价格也实惠。

下节预告。



李哥考研

生成任务，大模型

作业：



李哥考研

完成外卖数据集用bert分类。

用调试跑一遍bert的维度变化。



李哥考研

答疑和结束

THANKS

