* 语料（先只用知乎，因为其他平台上的大多也被手工搬到知乎里了）：

1.拿到知乎dump

2.提取相关话题下的问题，并将所有问题和回答爬下来

关键字seed：套路，异性交往，心理学，恋爱，婚姻价值观，两性情感，男性心理，情话，恋人相处，恋爱心理，审美，服装搭配，相亲，美食，健身，生活，婚姻，结婚，彩礼，笑话，有趣，心理，追求女孩，女孩，家庭教育，恋爱心理，情侣冷战，恋爱恐惧，两性关系，社会，90后，青春，婚姻价值观，相亲话题，社会现象，社会阶层，朋友，人际交往，择偶，价值观，泡妞，大学，生活方式，素养，女性心理，女朋友，求婚，生活常识，搭讪，撩妹，爱情哲学，分手，异地恋，关心，亲密关系，异性交往，恋人相处，恋爱学，恋爱心理，真爱，漂亮，女生，相貌，外貌，外貌协会，吵架，女朋友，学生，旅行，爱情，魅力，男生，情感，约会，单身，家庭，恋爱技巧，圈子，社会，表白，暗恋，北京，北京生活，旅行，吃喝玩乐，发型，穿搭，美食，潮牌，旅游…

策略和算法：尽可能大而全，话题下的问题也有其他话题标签，先把相关的话题做成一个全集，（手动去掉不太相关的），然后再把所有话题下的问题和回答爬下来，语料保存下来供以后retrieval-based QA检索最佳回答用

3.用2中预料训练一个word embeddings，供后面检索用

* 人工标注

分出三个人找潜台词QA文本（先1000条）

e.g. 知乎问题:你有哪些当初没听出来的弦外之音

Q：我感觉你这个人很贴心，给我一种闺蜜的感觉

A：咱俩也就闺蜜关系了，更进一步的事，你就别想了

Q：他的确不那么好，可能我真的傻，就是喜欢他

A：虽然他坏坏的，但是和他在一起，我每天都感觉到新鲜刺激，我喜欢那种感觉

Q：我现在什么都不缺，生活也比较充实，不想谈恋爱

A：我不是不想谈恋爱，是不想和你谈恋爱

Q：我这个人很老实，对女孩子也好，为什么没有人喜欢我

A：除了“老实”，你再也没有可以吸引女孩子的王牌了

* 固定人格类问题先加入 （关于调戏AI的问题，进行识别给予固定回复 比如：我是AI 恋爱大师，我诞生于创新工场的DeeCamp训练营, etc.）