***数据集调研与rebattle任务分析***

***关键词***

***数据集 review rebat 大模型 微调 预测 生成 伦理 学术主席 双盲评审 SM LLSTM 辅助驾驶 open review 论文 模型 rebuttal 任务 效果 辅助***

***全文概要***

***讨论聚焦于运用AI和大模型于学术评审过程，特别是review和rebattle的生成。涉及关键点包括：2023年提出的关于态度反驳生成的新数据集与任务，现有数据集与模型的局限性分析，如数据质量、规模及模型处理长文本的能力；模型生成方式、评估指标以及在大模型上的微调需求；关注模型生成内容的偏见、质量和伦理考量，提出通过添加水印表明来源以避免争议；强调AI系统应作为辅助人工决策的工具，而非替代评审过程。***

***讲话人1 - 0:33***

***Hello hello，请。***

***讲话人2 - 0:35***

***喂徐。***

***讲话人3 - 0:38***

***总好徐总好。***

***讲话人2 - 0:39***

***你好你好，学弟志豪。***

***讲话人1 - 0:43***

***我以为你叫志鹏。***

***讲话人2 - 0:45***

***我说志豪好吧。怎么说？要不志恒你直接共享一下你那个文档，然后讲一下。反正我们也是第一次看一下这个为的结果，然后决定一下我们下一步怎么做。***

***讲话人3 - 1:02***

***行，我看一下。***

***讲话人2 - 1:06***

***你那个课上完了吗？***

***讲话人3 - 1:09***

***还有两天。***

***讲话人2 - 1:10***

***好吧，你先上课。***

***讲话人3 - 1:14***

***看一下。***

***讲话人2 - 1:22***

***嗯。***

***讲话人3 - 1:36***

***能看到吗？***

***讲话人2 - 1:38***

***我好像现在看看不到。***

***讲话人3 - 1:41***

***我看一下。***

***讲话人2 - 1:44***

***线哪看到了。对。***

***讲话人3 - 1:46***

***了，主要是第一个是调研的是一个数据集。他这个数据集就是他他自己做了一个19年到20年的。***

***讲话人2 - 1:56***

***对。***

***讲话人3 - 1:58***

***然后他这里做的就是基于之前好像有一篇就是这里这个18年做的对和这个做的，他把它精细化了一点，就是把它这些全部都筛了一遍，然后进行一个分类。对，然后但他这里他这里就是他的一些优点，这是它不足。然后优点就是有点就是在这上面，他它是人工和那个机器同时分类，就是先让机器筛一遍，然后再拿人工再进行看一遍，然后就行不行。对，然后这一篇就是单单做了一个数据集，就就没做啥，也没有做配对，好像就全等一下。***

***讲话人2 - 2:55***

***我问个我我有个疑问，就是你说的那个他拿机器筛了一遍，然后再拿人工筛了一遍，是要筛什么？筛他的申稿人的态度吗？***

***讲话人3 - 3:04***

***这个就是好像有很多，就比如说他的一些有些审稿人的那些，他的review是不是很不是不是不是很清楚，或者是说的就不是很明确，然后他可能就是把他那个进行了一点数据的处。***

***讲话人2 - 3:27***

***理。***

***讲话人3 - 3:31***

***然后这一篇但这一篇就是比较老，我觉得我们做肯定就是这个当时可能就是刚刚起步，所以就他做的简单一点也没有关系。但后面几篇就会做的就是稍微可能会更详细一些。然后这一篇是看一下这一篇就是千里雪上推给我那一篇，然后这一篇就是相当于最新那一篇。***

***讲话人3 - 4:09***

***然后他对对对对对，他刚开始就是用了一个p df转成纯文本，还有一个latex，先把它把那个论文给转过来。对，然后他这个是直接拿这个拿p df转过来的，然后他我看一下他好像他有没有把反驳出去。嗯。他这个是很明显就是在大模型上微调，这然后来做一些就是做这几个任务，就是录用的预测。先预测用不用，然后再看他的分数，然后再把它生成就是评审生成这一指，就是根据它的现有的现有的论文，然后来是他自己先预测一下他这个review会回答一些什么问题，然后这样子就是能更好的帮他进行一些反驳。他自己自自己就是自己会进行一个预测。***

***讲话人2 - 5:21***

***他相当于他就那个读了PDF之后，然后自己先提了一些review。***

***讲话人3 - 5:27***

***对。***

***讲话人2 - 5:27***

***然后但是它是有真正的ground truth review是吧？***

***讲话人3 - 5:30***

***对，因为他爬了，他好像但他好像他这个瑞巴图到底爬没爬。***

***讲话人2 - 5:37***

***他是的什么会议呢？Paper是他叫。还是lp的。***

***讲话人3 - 5:48***

***好像可很多都有，这边好像很多都有。对，然后他这个review做了也做了一个数据集，他re好像也做了一个数据集，然后他review之前就是有一个就是这个是做的可能算比较好的，因为你容量看比较多，就是皮尔瑞瑞一，这好像是19年做，这个是皮尔瑞他这个version two它应该是21年做的，然后都同一个团队做。然后他就说他做的数量会比这个大，他就是这么介绍他这个数据集的然他瑞好像也做了一次。然后他这个好像有一个比较，我看一下。他这个就是有一些多轮回复合并为单轮对话，他就是嗯这一点它是比较有创新性。它说他先是有一个review，那个作者那个审稿人有个review，然后作者会对他进行一个rebattle。然后rebattle的同时，那那个审稿人也会回应这个rebattle，然后再提出一个新的review。然后他就把这些多轮对话合并成了一个文章，就相当于变成一个数据集，这样子就不用处理那么多组，就可能做一个总结。然后还有一些催促性的一些话，比如说就是叫沈稿龙快点，能不能快点回，或者是主要就是这个，好像就快追促的这种，他把它删掉了。***

***讲话人2 - 7:30***

***对，了解。***

***讲话人3 - 7:32***

***然后。***

***讲话人2 - 7:37***

***然后好，这边论文它是就做了一个数据集吗？***

***讲话人3 - 7:43***

***这篇论文我看一下这篇论文。对，好，现在就做了一个数据集，然后进行一个微调给大模型。***

***讲话人2 - 7:51***

***是它微调之后，那个大模型是用来做什么的？是用来出那个reviews的是吧？***

***讲话人3 - 7:58***

***对，就是来进行re它，它也会生成review，就是这些就是他那个任务类型。然后这。***

***讲话人2 - 8:08***

***对诶所以这个它就相当于他已经做了基于la马3.1听了一个模型，然后这个模型它可以分别听用不同的模型，这就不同模型又能review又能rebattle是吧？***

***讲话人3 - 8:25***

***对，就选择不同模型。但是对，然后他用的是好像用了四张的A180G的对，这上面。***

***讲话人2 - 8:38***

***合理的。***

***讲话人3 - 8:43***

***然后这个就是很多人也这个review有些review就是它里面说是有些大模型，很多人那些直接丢给大模型来做的。然后大模型就是有一些偏见或者是但我行会生成的那个质量不是很高，所以他们就可能要对这的这些数据进行处理。但好像这篇没有处理，他只是提到了这个，看一下。然后好像现在基本都是双盲评审，就是计算机里面好像没有弹盲的。对对对，所以很多做欧布图好像也没法做。然后这一篇就是后面我链他这一片的链接过来，我看到这一篇，我看下这一篇。这一篇是它主要是用。用SM来做的，但是。***

***讲话人2 - 9:45***

***我这是哪一年的论文？***

***讲话人3 - 9:47***

***对。他我看一下，他这个主要是他这个只罢了那个I clear的会议。对，然后他这个主要是想做一个那种review和rebat的一个匹配，他就相当于做这个一个小的一个子任务。然后。他这有个框架，然后提出来这个框架。他说他其中这个论文里面提出来这个框架他其实是他先提了第一个框架叫buy LSTM，然后by STM里面有两个比较重要的东西，一个是TLSTM，然后它是用来标记这些，这是也很长的论点里面的，就是很长的这些一些评论性的话，然后里面的一些重要的论点是用来标记这个的。然后S贴SLTSLSTM就是用来来进行一些协调。就是嗯论点和论据匹配之间他要做一个任务来来进行协调，把它们两个匹配起来，这个最喜欢做的这个了。***

***讲话人3 - 11:11***

***然后协调完以后，他又提出来这个框架。这个框架是主要是去掉了T的那一部分，提那部分就是标记论点的那一部分。那一部分就是去掉以后，他说在一个下游任务没有什么影响，他就说这个他就可以断定这个S是S是有协调能力的。然后好像基于理解能力，它就是理解的一个子人物，它就是会下降很多，就相当于做了一个那种骁龙实验。***

***讲话人4 - 11:49***

***嗯嗯嗯。***

***讲话人3 - 11:53***

***然后就是第一个世界是by LSTM，是一种分层的LLSTM它去掉以后就是没有分层的。我有那。***

***讲话人2 - 12:07***

***没关系，这个你要讲不清楚，也不用讲那么细，不用讲那么细的细节，我们先大概过一下，就是做re的他们的数据集怎么做的，他们做了什么任务？包括因为我觉得因为我后来想了想，你直接说自己是做p review里面这个过程里面的rebattle的话，很容易被喷。我感觉。***

***讲话人3 - 12:28***

***对对对。***

***讲话人2 - 12:31***

***特别是在现在这个时间点，现在。***

***讲话人1 - 12:33***

***为啥会被喷呢？***

***讲话人2 - 12:34***

***就是你像前段时间刚出的那个，虽然你的那是就比如说刚出的那个注入攻击，大模型那个注入攻击虽然你跟他没有关系，但是你用包括用大模型审稿和用大模型来回复，都是稍微涉及一点学术后端的这虽然没有关系，但我觉得。***

***讲话人3 - 12:53***

***对对。***

***讲话人2 - 12:54***

***我就我觉得怎么讲这个事儿比较微妙。比如说你可以讲我们不做的不是一个reital system，我做的是一个因为review它这个review其实最初初衷是沈广人帮你改你的论文，就是他给你提意见，然后你修改你的论文的叙述或者实验设计之类的，然后你把论文改的更好。到接收的那个boline。然后我们如果不讲我们是rebut怎么办？我们换个角度讲讲成那种比如说我们system，它可以根据提出的review，然后来提出一些论文的修改意见。***

***讲话人1 - 13:33***

***但是review本来就是修改意见。***

***讲话人2 - 13:36***

***review本来就是修改，也不一定。他review他可能只是提问题。***

***讲话人1 - 13:40***

***我知道，但是里面有个很关键，就是你要给沈可荣提给作者提修改意见，就不仅仅是问问题。***

***讲话人2 - 13:50***

***对。***

***讲话人1 - 13:51***

***所以你如果说我还是告诉你，那你直接看review好了。那你就是把review做一个summer，能做这件事情的意义是什么？***

***讲话人2 - 14:11***

***不知道我是不知道，就是做尾巴头会被喷。***

***讲话人1 - 14:14***

***被喷，没有什么可怕的，有什么。***

***讲话人2 - 14:19***

***做这个无罪声明就可以。就只是说然后在在论文里面写一句，并不用大模型的尾巴导。***

***讲话人1 - 14:26***

***致它可以你可以给模型的时候里面加水印，就是让别人知道这个东西是模型生成的。所以他只是给你一个参考。***

***讲话人2 - 14:37***

***行，之后你先讲，看一下其他的比较相关的工作，也不用讲太多技术细节。***

***讲话人3 - 14:44***

***对，因为我也自己也考虑这个伦理，我一直觉得只能说是帮助人工来进行V八头。对。***

***讲话人2 - 14:51***

***只能这么讲，当原理这块肯定是要重点强调的对，确实有一点那啥。***

***讲话人3 - 14:58***

***然后这一篇就是台大作，但这一篇我当时就觉得一个点是比较好，他就是做了一个做了一个三层的关系，就是他他作者审稿人做完以后，他还考虑到那个学术主席，最后进行了一个原评审，就是这一点。但是那个图我就没怎么看懂，他那个图它生成那个图就是这个，我当时就没怎么看懂。I do, 对我当时就我当时想是能不能做一个agent类似这种好像也没有做这种。***

***讲话人2 - 15:38***

***就是散散可以直接你可以直接点进去那个论文看一眼你说的是啥图。因为你说三层关系，其实我也没有听懂，是啥东西可看一眼。***

***讲话人3 - 15:46***

***看一好像这个有点卡，这个好像。***

***讲话人2 - 15:51***

***有点卡。那算了，我直接看你文档，要不。***

***讲话人3 - 15:54***

***应该挺快的，这个验证一下。***

***讲话人1 - 16:27***

***话说之前很多这种review的，没有人去challenge这件事情的能力问题。***

***讲话人3 - 16:35***

***好像大部分做了一个。***

***讲话人4 - 16:37***

***数据集。***

***讲话人2 - 16:39***

***review。***

***讲话人1 - 16:43***

***那review才是真正的，我觉得是问题很大的。***

***讲话人2 - 16:47***

***对。***

***讲话人1 - 16:50***

***我觉得rebat就跟驾驶一样，它就是一个辅助驾驶的这功能。但review就不是，review就有一种不负责那种感觉。***

***讲话人2 - 16:59***

***okay.***

***讲话人1 - 17:01***

***所以没有人谴责这件事情。***

***讲话人2 - 17:04***

***我觉得应该有。而且review重稿的论文不多。那些agent里面review中稿的好像真的不多，看了几篇都是archive，也没有更新自己种在什么地方。***

***讲话人1 - 17:15***

***Okay. 他们有open review吗？就是这篇文有没有人签了这件事情。***

***讲话人2 - 17:24***

***open review我想想那两篇论文是什么？***

***讲话人1 - 17:29***

***怎么图的？***

***讲话人2 - 17:32***

***就是我好像有一篇是ni是，ni只能看到最终的。***

***讲话人1 - 17:40***

***我记得ni中了之后不是是可以的，可以穿到所有，然后不中的话你是可以选择。***

***讲话人2 - 17:46***

***我我没有我没有印象，我有没有看到过那个open review上个对于这种论文的review。***

***讲话人1 - 17:53***

***Okay.***

***讲话人2 - 17:56***

***这咋打不开？打不开。***

***讲话人3 - 17:58***

***我看看就这个图，就我打开我是下载下，早就下载到这个上面。然后我就把它那个把它原言文件打开，就主的这个图。***

***讲话人2 - 18:14***

***这个图我现在看到是文档打开失败，是为啥？***

***讲话人1 - 18:17***

***行。***

***讲话人3 - 18:18***

***我知道我知道，我没有共享那个没有共享全屏，我只共享那个网易的。***

***讲话人2 - 18:23***

***行。***

***讲话人4 - 18:25***

***我把它调一下。***

***讲话人1 - 19:01***

***还有没有一种可能性就是如果我们做的是一个。辅助型的话，我可以先给一个初始的这种回复，对吧？怎么说？然后让让他去给这个初始性的回复去做一次完整的润色，就提出来他的问题在哪里，什么什么之类的。然这个初始性的回复，我们可以用大模型去生成。***

***讲话人2 - 19:31***

***你这个初始回复指的是最开始先生成一个比较粗糙的retle。***

***讲话人1 - 19:35***

***对然因为这是为了训练，但其实这在真实环境里面肯定是人先生成一个，然后我会去让这个大模型去指出来这个人的回复哪里有问题。然后再给你一个润色后的回复。但感觉本质没有变质。***

***讲话人2 - 19:58***

***对我原来本质好像也没有变，就是把你把work里面东西拆出来了。***

***讲话人1 - 20:01***

***不是就是不是，是说让人还是要先，他这波是强行要求人必须先有一个思考。***

***讲话人2 - 20:16***

***这个图overview model architecture，这个图我也看不懂。***

***讲话人1 - 20:23***

***啥玩意儿？***

***讲话人3 - 20:24***

***它这个记比较多，就好像是看一下。***

***讲话人3 - 20:36***

***他说红色代表是文档内的关系，但我也不知道他文档内关系是啥。然后蓝色是就是两个人讨论的一种关系。然后如果跨讨论，就是比如说谈到别的论文了，比如说我抨击你这个可能别的论文已经提出来了。它就是用绿色来连接。***

***讲话人2 - 20:58***

***这个A是什么？***

***讲话人3 - 21:00***

***AU就是它展示的一种方法。他说这个节点，如果这个论证就review是给他染绿色，然后瑞八头就给他染黄色。就这个对。***

***讲话人4 - 21:15***

***然后这个边是有颜色的。***

***讲话人2 - 21:19***

***考抽象，他们之前做的是设计了model architecture是吧？***

***讲话人3 - 21:24***

***对，他也是那种就是他是想生成那个原评审，最后的他想把这其他之前兴趣全部总结，然后设计一个child，相当于给这个做一个就是给他做一个最后的一个判断。***

***讲话人2 - 21:46***

***了解。Ok ok行，你先你要不先看其他的。***

***讲话人3 - 21:55***

***对我看后面。***

***讲话人2 - 22:03***

***Does my.***

***讲话人1 - 22:07***

***然。***

***讲话人3 - 22:09***

***这一篇做的比较早，我就觉得也是做的挺相当于只做一个机器学习，那够初步的一些数据处理。然后他只是给出了这些结论，所以这篇我看这篇。***

***讲话人2 - 22:22***

***这个数据爬的是AL2018的配装AL了吗？居然。***

***讲话人3 - 22:28***

***对中了好像，但反正官网是有，不知道他中文是什么赛道的。***

***讲话人2 - 22:38***

***Is out.***

***讲话人3 - 22:43***

***Oh Oh is.***

***讲话人2 - 22:47***

***但这个确实好找，数据都是2018年的对对。***

***讲话人3 - 22:56***

***这。***

***讲话人2 - 23:00***

***嗯。***

***讲话人3 - 23:02***

***对，这一片当时可能没什么人做大可能就比较好。***

***讲话人2 - 23:07***

***我做了个分析，之前方法媒提对，好吧。***

***讲话人1 - 23:10***

***就是我因为没一直没搞过AN他么感觉ACR我一直感觉ARP会很奇怪。是的，很抽象。***

***讲话人2 - 23:19***

***是很抽象。那天浩博不是说说有什么做情感分析，加上个十六星人格分析，那那都能中，就很难评价。***

***讲话人3 - 23:28***

***对，就故事讲的好是。然后他这个。他这里也是其实分析，然后他就搞了一他方法讲的是这个就是做一个这种一个网络，就是说他他从心理学来介绍这个。对，他说有个从众心理，就是如果一个人给你打打打那个打那个低分，一个人给你打高粉，他可能就会影响到其中一个人的判断。然后对，但这种能看到吗？我不知道能不能看到，这种就是在实际上打分的时候，因为是多个评审人给一个作品打分。对，我就不知道这个能不能就这种他说有四个关系，这两个是有风险的，然后这两个是没有关系的，所以他就用这个方法来分析。***

***讲话人2 - 24:26***

***这个论文。它甚至分类都不是cs的，能看什么social network。***

***讲话人3 - 24:30***

***对，感觉就是这种那种社科类的。然后对，但其他好像又没人做了，基本上但这个好像居然被引用了，这个好像是被这一篇给引用。***

***讲话人2 - 24:43***

***对，我们我们要是做的话，估计也都得引一遍。对，没错。***

***讲话人3 - 24:50***

***其他就没啥。然后这两天是我看了一个关于那个片见的，就这一篇引的好像，然后这一篇是提取论源，就可能。***

***讲话人2 - 25:02***

***对。***

***讲话人3 - 25:03***

***好。目前我就觉得可能做个，但A真的可能就写可能不是很好写。因为A真的就相当于没有基本上没有给参考，相当于可能就是比人类更要直接来决定这个瑞巴头的事情了。所以这个可能伦理就对于agent伦理就可能不是很好写。***

***讲话人1 - 25:41***

***Ok所以总结一下，现在到底有多少数据是用直接用的，还是说要重新收集数据。***

***讲话人2 - 25:51***

***我感觉能直接用的可能就是他前面提提的两篇。第一篇我记得有一篇是主要收集的是ERP还是acc l的数据，然后有一篇是爬了很多。***

***讲话人3 - 26:02***

***对。***

***讲话人2 - 26:02***

***好，这篇是不是？这篇也爬了艾克里尔的那是哪一篇？只拍了ACL还LP的数据集，就是找能用的数据集，这个就很好找。对着这两篇找他们的来源，然后找引他们有没有新的数据也要找出来了。然后就是我我们要讨论一下开座的话。***

***讲话人1 - 26:25***

***作画数据也是第一位。***

***讲话人2 - 26:26***

***这比如他数据是第一位。***

***讲话人1 - 26:30***

***他的数据里边是不是年份、会议都是比较那啥的，比较清晰的是吗？不是比较老的。***

***讲话人2 - 26:38***

***比较老的。***

***讲话人1 - 26:39***

***就我意思是说他是不是不够新，那你肯定小心。以及比如说我们之前讨论过这个数字本身是不是真的是我们之前不是说过一个话，就是那种降的retle，不应该在我们那里边，对吧？你肯定要做数据清洗。然后这些东西我觉得应该是第一个就是你用哪些数据去做这件事，这是应该讨论的第一步。***

***讲话人2 - 27:06***

***嗯。***

***讲话人2 - 27:12***

***我有个问题，就是在open review上能看到降分和提分的那个过程吗？只能点开那个history是吧？***

***讲话人1 - 27:21***

***Open review就是我。***

***讲话人2 - 27:23***

***记得因为最后给我们看的那个界面，就是最后最终那个分数，然后我们知道他讲涨分或者降分了，还得知道他之前是什么分数吗？***

***讲话人1 - 27:33***

***对。***

***讲话人3 - 27:33***

***这个我还真可好对吧，我的这个。***

***讲话人2 - 27:36***

***要这个revision history好像。***

***讲话人3 - 27:39***

***这里有批吗？这里有一片就是说他能找到。***

***讲话人2 - 27:44***

***ok那得看他们怎么找的这个对。***

***讲话人3 - 27:46***

***我看一下。***

***讲话人2 - 27:48***

***能获取也行。我想的因为我想的就是如果你直接爬网页的话，只能看到最后的分数。***

***讲话人3 - 28:08***

***After.***

***讲话人3 - 28:37***

***好，那这里没有这一部分没提到。***

***讲话人1 - 29:17***

***现在是在查什么东西吗？***

***讲话人3 - 29:20***

***我想知道他那个是怎么他那个分数是怎么截下来的。***

***讲话人1 - 29:25***

***我觉得这个可以下面再查吗？***

***讲话人3 - 29:26***

***再好像他这个是他好像直接问他这个人我我也不太清楚这。***

***讲话人2 - 29:34***

***我觉得可以先考虑一下比较high level的事情，就是我们要做一个什么样的任务。***

***讲话人1 - 29:41***

***我想想问我们跟他的现在里边应该没有这种完整的做re的对吧？***

***讲话人1 - 29:54***

***就比如说你列的1.1到1.6，它里边有哪个是跟我们的。***

***讲话人2 - 29:59***

***任务是一，我们跟我们那初衷是一样的。***

***讲话人1 - 30:02***

***对。***

***讲话人2 - 30:03***

***1.1和1.2应该最相关。他们俩是做的啥？你点开点1.1，我看一眼。说法哎呦。Exploring什么argumentation for writing peer review rebattles这个估计也是生成。行，点还挺快的。***

***讲话人4 - 30:26***

***你可。***

***讲话人2 - 30:28***

***以直接看到abstract做个什么任务。In this work, we are the 1​st to explore这个是啥意思，什么argument for peer review，by proposing the novel task of attitude rebuttal generation.***

***讲话人2 - 30:48***

***我操工具在哪？这个单词是啥意思？这个动没吗？不是我我的我看一下我的笔在哪儿，这个单词啥意思，能看到。***

***讲话人3 - 31:03***

***吗？这是他那个名字，我也不知道他为什么要叫这个名字，对吧？就只那个对，是它就属据极致，好像就这。***

***讲话人2 - 31:13***

***你看他们反正在abstract里面说in this work为巴拉巴拉巴拉，然后做说最后一句，我们是一个啥rebutgeneration的任务。***

***讲话人3 - 31:26***

***你往，对我看到了，看到。***

***讲话人2 - 31:28***

***对对对。***

***讲话人3 - 31:32***

***然后有两个字，然后。***

***讲话人1 - 31:40***

***这二三年的。***

***讲话人2 - 31:42***

***这也是发这中了ELP对。***

***讲话人1 - 31:53***

***Ok那他这个好像也是个button的这个是吧？***

***讲话人2 - 31:57***

***对。这篇可能这篇，然后另外一篇蕊芳1.1和1.2，这两篇可能是比较相关的。***

***讲话人1 - 32:12***

***就是我我最大的问题是，如果别人做过的话，我们做这个还有什么意义？***

***讲话人2 - 32:18***

***如果别人做过的话，他们就你得看他做成什么样了。因为这个事儿你想这篇论文如果这衡部调研的话，根本不知道有人做过a resystem。***

***讲话人1 - 32:28***

***你意思说他们相当是做到效果还不够，不是足够好是吧？***

***讲话人2 - 32:32***

***我不知道，如果我我们把效果做到能用。***

***讲话人1 - 32:35***

***就能真实用。***

***讲话人2 - 32:37***

***能真使用。***

***讲话人1 - 32:41***

***Ok.***

***讲话人2 - 32:42***

***因为我就是我今年其实在那个AL稿论文上，我也看到了很多那种review。像就是关那个论文标题里面带review的，然后大家都中了。我就很好奇他们是以什么角度去写这个论文的。***

***讲话人1 - 32:55***

***那不是能写下来看吗？***

***讲话人2 - 32:57***

***对，可以看我看了就我搜了一下关键字，大概十几篇。***

***讲话人1 - 33:00***

***我看看过。其实我感觉就是从侧面，比如说去说理解的这个问题，然后你跟真人去那个啥，就是能能对对对，兑现的足够好，才说明你的这个能力更像人，对吧？女巴特别是这种就是你是真的跟要跟审稿人，然后那个啥对他不是一个固定的，之前的别人打的那身上打的那些脏标签什么之类的，他就是跟真人对话，然你要还，而且你要说服他，而且这个东西本身逻辑性就很强，这毕竟是re巴通。对，也可以从这种角。***

***讲话人2 - 33:41***

***度去写。***

***讲话人1 - 33:44***

***一个方面。然后另外一个方面就是辅助这个科研人员的但这个伦理问题肯定要考清楚就好。但我觉得这些写还是另外一方面，主要是还是说我们要做成什么样的，对，就他这个现在是一个什么样的一个level，然后你想要的一个什么，就是他现在的这个re然后有什么问题，就是他为什么他这个东西没有人就没就就就没有没有没没有被广泛的使用了。是因为它性能很差，还是什么情况？我觉得这个问题可能要分析，比如说他是不是他有，如果有这个的话，你是不是得看一下，他可能也会有一些demo，或者说可能会有甚至会别人的讨论。这个可能得看一下。***

***讲话人2 - 34:29***

***他们这个任务，他们这篇好像就提了个数据集。***

***讲话人1 - 34:32***

***没，就是这个数据集。***

***讲话人2 - 34:34***

***你看就是他们inconsistency，就是他们的abstract里面写三点都是关于数据的，limited data set、inconsistent loquality data, insufficient support, 全是关于数据的。Ok所以能比较大，就是提了个数据。***

***讲话人1 - 34:53***

***我看另外另外一篇。***

***讲话人2 - 34:55***

***另外一篇就是EMRP那篇，他们也有一个配套数据集，但是应该是做了regeneration这个任务，所以效果但是效果怎么样不知道。***

***讲话人1 - 35:04***

***对所以问题是他的generation是单问题的解决，还是说一次全给你那个啥，就是你去把问题全都喂给他，让他一次出全部。***

***讲话人2 - 35:18***

***对。***

***讲话人1 - 35:20***

***那这是普recall。还有比如说他的这个指标是怎么算的？***

***讲话人2 - 35:25***

***Recall就很那啥就相当于是他们爬了之后，然后就觉得作者的re就是光truth，然后就使劲往上面拟合。也好像也挺合理的，因为我们也是到时候要看。***

***讲话人1 - 35:39***

***但是你算这个S是怎么算的呢？是按这种，但因为单词上一般我们也不算Proc和record对吧？***

***讲话人2 - 35:45***

***对。***

***讲话人1 - 35:46***

***但是所以我不太清楚像position record是怎么算出来的，也很奇怪，我也可以看。***

***讲话人2 - 35:52***

***对。***

***讲话人1 - 35:55***

***对我甚至还在想。就是能不能让那个模型去学习。如果它是辅助我们把这篇paper做的更好，那是不是这件事情不是应该等到那个review已经出来。那有没有一种可能性，就比如我们做一个single，我说我们写完了篇paper，然后这让这个paper pper给我。***

***讲话人2 - 36:21***

***们提问题。***

***讲话人1 - 36:22***

***其实就是review的过程。对，然后再再再让再再有一个模型去给我们做retle。对让就是让模大模型帮我们发现我们论文里面的问题，以及的解决的方案。那相当于是我们有一个我式，同时要做review和re对，对对对，而且是要做多轮。对了，你比如你改完之后，他可能会发现新的问题，然后直到这个东西能收敛，有点像干一样那种感觉。对，然直到这个收敛，然后这样的话就相当于把这个东西真的是辅助到了科研。在在前期就可以辅助科研，而不是仅仅是我在后期做了一个头的这个工作。就相当于是我做一个system，这个system可以同时做review rebattle和re这一paper的循环。***

***讲话人2 - 37:14***

***对的，就是。并可以弱化re的功能。***

***讲话人1 - 37:21***

***不但我觉得第一次肯定做出这么多，就是你一次肯定很难把这三个同时做完的。***

***讲话人2 - 37:27***

***对。***

***讲话人1 - 37:30***

***要做就可能得做成一个类似于系列的一个东西。你先做retle，然做完re之后再做一个review加re的功放的那种感，然后再去做一个，再加上原文refine的这种。***

***讲话人2 - 37:51***

***对。***

***讲话人1 - 37:58***

***Ok那觉还是先说button这件事情。对，那button这件事情的话就是我觉得一个是得先让志恒去分析一下这个东西，就刚才说这篇文章他的问题是什么，对吧？那你肯定得发现，毕竟他做了简单的事，因为他是怎么样生成的，以及是什么样形式生成的，生成什目的什么。然后大概的回复，我不知道他有扣的没有，你可以拉上面看看有没。***

***讲话人3 - 38:28***

***有这。***

***讲话人2 - 38:29***

***好像是有githugithu的。***

***讲话人1 - 38:32***

***Ok你以拉拉就是别人别别别人别人提出来的那个你说意见也好，还是什么？Star star就五个。***

***讲话人2 - 38:43***

***对，这个工作那天我搜到的时候很震惊，就没有什么影响力。***

***讲话人1 - 38:48***

***那说明他这re肯定做的不够好吧。***

***讲话人2 - 38:53***

***我没有仔细看，我都不知道他做的是什么，你也不知道他做的是什么re头.***

***讲话人1 - 38:58***

***ok所以我觉得正好得把这细节。***

***讲话人2 - 39:01***

***就是你想他们现在都是好像是做了一个模型。单模型的话你很难也很难接受是的，用文这种长文本的输入。所以你其实是缺少应该是它大概率缺少context。那他就是他说分析的那给的那个结论可能也就是硬去拟合那个数据集。***

***讲话人1 - 39:19***

***就变成发化性很弱呗。***

***讲话人2 - 39:22***

***对我我是这么觉得可以，但是得仔细看一下。***

***讲话人1 - 39:24***

***对，对他这个代码我怎么感觉上面没有代码，只有read me。再再往前一点。***

***讲话人2 - 39:29***

***你再往上拉一点之后，扣子可能里面有。***

***讲话人1 - 39:32***

***这个真的有吗？***

***讲话人2 - 39:34***

***那五个star说明确实可能就要么不能用。***

***讲话人3 - 39:39***

***对，感觉没优。***

***讲话人1 - 39:41***

***是吧？***

***讲话人2 - 39:44***

***还是有代码的。***

***讲话人1 - 39:47***

***它还是T5，什么bu啊这些都是很老古老的东西。***

***讲话人2 - 39:51***

***是的，已经是上个时代的产物了。***

***讲话人1 - 39:54***

***对，所以肯定他应该不太好。这个其实跟他对比很好对比。对，因为就他的就是你随便拿出来两个问题。***

***讲话人2 - 40:04***

***1 generation就知道。***

***讲话人1 - 40:06***

***对，就是你自己的模型跟他的模型就区别肯定很明显的，就能对比出。Ok ok我觉得那你第一步应该是一个是把这个东西再仔细看一下，这篇就是两篇非常相关的论文。对比如数据那篇文，它数据有什么问题，然后这篇填model的，它model有什么问题？对，然后再去看完这两个之后，再去决定一下要不要去收集新的数据集，或者说还是直接把他的这些数据直接用起来，让我们去做一些要不要做一些过滤这个对吧？***

***讲话人2 - 40:47***

***对你主要还是看前两篇，第一篇效果肯定是不太行。***

***讲话人1 - 40:52***

***第一篇不是没有写。***

***讲话人2 - 40:53***

***第一篇不是只有数据集吗？第一篇只有不是，第二篇是个只是个数据集1.2只是个数据集1.1是有有扣的。***

***讲话人1 - 41:00***

***Ok ok对，所以就是得把前两篇仔细看一下，然后去列一下这个数据集。我们要他是这个税，这是24年的是吗？***

***讲话人2 - 41:11***

***这个是刚投的，好像瑞方。***

***讲话人3 - 41:13***

***这个吗？***

***讲话人2 - 41:15***

***我记得是刚投的，是刚才挂出来的2K5。你点一下。***

***讲话人3 - 41:19***

***对这个论文按12505.***

***讲话人2 - 41:22***

***刚投的应该是ni投的。***

***讲话人1 - 41:24***

***Ok那我觉得这个可能他还不。***

***讲话人2 - 41:27***

***这个是最新的。***

***讲话人1 - 41:29***

***Ok那我觉得他有可能他它release是吗？***

***讲话人2 - 41:34***

***我记得release这个是还是浙大做的，浙大计算机系的主任。***

***讲话人3 - 41:39***

***我那。***

***讲话人1 - 41:41***

***ok他如果有数据集的话，那很有可能我们确实是不用从头收集手机，可能只用对查一下就行。***

***讲话人2 - 41:50***

***Agent review clear这是哪？就这个table里面比还挺全的，可以再把table里面那个看一下。***

***讲话人3 - 41:57***

***这些吗？***

***讲话人2 - 41:58***

***对，就table里面比这些，什么agent review，然后deepreview这个中了acl我刚看到了。***

***讲话人3 - 42:05***

***对，这一篇我之前看，但是他之前这这边这边纯做review的。***

***讲话人1 - 42:10***

***对，所以review都能做的话，你不用担心re 8行，我觉得你还不如担心能不能把这个东西做到让别人惊艳。***

***讲话人3 - 42:22***

***这怎么去。***

***讲话人1 - 42:25***

***Ok ok那我觉得就先看一下个那个那个方法本身的问题。对，然后这个数据集我觉得可以先那个啥先也看一下这数据集大概多大。***

***讲话人2 - 42:40***

***它这数据节规模是多大。***

***讲话人3 - 42:41***

***你看一下。***

***讲话人1 - 42:49***

***我记得是几万篇票票应该也没多大。***

***讲话人3 - 42:55***

***Thank you.***

***讲话人2 - 43:07***

***有多少？我这个像不到。***

***讲话人3 - 43:09***

***差不多2万，这里写的2万篇，然后这是V6的这是V8头的，有一些很多人没有做对巴头的数据。***

***讲话人2 - 43:20***

***确实。***

***讲话人1 - 43:28***

***不过说实话就是没有数据集就会降低了一些过性。***

***讲话人2 - 43:34***

***是。***

***讲话人1 - 43:35***

***啥意思？就是如果有数据集的话，自己能提出一个数据集是最好的。就数据集加算法，那纯粹是一个算法的话，肯定是没有数据其他算法的那事。***

***讲话人2 - 43:47***

***对，所以还要在他数据集上做进一步处理。***

***讲话人1 - 43:50***

***对，所以就可能就这还是得给一个新的税籍。所以你得看一下这个数据有什么问题，你觉得。***

***讲话人2 - 43:56***

***对他们处理器真的太远了，看一下。***

***讲话人3 - 44:01***

***还这一批。***

***讲话人2 - 44:12***

***好像是开了。我记得在那啥上面可以看到，在给face上有他们的数据，就他们代码放在那个匿名仓库里面了。有的71.9K的train数据，conference还有年份什么的都有。然后。有的。***

***讲话人1 - 44:39***

***对我觉得可以先下载到本地看一下，这个东西应该不会很大，因为它都是文字。***

***讲话人2 - 44:44***

***对，三百多兆很小的。***

***讲话人1 - 44:46***

***所以我觉得那个谁志恒可以先下本地先看一下这玩意儿。***

***讲话人3 - 44:51***

***这一篇。***

***讲话人1 - 44:52***

***就税及那一篇。***

***讲话人3 - 44:53***

***书籍不是只那一篇吗？那就是那一篇。***

***讲话人2 - 44:59***

***姐我那个链接我放群里了。***

***讲话人3 - 45:02***

***好。***

***讲话人1 - 45:31***

***Ok行行，我们要不快点。中中那就先看这个数据集，先看这个数据集里边的问题，还有模型本身的问题。对，然后模型它里边应该会有论文，里边应该也会给了一些general的结果。然后可以看下它那个general结果，我在想要不要测一下这个模型。***

***讲话人2 - 45:52***

***他这是个模型吗？你说NRPEMNRP那一篇。***

***讲话人1 - 45:55***

***就是有一篇不是有个模型的吗？***

***讲话人2 - 45:57***

***对，有一篇有。***

***讲话人1 - 45:58***

***对我就想要不要测这个模型。***

***讲话人2 - 46:02***

***就可以，你可以拉下之后跑几个例子看一下，感觉可以试试。***

***讲话人1 - 46:06***

***对对对，你本地是没有那个，算了，还是给你一个啥。***

***讲话人2 - 46:11***

***现在有卡了是吗？***

***讲话人1 - 46:13***

***我在想要给他开学，因为学校这个月月底会再租16张，我在想给他这个还是给公司的开一个那个。你还用过服务器吗？***

***讲话人3 - 46:22***

***我用过云就是语音服务器是吗？对对对，我用过。***

***讲话人1 - 46:27***

***对我我觉得直接啥也行，因为现在我们刚好有一个卡在试用，然后直接使蹭蹭一下，这个也可以，因为他之前不是现在临时的。就是有一国家创作中心，然后我们在后面得考虑租他的卡做仿真。因为那个仿真发现很多服务器上没有界面，跑不起来，所以可能那啥，那到时候让让那个盐泰，然后因为盐泰在申请那个，因为就我拉着他那边让先试用一段，然这里边有1000卡时的这个时间。对，可以先用这个还挺多的是吧？***

***讲话人2 - 47:00***

***这是啥卡？***

***讲话人1 - 47:01***

***4090 100都有。***

***讲话人2 - 47:03***

***Ok拉我也进去。***

***讲话人1 - 47:05***

***你你好好你好好用你la吧。对，当时是主要是想试一下他能不能部署那个仿真环境，您系器人，还有那个自动驾驶的。***

***讲话人2 - 47:19***

***对。***

***讲话人1 - 47:19***

***然后我觉得他对他应该是明天申请好，然后申请好之后，然后到时候再那个啥一下，让闫泰把估计会有账号密码之类的东西，到时候登录一下试一下，或者说等颜菜先试，等颜菜先试试。都走走通之后，然后再再我把这个给志恒，让他给志恒一个账号。***

***讲话人2 - 47:39***

***我感觉行。***

***讲话人1 - 47:40***

***ok ok行行好，行，那我觉得可以先洗涤的，先看一下那两篇。***

***讲话人2 - 47:45***

***对对对，再仔细看那两篇，以及他们表格里面对比的那些。***

***讲话人1 - 47:49***

***对，你可以把那个数据集大先的下来先看一下。对，因为数据集里面可能也会，我不知道它里边全不全。就比如说这个reconfidence，这些东西其实也是一个将来你做这个回复的reference确实对，所以你得看一下那数据机里边存的全不全，它如果存在不全的话，其实也是你后面再去弄这个数据集的一个什么一个一个理由，对吧？对，ok行，那千里还有什么其他的。***

***讲话人2 - 48:22***

***我没啥我我我没啥了。我其实我还是在想这个任务该怎么怎么去从哪个是角度去切入。***

***讲话人1 - 48:29***

***你讲吗？***

***讲话人2 - 48:31***

***就是现在现在我们想的不是非常暴力，就是做rebattle。对，rebattle做。***

***讲话人1 - 48:35***

***我觉得想的太早了，你现在reto能有一个好一点的结果之后再去想其他的。***

***讲话人2 - 48:40***

***好吧行吧，行。***

***讲话人1 - 48:42***

***不要想那么长远，就是长远的感觉。***

***讲话人2 - 48:45***

***好，行。***

***讲话人1 - 48:45***

***没事，可还有可能想半天就发现米巴都就做的很不，还剩AA味道很重，东西也不好，那就没什么意义。***

***讲话人4 - 48:53***

***对。***

***讲话人2 - 48:55***

***对。***

***讲话人1 - 48:57***

***对，所以先把东西做好再说怎么去那个啥。***

***讲话人2 - 49:02***

***对。***

***讲话人1 - 49:05***

***o ok那先这样，先这样。行，拜拜。那那那个那个倩你不是还是跟我那啥一下？***

***讲话人2 - 49:11***

***对对对，行，我再跟你说一下。***

***讲话人1 - 49:14***

***行，志恒您嫌麻烦，大家先退一下。行。***