

Diamante 数据标注规范与流程

提交时，后台返回数据会有时延，**请勿连续点击**『提交』按钮，造成重复！

若多个候选回复均符合要求，请选择符合自己偏好的回复进行复制或改写操作

规范更新

----- 规范更新 v5 -----

- 1、标注时，注意标点符号的正确使用，避免错别字；
- 2、标注的对话内容较为集中，请结合日常生活，发散思维；
- 3、标注时，请看完展示的候选回复再进行标注；

----- 规范更新 v4 -----

- 1、模型会生成看似信息量充足但是啰嗦或逻辑不通的回复，请带入语境仔细判断，对回复进行选择、修正精简或重写；
- 2、对于某些话题(科技、文化等)，模型能力欠佳，不应完全依赖模型的回复，需要您引入自己的知识，不要显得对话内容很空洞；

----- 规范更新 v3 -----

- 1、标注时，减少使用**语气词**的频率！（哈哈哈；对对；嗯嗯好的；呜呜；哦）
- 2、避免简单的你问我答。标注的尽量是日常闲聊，而非知识问答；
- 3、标注时，注意对话角色的区分，保持对话角色的人设一致；

----- 规范更新 v2 -----

- 1、聊天话题尽量**发散**，不局限于当下时间节点；
- 2、避免过多地**相约做某事**（今晚来我家吃饭；周末一起爬山；一起买水果吧）
- 3、减少**热点类**的聊天（时事新闻）；避免聊及**过于小众**的事情（隔壁的《唐记甜水铺》开了）
- 4、不当**谜语人**。通过聊天内容，其他人能知道在聊什么（安修打不过王亮青的；你说嘉灿最后会成功吗？）
- 5、避免**尬聊**。感觉聊天不对劲、没内容时，及时转移话题、跳出；

----- 规范更新 v1 -----

- 1、避免涉及**当下时间**的聊天（现在几点了？；啥时候期末考试？）
- 2、聊天具有**普遍性**。避免涉及个人信息的聊天（把你的手机号给我？；你的网易云账号是什么，我关注一下？；我加你吧；）
- 3、措辞**凝练**。标注内容是日常聊天，但不是微信类的闲聊，能一句话说完的事情就不要分成多句；

任务介绍

- **标注一个对话。** 撰写对话起始句，随后的对话便由标注人员展开：从模型产生的候选回复中进行复制、修改或自己撰写等操作，使得对话可自然地进行下去！
 - 若模型回复质量较高，则可直接点击进行 **复制** 操作
 - 若模型回复意思准确，但表述欠佳，则可进行 **修改** 操作，完善模型回复
 - 若模型的多个候选回复均欠佳，则可进行 **重写** 操作
- 整体要求
 - 1) 对话以有内容的日常聊天为主；
 - 2) 整体对话应流畅自然，有内容，能深入；
 - 3) 回复语句需要：表述规范、符合语境、有信息量且符合逻辑。

标注要求

- 对话起始句
 - 不建议涉及功能询问（如，播放小猪佩奇；4 加 5 等于几？）
 - 选择自己感兴趣的话题开启对话，直入主题；也可以结合自己感兴趣的话题，不做限制（如，你最喜欢喝什么饮料呀？；三亚挺好玩的，有机会再去一次！）
- 对话内容
 - 需要从模型产生的候选回复中进行复制、修改或自己撰写等操作，使得回复自然、有信息量且符合逻辑；
 - 如果机器的回复质量欠佳（**存在句子不通顺、违背常识、缺乏内容、消极回应、或与上文冲突 等问题**），标注人员需要自己撰写合理的高质量回复；
- 提交要求
 - 如果感觉对话质量欠佳（**存在句子不通顺、违背常识、缺乏内容、消极回应、或与上文冲突 等问题**），需要点击『重新刷新』按钮，舍弃当前对话、开启新的对话；
 - 回复的长度不少于 5 个字；
 - 完全使用模型的回复的比例不超过 70%；
 - 一组对话的聊天轮数 7-10 轮，至少 7 轮。

特别要求

- **话题转移**
 - 发现对话：趋于平淡、无法有效进行下去的时候，需要标注人员**主动自然地转移到合适的话题上**！
- **积极主动**
 - 聊天双方至少有一方对聊天有渴望，展现出**积极主动**的态度，不要双方都冷淡！
- **有内容，能深入**
 - 不应局限于日常闲聊（吃了没；你好我好大家好），语气词慎用（哈哈；呜呜呜），对话应该有**内容、有深度**！
- **人设一致**
 - 聊天时注意区分对应的人设（说话风格、对应的人物信息），在对话中需要**保持人设一致** 且有**区分度**！并减少附和！

标注流程

- **Step 1 开启对话**：点击『开始对话』按钮，选择并补充感兴趣的话题（可以点击『更换话题』按钮，寻找更多灵感），也可以自己直接发散。

#	Dialogue

Example: 你喜欢喝什么饮料呀？

Submit

Task Description

[标注前请阅读标注流程及规范](#)

与机器一起标注一个对话

首先，根据兴趣撰写对话起始句

随后，你将从模型产生的候选回复中进行复制、修改或重写等操作，使得对话可自然流畅地进行下去！

开始对话

Topic Suggestion

- 游戏 领域，你比较感兴趣的是 电竞主播
- 音乐 领域，你比较感兴趣的是 华语流行
- 体育 领域，你比较感兴趣的是 国内足球

更换话题

- **Step 2 进行对话：**标注人员根据话题兴趣开启对话；随后根据模型候选进行复制、修改或自己撰写等操作，使得对话可自然流畅地进行；

#	Dialogue
1	春天我家猫开始掉毛了，有啥好办法吗，到处都是
2	我都是买吸尘器吸的，没办法
3	你家猫掉的多吗，是啥品种的
4	英短银渐层，现在掉毛，我每天都得吸
5	我猫有时还吃自己的毛团，我好担心他消化出问题
<div>你买个化毛膏给猫吃就好啦，问题不大</div> <div>Submit</div>	

Task Description

[标注前请阅读标注流程及规范](#)

与机器一起标注一个对话
首先，根据兴趣撰写对话起始句
随后，你将从模型产生的候选回复中进行复制、修改或重写等操作，使得对话可自然流畅地进行下去！

保存对话

重置刷新

Candidates

- 我家的也吃
- 不会吧，你家的多大了？
- 是不是缺东西啊？给它买点猫草吃？
- 吃点化毛的
- 吃毛球，你买个化毛膏给猫吃
- 吃毛没关系啊，我猫也吃
- 这个你要注意了，可能缺什么微量元素，去看看医生

注:点击相应的回复，即可复制到输入框！

- **Step 3 保存对话：**点击『保存对话』按钮，完成一次标注，并记录相应的对话 id（作为工作量证明）

本次聊天已完成，请记录对话 id：

0-yOjG25op

点击确定开启新一轮对话

开始对话

Good Case 示例

- Case 1 (对话积极主动)

天天下雨，心情都不太好了

我也不喜欢下雨天，湿漉漉的，也无法正常出门

对啊，一下雨就不想出门，就很down！

但是下雨天比较适合宅在家里看剧~

是的，下雨天躺在床上看剧、聊天什么的最舒服了

哈哈，你都看什么剧

现在在看，请回答1988，强推

没看过，我有时间也去看看

你都看一些什么剧呢？

我喜欢看悬疑剧，《白夜追凶》不错，你可以看看

好啊，我记下了，谢谢啦

- Case 2 (话题转移自然：过敏→疫情)

花粉过敏伤不起

那就别出去了，呆家里吧

春天到了，外面各种花开，出门都只能一直戴口罩

我也过敏，不过我不是花粉，是柳絮，那感觉简直要窒息

是的，鼻子痒的难受

你可以带那种防毒面具，哈哈

那个太夸张了吧，我戴个口罩就可以了，反正现在疫情，大家也都戴口罩

是啊，前段时间因为疫情，我们居家办公了好几周，刚回来公司办公

你们是在上海吗？

不是，我在深圳，最近上海也很多公司居家办公了

- Case 3（有内容，能深入）

你怎么看虚拟现实VR？

VR挺有意思的。我体验过一些恐怖游戏，身临其境的感觉，太吓人了
体验的什么游戏？

一个VR专门的游戏，叫生化病院
我玩过一个叫寂静岭的，吓得半死，后来再也没玩过

那个我玩过，还玩过一个叫死亡空间的，也挺吓人的
你是不是很喜欢恐怖呀？我就很怕

我也挺怕的，可是又很想尝试，就是这种感觉吧。喜欢和害怕是两回事
你有没有看过类似的电影或是什么的？

有呀，电锯惊魂，致命弯道，死神来了都挺经典的
致命弯道讲的什么呀？有点好奇，其他两部我都看过

它讲述一群人在丛林里与畸形食人魔生死角逐
有时间我也要去看看，谢谢啦。你看过咒怨么？

还没看。。。那种恐怖对我来说比血腥恐怖更可怕

Bad Case 示例

- **Case 1** (没内容，陷入无意义的聊天循环)

吃饭了吗？

刚吃过，你呢？

刚吃完，你吃的什么？

食堂吃的，你呢

食堂的饭菜好吃吗？

还行吧，不难吃的

噢噢，那就好。

那你吃的什么？

我吃的快餐

嗯嗯，但是我不喜欢吃快餐

- **Case 2** (常识错误：青蛙腿上没有眼睛；逻辑矛盾：不是瞎子 → 靠视觉)

你养宠物吗，

想养狗，可是我连自己都养不好

我养小青蛙，比狗方便多啦，

你的青蛙腿上有几只眼睛？

两只眼睛

它怎么看路的呢？

青蛙又不是瞎子，他靠的是触觉

青蛙的触觉哪里来

- **Case 3** (人设不一致)

你篮球打的怎么样？

还行吧，参加过学校举办的新生杯篮球赛

不错噢！那你一般打什么位置？

啥意思？位置是啥意思

我问你打篮球一般是打控卫、中锋还是前锋

我不会打篮球呀！

- **Case 4**（不知道合适地转移到有内容的话题）

你别每天和我吵架

我也不想，可是控制不住

你就不能让着我点

我让了，我都不知道让过你多少次了，你还不满足啊

你每次都不让

那是因为你太过分了

我哪里过分你说

你每天叫喳喳

你才喊喳喳

- **Case 5**（语气词过多）

夏天快来了，我又能吃荔枝了！

我也喜欢吃荔枝，不过吃多了会上火

是啊，所以不敢多吃，一次吃几颗，过过干瘾

是啊！每次看到都想吃！不过不敢吃多

好像我喜欢的水果都是容易上火的哈哈哈哈

怕不是菠萝芒果榴莲

对对对，都是容易上火的哈哈哈，但是真的好好吃

呜呜呜流下了不争气的口水，榴莲真的一生推

我都是我姐买的，她是真土豪，一买就买两三个

有个土豪姐姐真好呜呜呜，请问她还认妹妹吗

你可以认我这个姐姐哈哈哈哈，然后我从她那里抠出来给你一口

你走开啊哈哈哈哈哈哈