BIRD的问题，数据太脏，即使是人类也难以分辨，不适合用于判断语言模型的逻辑能力， 改进后问题表意清晰无误且与数据库描述和数据库本身匹配。

例如问题36，明明在Adamname3上写了not useful 却仍然在回答中使用了；要求full name 但回答却是分开给出的

问题83，magnet在表描述中是Y或N，到了实际数据中却是0或1.

有些问题目标是percent但不明确描述，有些会则明确描述

问题94，问题较含糊，目标sql的选择条件是district\_id = ( SELECT district\_id FROM client WHERE gender = 'F' ORDER BY birth\_date ASC LIMIT 1 ) ORDER BY T2.A11 DESC LIMIT 1作为条件，意为：从指定的表中检索地区 ID 与表中年龄最大的女性客户的地区 ID 相同的记录client。在这些记录中，选择列中值最高的记录A11。”这与question严重不符。

问题116，有歧义。目标sql求的是所有满足条件的client的交易总额的变化率，问题问的却是个体的。修改问题。

问题125，问题要求列出district但sql中并未列出。

问题149，答案可以有多种格式，未限定DISTINCT。修改问题和sql。

问题169，要求1996到1997之间的增长率，目标sql的统计终点却并非1996-12-31而是1997的任意一天(更新：没有问题)

问题198，要求的是molecule的数量，结果却是求所有满足要求的molecule中COUNT（bond\_type）的average。逻辑不对

问题212，模型无法分辨non-carcinogenic molecules指的其实是每一种分子而似乎错误的理解为分子数量来自于atom表中，这要求模型具有相关知识。并且事实上并列答案的有四个，但这不可能在查询前得知，因此模型无从得知需要四个输出。很难的问题。

问题220，又是涉及术语。Top 3 Elements不可能包含两个一样的项，这是基于我们对element的理解，然而llm未必具有相关知识。

问题231，molecule表中根本没有TR018，因此问题的后半段根本无法回答，目标sql中也未给出询问。需要这样的问题吗？？

问题263，同样是考察element的理解，到底何时需要distinct何时不需要。本题问的实际是原子数量。

Toxicology系列的难问题过多考察同一个点：对element这个词的理解。问题268，277，281，285，290，302，304，306，307，319，322，328等

问题330，考察对相似列的辨别能力

Cardgames系列的legalities表的status列值描述有问题，真实值为全开头大写，给的文件里则是全小写

问题416，uuid的描述有一个表有一个表没有，并且在某个表写了not useful。并且此问题涉及卡牌规则，unknown power被认为是without power，这不属于一般知识。

问题431，发现表描述文件名是错的（translation变成了transaction）。并且如果只看问题无法得知问题同时需要name和id（暂定）

问题494，语序混乱，并且说明文件的ruling也不对应rulings。

问题507，要求llm对表中值有所了解（其实就是value description没给）

问题586，问题中未要求Title但SQL中有，很奇怪，感觉是生成的数据。

问题701，题目中要求“score above 50”，SQL中却是“score >= 50”.已修正

问题724，缺少表中值的描述，而要求llm能理解（Blue，blue，Blond， blond）

问题·730，第一没有DISTINCT，同样的hero id被重复计数。第二联表时INNER JOIN hero\_power AS T2 ON T1.id = T2.hero\_id 忽略了没有superpower的hero。（super hero中有750个id，但heropower中只有652个hero\_id）

问题743，问题中的superheroes who act in their own self-interest or make decisions based on their own moral code直接对应表描述中的neutral，但目标sql中却是bad？？？

问题772，问的是color，目标sql中却是color id

问题829和问题34同一个问题稍微不同的问法

问题835，联表alignment之后反而少了值（387变384），说明有英雄没有alignment

问题880，问的是最快圈速之差，回答SQL是最快圈速之和之差

问题896，results和driverstanding两个表中有着position这列且意义完全一致，只是result中的数据更少，这种情况如何判断？甚至前者的描述更准确，但目标SQL要的却是后者的数据

问题955，time指代不明（finish time or fastlaptime）。此外，sql过于复杂，有更简单的方式得出结果但只是在小数点后4位以上的数有不同。

问题962，问的是what percentage of driver，sql取得却是满足条件的result的percentage

问题1006，目标sql不知所云（time和milisecond两列完全一致），查询结果也不对（2：00：00，实际67411ms）。

问题1011，和1006一样的问题。

问题1014，目标sql过于复杂，SELECT results.fastestLapTime FROM circuits JOIN races ON circuits.circuitId = races.circuitId JOIN results ON races.raceId = results.raceId WHERE circuits.country = 'Italy' AND results.fastestLapTime IS NOT NULL ORDER BY results.fastestLapTime ASC LIMIT 1即可；并且问题可能有歧义，个人会认为是要取得每个赛道的赛道记录。

问题1016，用 Austrian Grand Prix Circuit 迷惑llm，circuit应该小写，不然人类也会认为Austrian Grand Prix Circuit是一个赛道（这条赛道事实上是red bull ring），迷惑不能违背人类语言逻辑，小写一样构成迷惑。

F1的系列问题有个通病是使用过于复杂的公式计算时间（时分秒格式转化为秒后做比较），然而表中直接给出了等价数据（miliseconds）。

问题1037，目标SQL计算球员数量时没加DISTINCT，导致同一球员被重复计数。

问题1041，又是问名字没distinct

问题1058，无法设置标准答案（问题问的是哪个更好），可以考虑改成给出名字

问题1071，又是考察distinct name，此类题目太多

问题1084，只考察DISTINCT name太简单，考虑进moderate

问题1094，问题含糊不清，要求的其实是sum，但易被认为是求avg（语境导致）

问题1114，1115，1139全部考察对非标准形式的date的获取能力（1139未入题库）

问题1168，问题中The oldest SJS patient's medical laboratory work，sql中 T2.Diagnosis = 'SJS'没有考虑到除了SJS还有其他病的可能性（比如diagnosis为RA，SJS就抓取不到），缺乏对表中数据的描述

问题1171，SQL的执行结果完全是错的，，根本没有满足小于18岁的要求

问题1183，过于简单： SELECT `Diagnosis` FROM Patient WHERE `Diagnosis` = 'SLE' AND `First Date` = '1997-01-27'

问题1185 网络上的对decrease rate的一般定义为Percent Decrease = [(Old Value - New Value) / Old Value] × 100.，但此题求decrease rate for his/her total cholesterol from November to December in 1981?却是将12月的数据作为除数（即本例中的New Value），不合逻辑

问题1189，问题中的have a level of anti-Cardiolip in antibody (IgM) 20% higher than average?并未描述需要的是满足Thrombosis = 2 AND `ANA Pattern` = 'S'的群体的average，因此逻辑不对。

问题1190，uric acid level below the normal range?目标SQL未分男女讨论，但实际上男女的低UA标准不一致，表描述文件中有体现（男小于8.0，女小于6.5）；目标SQL中简单的取了所有UA小于等于6.5项(CASE WHEN UA <= 6.5 THEN 1 ELSE 0 END）

问题1191，较简单，考虑放进moderate。

问题1192，目的之一是考察DISTINCT name，但结果中没有体现出来。

问题1202，ID缺少DISTINCT但未影响结果。

问题1231，1232，1236，1307全部考察对DISTINCT ID的认知

问题1239，表描述中明确写了29<HCT<52为正常值，但目标SQL为 exceeded the normal range选择的条件仅为HCT>52，而没有考虑HCT<29也是超出范围。

问题1241，表意不明，How is it compare to...无法直接回答（答案可以是low，high，big，small等），但目标sql给的是两者的差值，并且没有给出问题中state the number of patients要求的值

问题1247， 目标sql挑出来的数据很多是不合要求的（SEX应为M），因为OR和AND语句之间少了括号

问题1257，，又是挑出不合要求的数据（生日至今小于18年），因为目标SQL的年龄判断用的是<‘18’ (应为<18)

问题1270，答案设计不好（0），容易误打误撞

问题1292过于简单，考虑进moderate

问题1295，examination的ANA Pattern 列缺少表描述，无法得知问题中的peripheral对应怎样的数据。且求数量缺了DISTINCT

问题1302，问数量缺DISTINCT

问题1339，逻辑不通，SQL的回答只看字面是某人在9月和10月花的所有钱的总值；无法从已给出的信息中判断出需要的是expense表中的对应项cost的均值（问题问的是某人在event上花的钱，并但是未给出所有expense都来自event这一信息）

问题1437，考察DISTINCT 但过于简单，考虑放进moderate并把sql改成SELECT DISTINCT E.link\_to\_member, A.link\_to\_event FROM Expense E JOIN Attendance A ON E.link\_to\_member = A.link\_to\_member WHERE E.approved = 'true' AND E.expense\_date BETWEEN '2019-01-10' AND '2019-11-19'

问题1448，过于简单，考虑放进moderate

问题1451，在没有明确要求回答ID的情况下目标SQL求了ID而不是人名

问题1457，问题中明确指出要全名，那就应该只回答一个全名，而不能把first和last name作为两个分开的项目给出(有待商榷)

问题1464，情况同1437，且也涉及full name

问题1481, 有错误，目标sql计算Consumption的方式是T2.Consumption = ( SELECT MIN(Consumption) FROM yearmonth ) AND T2.Date BETWEEN 201301 AND 201312，这事实上最终只能选出一个条目；然而根据问题，符合 least amount of consumption paid in CZK for 2013的条目显然不能这么算（需要将每个CustomerID的所有consumption累加）。事实上最终ID为43110的这位客户在这个时间段内的consumption为170116.39而非运行目标sql得出的582092.86。

问题1482，问题中的paid in EUR在sql中没有体现，并且也没有给出增加率最大的segment而是直接给出了每个segment的消费增加率。

问题1526，目标SQL联表gasstation完全不必要

问题25，只考虑了District name = Riverside的情况而没有考虑Countyname = Riverside 的情况（问题也没明确告知要的是哪个riverside）

问题27，问题里问的or，目标SQL里给的and

问题34，明明给了percentage的列却还要硬算（结果有细微差别），在题目条件不给出的情况下完全不合理

问题89，93，题出的不好，答案是none容易蒙对

问题95，目标SQL给出的实际是average salary最高的人里youngest的，问题问的是 who are youngest and have highest average salary，当然也可以被理解为youngest的人里salary 最高的，有歧义。

问题100，答案为0，容易蒙对

问题217，答案sql画蛇添足，问题已经是在问确定的分子中的两个确定原子之间的连接是如何的，并不是问这两种原子之间的化学键的可能性.

问题234，题目问的是bond involves atom 12 does molecule TR009 have，目标sql求的是bond involves atom 1 and atom 2 at the same time does molecule TR009 have, 无逻辑对应

问题408，答非所问（问的是数量回答的是T2.text）

问题881，有歧义：每个选手的个人生涯完赛率，还是所有人当场比赛的完赛率

问题1219，WHERE。。。OR。。。AND。。。与问题逻辑不对应。因改为

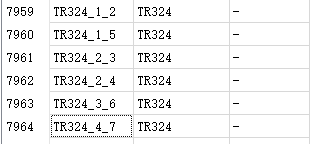
WHERE（。。。OR。。。）AND。。。此类问题频发

问题1275，表描述信息与实际不符（实际值为negative或0，表描述为N=-or+-）

debit\_card\_specializing系列问题过多考察同样的非标准日期格式的理解

类似CZE = 捷克共和国 的缩写识别，是否为默认知识？

问题259，联表时损失数据



目标sql漏选。它确保bond中的分子在molecule数据库中存在，但这并非必要，反而错失正确结果

Bird经常会出现进行不必要联表而缺少正确结果的情况，如问题259

问题357，promotion 类型为NULL理论上也应该进入答案，不能排除，这里就产生歧义了

问题556，有歧义，问题中的views指代的是user的views还是viewcount呢（gpt4可以理解）

问题741，题目问的most powers，目标sql的判断条件却是ORDER BY COUNT(T2.hero\_id) DESC LIMIT 1，即hero\_id最多的，这导致Captain Marvel 因为有两个id而被重复计算了两倍的power\_id，从而得出错误答案

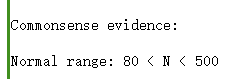




Bird中存在不少问题要求不明确的情况（如是否类问题，答案往往不唯一，无法验证）

问题1199，问题问的protein level，回答却给的ALB level 而非Total Protein level

问题1204，问题问how long，sql的结果却为负数，不合逻辑

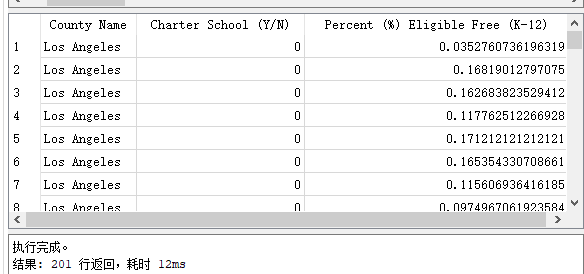
问题1253，表描述中写明正常范围为，但目标sql中的条件描述却使用了between 80 AND 500（包含上下界），不相符（此问题多次出现）

问题1446，目标sql条件是错的，Po\_Box它写成了Po\_box

问题1477，不合理，从表描述中完全无法确定目标是yearmonth表中的consumption列

问题0，question中要DOC type，目标sql中却给T2.DOC

问题2，事实上有201个满足条件的条目，但目标sql只能给出一条



Devnew中：

问题9，目标sql考虑与=键属于同一分子的原子而非=bond中的两个原子，结果就是：



TR005\_11\_12中根本没有h元素，但却联立了，这就是误解了问题。

问题12，问的是所有不全是f的占比

问题18，不存在所谓single bond molecule的公认概念