目前改过：california的magnet，superhero的color，F1的result的fastestlapspeed、time

问题0，考察推理复杂询问的能力

问题1，考察结果为复数，但问题中全部为单数格式时的推理能力（基于列名和词义的推理）

问题2，表描述中对magnet这列描述有问题，修正。（难度源于微妙词义理解，并故意引入serving作为干扰，其实是要offer的数据）

问题3，考察联表能力（复杂询问的能力）

问题4，考察公式推理（复杂询问的能力）

问题5，考察复杂公式推理（复杂询问的能力）

问题6，考察推理能力（复杂询问的能力）

问题7，考察联表能力

问题8，考察对表中值的分析能力

问题9，考察存在多个等价列时的能力。

问题10，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11 ，存在歧义可能，但仍在可识别范围内，需要一定相应知识。（atom不可能又是s又是i）

问题12，考察推理能力

问题13，考察llm对element这个词（可引申为某些常规语言概念）的理解。

问题14，考察llm是否能理解某些术语间的联系，比如element，atom和bond

问题15，考验llm的容错率，问题的一半是无法回答的然而另一半可以。

问题16，考验llm对格式错乱的鲁棒性

问题17，考验llm对格式类似的数据的辨别能力（依赖表描述）

问题18，考察llm对整体数据的理解能力（目标是atom数量但问题中未直接给出该要求）

问题19，联表能力

问题20，考察llm对疑似歧义的句子的理解能力（人类判断时认真分辨则不存在歧义）

问题21，考察llm对要求的理解能力（which）

问题22，考察llm对相似列的分辨能力（题目条件疑似属于多个列）

问题23，考察公式推理能力（需要增加对表内容的描述）表内容理解

问题24，考察对词‘status’的理解能力（是否需要要加distinct有待考量）

问题25，考察是否能根据语境判断出需要distinct（）

问题26，考察which player 词义理解（要的是displayname而非id）

问题27，考察llm推理能力（表中数据认知，哪一列有student）

问题28，考察对相似列的分辨能力（user.displayname 和OwnerdisplayName）（也是考察对表中数据的认知）（特殊问题Which one）

问题29，考察llm对表中数据的认知

问题30，公式推理能力

问题31，考察对表中值的认知（DC comics）（需要增加对表内容的描述）

问题32，公式推理能力

问题33，考察对表中值的认知（Dura or dura）

问题34，考察句意理解

问题35，考察句意理解（color指什么color）以及表中值（female，Female和blue，Blue）

问题36，公式推理

问题37，公式推理，联表加表中值理解（female）

问题38，考察句意理解以及表中值（blue）

问题39，考察对表中值的了解（female）

问题40，考察公式推理能力

问题41，考察公式推理能力（next race 以及No.853之间的联系，句意理解）

问题42，考察在不同表中有相同列名的情况下llm的分辨能力（driverstandings和racing中都有position这列，但driverstanding中的数据更多）

问题43，考察公式推理能力与词义理解（completion percentage）

问题44，考察公式推理以及seconds和millisecond的关系理解

问题45，考察联表能力（考察如何规避多选）

问题46，考察联表能力

问题47，考察联表能力

问题48，考察联表能力

问题49，考察联表能力以及考察对表中数据认知

问题50，考察对表中数据的认知和理解（seanson 2010 = season 2010/2011， league name）

问题51，考察联表能力以及基础公式推理能力

问题52，考察联表能力以及句意理解（问name的时候要DISTINCT）

问题53，考察联表以及基础公式能力（高复杂度查询）

问题54，考察公式推理能力

问题55，考察公式推理以及对错误字符的鲁棒性基于对表内容的认知（\n）

问题56，考察对数据的认知能力（如何获取正确的时间范围·）

问题57，考察对数据的认知能力（如何获取正确的时间范围·）以及基础公式推理

问题58，考察公式推理能力（可能降为moderate）

问题59，考察公式推理能力（如何在一个需要COUNT的问题中正确应用除号而不会导致精度损失）

问题60，考察语义理解，诱导llm回答多余

问题61，考察词义理解（下降率）以及公式能力，考察求百分比数能力

问题62，考察读取数据描述能力以及处理复杂查询的能力

问题63，考察联表能力

问题64，考察联表能力以及时间读取能力

问题65，考察按要求回答的能力以及选择正确查询对象的能力（选DISTINCT patient ID而不能选case）

问题66，语境理解，DISTINCT ID

问题67，语境理解DISTINCT ID加考察时间计算

问题68，语境理解，50岁以下指的不是现在而是1984年实验进行时的年龄，以及时间读取

问题69，考察语义理解能力

问题70，考察语义理解（yet = time now，name = DISTINCT ID）

问题71，考察语义理解（此处times是指倍数而非次数）

问题72，考察对表中数据的了解程度（Social和Vice President都要首字母大写）

问题73，罕见语法检查（COUNT和SUM不可以用在WHERE里，要用HAVING COUNT），以及联表

问题74，考察求full name 能力（如何结合first和last name）以及DISTINCT name

（关于full name, 有待商榷，是）

问题75，考察对表中数据的认知能力（Water, Veggie tray, supplies需要作为一个整体而不是分开）

问题76，考察如何在不多SELECT的情况下算出答案

问题77，考察非标准格式时间读取能力，联表能力（不过多联表），公式能力，理解能力

问题78，语义理解（Consumption decrease rate指的其实是consumption之和的下降率），联表，公式

Moderate

问题79，考察语义理解（direct charter founded school）

问题80，考察词义理解能力（CDSCode = cds）

问题81，考察规避多选目标的能力

问题82，考察对表中信息认知能力（给出了注释）

问题83，考察CAST AS REAL以及百分数处理（先要乘100）（percentage要求取小数）

问题84，考察查询目标为NULL时的句意理解

问题85，考察特殊问题回答能力（YES or NO）以及对连续问题（多输出）的处理能力。

问题86，考察对most common的词义理解

问题87，考察基础的联表能力

问题88，考察基础联表以及表内容认知能力（NULL 也意味着没有）

问题89，考察表内容认知（未给出任何信息）

问题90，考察推理及联表

问题91，考察基本联表和抗符号干扰能力以及last edited 词义理解

问题92，考察对表中内容的认知（未给出）

问题93，考察联表能力

问题94，考察联表能力

问题95，多信息联表

问题96，公式推理

问题97，信息了解以及词义理解（out of europe and asia）（特殊考点）

问题98，公式推理及联表

问题99，考察条件复杂的联表

问题100，考察信息认知以及复杂联表，词义理解

问题101，考察对draw的词义理解

问题102，考察基本公式推理

问题103，公式推理加规避多选

问题104，语意理解（女人中的百分比）

问题105，考察对age的词义理解（考虑进challenging）

发现对涉及年份年龄的词容易不理解

问题106，考察联表和时间（年份）确定

问题107，基本公式推理加时间确定

问题108，考察表中内容认知（food 应为 Food）

问题109，考察联表能力以及对difference的词义理解（改challenging）

问题110，考察表内容认知（时间格式）一阶对monthly的词义理解（改challenging）

问题111，考察联表以及时间查询能力

问题112，考察表内容认知（fresno这个信息对应哪个列）

问题113，考察表内容认知（Los Angeles这个信息对应哪个列）

问题114，考察如何规避多选

问题115，考察简单联表

问题116，考察对district的词义理解（要name而非id）

问题117，考察基本公式推理以及联表

问题118，考察简单联表

问题119，简单公式推理（compound词义理解，指的是bond而非molecule？？？）

问题120，简单联表

问题121，考察简单公式推理

问题122，考察简单联表（表内容认知Angel）

问题123，考察目标隐藏时的选择能力（目标不在name中而是在另一个表中）

问题124，考察公式以及联表（WHERE条件判断）

问题125，考察联表条件判断

问题126，考察对表内容的认知

问题127，考察最简单联表

问题128，考察简单联表（联立条件）

问题129，考察简单公式和WHERE条件

问题130，考察NULL数据抗性

问题131，考察对表中数据认知（非标准时间格式，as词义理解）

问题132，考察对finish the game的词义理解（time不为NULL）

问题133，考察年龄计算（句意理解）

问题134，考察AVG使用以及WHERE条件判断

问题135，考察时间计算

问题136，考察内容认知

问题137，考察DISTINCT能力

问题138，考察AVG使用以及WHERE条件判断

问题139，SUM和COUNT区分

问题140，expense词义理解（expense = expense description）

问题141，考察简单公式计算

问题142，考察简单公式计算

问题143，考察规避多选能力

问题144，简单计算以及表中信息认知

问题145，考察DISTINCT相关推理能力