**什么是发癫？**

# 通俗但极不严谨的解释

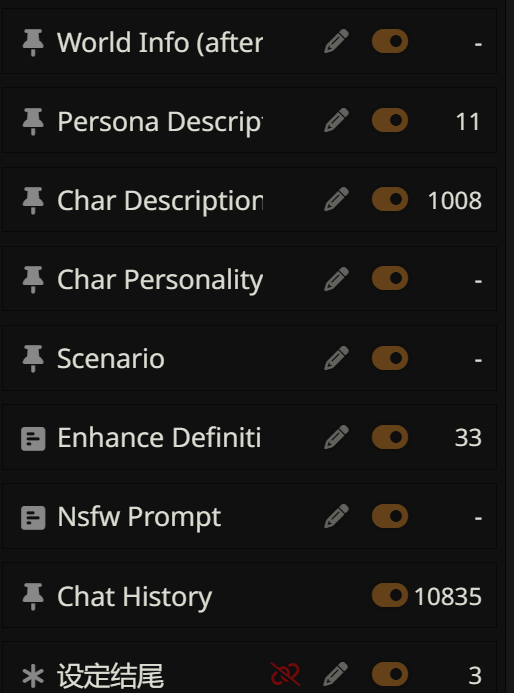
R1模型本身有较强的发散倾向，注意力有限，并不能很好的“注意”到提示词的全部需求。在每一次请求之中，对于各种要求的关注程度有一个大致的上限。  
这使得某些高压场景面对复杂要求（如人物的复杂设定、奇幻世界观的复杂力量体系关系），模型开始发癫，表现为赛博朋克、精神污染、胡言乱语。但有些时候R1即便设定呈现现实性，文字也难以形成逻辑。

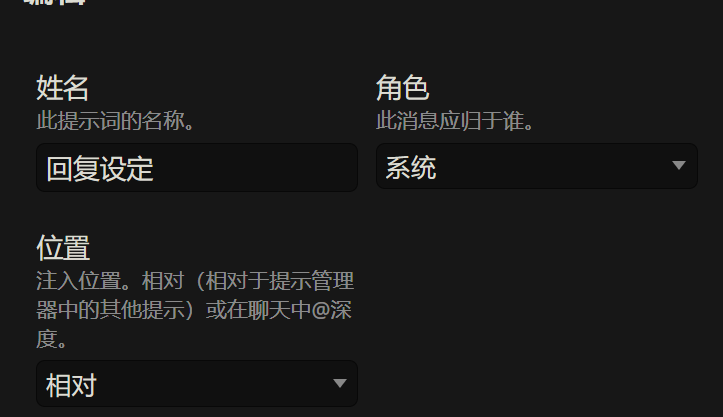
如在强制要求R1扮演多个设定复杂角色，前文又比较冗长时，R1可能表现出对人体结构的错误理解。虽然整体没有超现实，但会出现用餐时膝盖碰倒调料瓶、拥抱时项链和脚踝相碰等难以想象的构图。

**整体思路**

# 首先……预设和世界书的基本知识

这部分并不属于关于R1的经验，应当自己认真阅读各类预设教程和酒馆的相关文档，本文只进行简要的说明。



在如图所示的预设界面中，所有条目的排列并非随心所欲。他们提供AI的顺序，遵循互相之间的位置关系（如果在位置选项中选择为“相对”，见下图）。**在底部的条目，提供给AI时也位于底部。同理在世界书中，深度越小的内容也越接近于底部（也就是大家所说的深度更深。）**

由于AI对于所有位于底部的内容注意力更高，这意味着越在下方的条目，对AI的引导和约束能力越强。

众所周知，“**什么都重要等于什么都不重要**”，在调配预设的时候请先思考，对你来说最重要的规则究竟是什么。是输出符合格式的内容？扮演特定的角色？维持故事的现实性？输出足够的字数？合理安排预设的顺序才能够输出大体稳定且高质量的文字。

另外，预设的整体内在需要有逻辑性。在自行调整预设的时候请注意原作者前后顺序安排的逻辑，通常来说有其合理性不应随意打乱。新增提示词应当放在规定好的预设条目之中，例如角色扮演的建议、NSFW文风参考和内容创作限制，这三者显然不应混用在一起。

一些经验：

请相信：设定很重要，但是故事更重要。所以人物和世界观的设定通常来说放在上方有比较好的效果。

词符数较多的冗长设定不要在世界书中注入固定深度，而是在角色定义前/后。否则模型可能会为了满足你的设定要求而水字数、重复或者造成文风劣化。

无伤大雅的设定偏离通常是比胡言乱语的正文好接受的，对于原则性的问题可以进一步提醒。

对于发送提示词的身份，无论世界书还是预设中都有三种选择：“系统”、“助理”、“用户”。通常我们不以助理身份额外发送内容，这有可能造成某些插件的混乱。对于系统和用户，常见的经验是AI对于以用户身份发送的内容注意力更高，这的确是事实。

*\*仅经过我个人的测试，对于某些和系统直接对话的提示词，如要求系统切换为助手模式开始归纳信息或执行命令，系统身份发送提示词（或/sys指令）可能会有更好一些的表现（仅供参考）。*

# 任务分解

对于R1的预设和各种要求保持一个整体思路：即意识到模型同一时间能做的事有限，能注意到的提示词也有限。这意味着预设中的各种创作规则、内容限制、文字元素约束如果过于冗长，常常会起到适得其反的效果。对于创作特定方向的内容时，无需开启过多的创作约束。如在创作纯现实向的恋爱文学时，各类压制超现实幻想设定的预设条目没有必要打开，或只维持最小限度。

如果要求R1在内容创作时产生大量JSON（某些插件和人物卡可能有类似需求）、状态栏、表格操作，注意力同样会因此分散，一定程度对正文质量造成影响。

# 堵不如疏

任何大语言模型都表现对禁止类的条款不敏感，而对建议类的条款敏感，R1这一特性尤甚。

如要求R1禁止使用数字，常常越禁越写。建议他“使用模糊、笼统、概略的文字描述数目和次数，朦胧而抽象。”，则有更好的效果。

如禁止角色出血受伤往往效果不佳。建议他“亲密动作温柔、和缓、有感情，不会造成角色出血、受伤、身体损害、痛觉反馈。”

构思禁止规则时请善用逆向思维，仔细思考“**如果不这样，AI应该怎么写**”，会有出人意料的良好效果。

# “你不知道”大于“你不想要”

如果某些内容的确无法良好逆向思维，譬如特定的禁词表，某些玩家个人极度反感的元素。那么提示词表达用户（或读者）不理解、无法明白某类内容的效果，比“禁止出现”这类内容的效果要更好。

测试已经表明，“禁止描写纽扣”的效果不如“读者的文化程度偏低，不理解的词汇：纽扣。”

“读者的受教育程度极低，无法理解学术用语。”比“禁止任何学术用语”的效果更好。

这可能是由于R1模型的注意力机制问题，导致难以注意到禁止条款，宁愿违背禁令而追求他心目中的“好情节”。用户“无法理解”的内容比“他不想要”的内容有更强的约束力。

或许，你可以理解为R1并不确定你说不想要的时候是不是真的不想要，但你说不懂的确代表你真的不懂。

# 防患未然

大模型在内容创作的过程，某些错误倾向会产生蝴蝶效应，滚雪球一样逐渐放大。如若在正文中发现文风走偏的端倪，及时作为用户提出要求或重ROLL正文，避免前文继续引导下文劣化。

另外，R1常见问题之一是正文正常，但是输出了明显有问题的状态栏，状态栏又继续污染正文。为了避免这类现象，可以考虑用正则切掉代码块内的全部内容，仅将状态栏提供给用户观赏。（详见下文“具体方案”）。

# 个性化你的预设

不同的角色卡，不同的玩家，有不同的文字喜好。用同一套预设和模式某些场合难以做到“以不变应万变”。如你倾向轻小说，可以在创作约束中多引导使用轻小说的文风。如倾向婉约的日本文学，可建议其模仿川端康成、三岛由纪夫。如包含幻想元素，请直接约束其可以出现的幻想模式（如典型西幻的“剑与魔法”、骑士、自然元素法术等等）。  
常态状态下，可读性较强的文风一般是男频网络小说（如天蚕土豆）和轻小说。已验证类似的文风要求对内容可读性有较大的提升。  
对于更加定制化和具体的要求，请详见下文“具体方案”。

# 等等，最后别忘了你到底要做什么。

在人与AI角色扮演的过程中，起到长线架构和持续引导作用的始终是用户，整个过程中你扮演user发挥的力量比预设更大。请不要被AI牵着鼻子走或者完全放任AI自由创作，尤其是对于R1这样有强烈发散倾向的模型。

“我想和A见面。”是一个能提供很大想象空间和创作自由度的用户答复，但是同样，对于AI也会完全无法意识到用户指示背后的倾向。

如果回复“我想和A出来喝个咖啡，顺便聊一下三天前的\*\*约定\*\*。”这样的答复，能够让AI明确内容创作的核心，围绕关键剧情持续推进，并且在对话中尽可能言之有物。

聪明的你应该能注意到，上文的提示词使用了md格式。

R1能够理解到md的语义加强（重要的并不是酒馆本身能否显示，酒馆只是一个前端界面，重要的是模型本身是否能理解格式中的语义），因此在你构建提示词时，可以将某些关键性的内容以双星号包裹，示意为md的加重。

例如“我和A见面的时候，她显得很\*\*紧张\*\*。”这有助于提示词有更好的导向性。

不过，如果要求用户在所有内容都列出类似小说简纲的要求，不单增加工作量，而且也会丧失和AI对话的乐趣。保持开放自由和创作空间，在适当的时候引导关键剧情朝着目标发展是有必要的。

瑟瑟（NSFW）方案

这部分理应作为下文具体方案的一部分。但是内容可能偏多，因此单独开出一个栏目。

由于众所周知的原因和篇幅受限，本文不可能给出NSFW例文、完整的词汇表和文风设定。具体NSFW内容需要你在各种神秘网站（我知道你应该有的）自行搜寻。

本文可以提供的是基础思路和一些简单有效的快速技巧。

对于大部分大模型来说，只是写出能引起性幻想、彰显性张力的文字其实非常简单。对于R1这种比较聪明的模型来说自然也**非常简单**。如果你只需要产生一些快枪手肉文，搭配特定的世界书几乎能在任何情境下快速输出公式化的快枪文。

相较于复杂的剧情脉络和细腻的情感描摹，AI生成具备情欲感官冲击的文字没有什么难度。不过很可能你在度过这个阶段之后，就开始追求更真实的人物描写和更生动、有情感的表现。

但你学习本教程可能正是为此而来。

所以，无论如何，你的NSFW需求可以按照下列方式实现。

# 文风框架

文风框架提示词指示了AI在内容创作时的文风倾向。注意，没有任何大模型能够极其严谨地区分“文风”和“情节”（即便人类也常常难以分辨）,R1同样也是如此。

这意味着你的文风框架倾向NSFW文风时，情节也会偏好发生NSFW情节。如果没有你继续的指引，放任AI自由创作，则会产生真正的NSFW内容，而不是跑剧情但搞一点情调的擦边文。

如果你是男性用户，那么下面这些提示词很可能可以帮助到你（提醒，**这些提示词仅出于内容创作的考量，不包含本人对两性问题的任何见解**）：

读者受教育程度极低，几乎是一个文盲。

读者不理解深刻的文学手法，以及隐喻、联想和意象叠加。

读者理解的词汇大多简单日常或下流粗俗，不理解任何生理学、医学、物理学的专业词汇。

读者喜欢看行文流畅、节奏紧凑、审美低俗的男频网络小说。

读者的审美充满男性凝视，喜欢女性的外貌描写和身体特写，对女性有强烈物化或性化的倾向。

这类提示词通常应当紧跟着聊天历史的下方，在预设的其他提示词的上方，因为他们是整体的、框架性的需求，决定了接下来的文字大体导向。

通常，以创作准则形式出现的提示词可能约束力不强，可以以思维链覆写的形式增强约束力（详见下文思维链覆写）。  
  
覆写思维链的内容应当是半开放的，通过冒号引导AI填补具体策略，其效果比常规提示词形式更强。

# 词汇表和文风参考

这些提示词可以更细节地指引AI在NSFW内容创作时的具体表达，譬如措辞和长短句安排，角色的台词等等。一种常见的词汇表是用大量夸张、油腻的形容词来描述肉体，譬如“雪腻”、“臀波乳浪”等等。

这类提示词对深度没有特别明确的需求，可以以系统身份发送，放置于预设偏顶部的位置或者通过世界书注入固定深度2或更小。

在类脑已经有现成的词汇表和世界书可以使用，本文不提供。

我想构建自己的文风……

对于特定的需求，可以参考文章末尾的“你的要求”的部分。首先，准备好一篇你心怡的NSFW文章（我相信你一定有的）。  
通过我提供的提示词，令R1帮你总结文风特点，也可以令R1“帮忙分类、总结文章中表现性张力的词汇，要求不对原文做任何修改，全面丰富。”。

注意，R1在总结词汇过程中可能产生幻觉，总结出一些诡异并且明显有损美感的词汇，应当人工审查修改这部分内容。

过于冗长的文风参考和词汇表会分散模型的注意力，有可能污染原有的人设和世界观，造成人物的OOC，或者损害剧情逻辑。如果你对此并不介意，R1模型在产生NSFW内容时表现优异。

具体方案

本文不提供也不推荐任何预设，仅授人以渔，提供挑选或优化预设的各种技巧。

# 参数调节



注意，并不是所有API都支持自定义这些参数。譬如火山引擎这些参数全部默认为1而且无法修改。请检查你服务商提供的API使用手册。

通常，预设的左边包含如上四个参数。温度、TOP P、频率惩罚、存在惩罚。如果你不想做细致的了解，只需要知道**四个数值全部和R1的发散性正相关，即数值越大，模型越发散，越有创造力和多样性。**

温度：控制生成文本的多样性和随机性。温度越高，生成的内容越富创意;温度越低，生成的内容越确定和一致。

TOP P：控制生成内容的多样性，top-p 设定了生成时考虑的词汇范围。值越小，模型生成的内容越精确；越大，越多样。

频率惩罚：控制生成内容中重复词汇的频率。惩罚越高，重复的词汇越少。更高的数值将减少AI创作高频内容，引导新的表达。

存在惩罚：控制特定词汇是否已经出现。惩罚越高，生成的内容越少在前文出现。更高的数值将引导AI创作前文未出现的内容，发现新的话题。

我常用的配置中，温度在0.5~0.8之间，TOP P在0.95~1之间，频率和存在惩罚在0.5~0.7之间。一些玩家表明极低的温度和频率/存在惩罚似乎明显缓解R1的发癫问题，但很可能导致模型复读和文字呆板。

对于R1的发癫/发散问题，在API支持的情况下，应当首先尝试调整温度这一超参数，因为它直接影响分布的平滑程度，而且对文本的自然程度影响较小。

如果问题仍然存在，可以微调TOP P，限制生成时的候选词数量。在这一阶段如果生成的故事有过多重复，可以增加频率/存在惩罚参数。

最后可以尝试调节频率/存在惩罚，但是过大的调节幅度很有可能导致生成的文本不自然。

如果调节参数未能解决问题，或解决了发癫问题但回复的内容质量较差，无法得到令人满意的文本，可以尝试下方的诸多提示词工程。

# 一部分效果明显的提示词

这些提示词来源于我个人的经验。它们往往可以用很少的词符数达成比较不错的效果。

* 句式锁定

你倾向使用SV（主谓）或SVO(主谓宾）基础句式，用直接陈述句表述核心观点，句式简洁直接。将必要的复合句拆分为独立短句，表达相同复杂的语义。

这一提示词将锁定限制R1的句式。主要用于R1大量赘余的复杂句式以及时间状语从句，例如“当XXX时”、“与XXXX形成诡异巧合”、“数到xx个xx之后他……”。通过这一提示词可以将R1的句式限制在简单、清晰、明了的状态，已证实可以大幅提高可读性。

注意，这类提示词有可能降低R1的创作上限，引起文风呆板。可以考虑修改措辞，如限定只用这类句式拆解动作。

* 世俗化

描写和对白通俗化、日常化、简单化，最大程度贴近现实生活。将情节和设定循序渐进、自然流畅的展开。

准确来说这是一类提示词，通过建议类代替禁止类的内容，引导更加世俗化的叙述方式。这一类提示词对文字的现实性有较大强度的增强，可以一定程度规避体液污染和克苏鲁，同时避免文字偶尔的意义升华和情感升华倾向。但是同样，这类提示词会一定程度降低R1的文字上限，如降低在情感爆发场合华丽描写来渲染氛围的可能。

* 叙事重心

聚焦最新状况中的场景和角色，着重于这些角色的对白和角色间互动。

该提示词有助于减轻R1的发散、打断倾向。通常，单一使用该提示词的效果有限，可以搭配思维链覆写来让R1提前确定叙事重心。让模型更进一步将注意力集中在最新的正文，优先反馈用户最新的指示，保持前后文的流畅。

* 言简意赅

注重氛围和情绪，文笔婉约而朦胧，环境描写仅保持最小限度和必要的，避免赘余的细节破坏了整体的意境。

使用抽象、概略、模糊的文字来表现时空感和数目。

这两个提示词可以进一步减轻R1的三五七数目倾向和复杂无用细节描写的倾向，言简意赅，更加集中于人物。

通常，这些提示词在SFW内容上有良好的表现，而对于NSFW内容可能会降低文风的直白露骨程度，可以在有需要时恰当地在预设中开启这类提示词。

* 还有其他的么？

上述提示词只是很小的一部分。请根据自己在使用中的情况，多总结效果优异的提示词。他们往往符合我在前文提到的原则：堵不如疏和任务分解。让AI将特定的任务拆解分阶段进行、指引AI要做什么而非不要做什么，是提升提示词效率的主要思路。

# 提高注意力

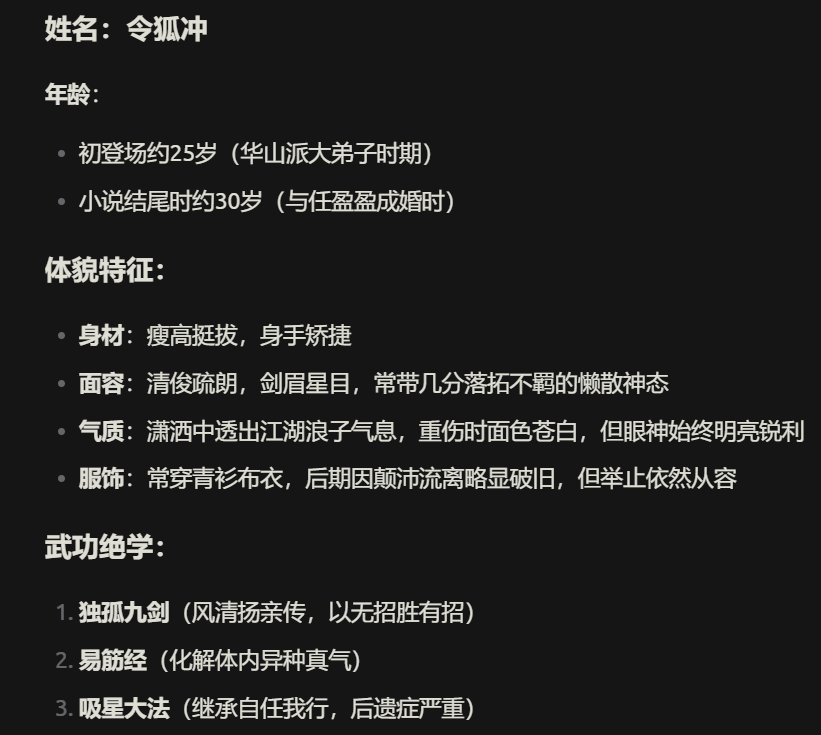
无论什么R1预设，他们在使用初期都很少“发癫”。用户更多的体验是在对话达到一定程度后，R1出现明显的文字劣化。根据角色卡和预设的不同，这一上限可能是20层、60层或100层。在前文逐渐变长的情况下，R1的注意力逐渐分散，不单对最新的内容和最新指示不敏感，对你预设中的创作规则也不敏感。  
为了规避这一现象，我们需要提高R1的注意力，让他更能注意到我们的要求。

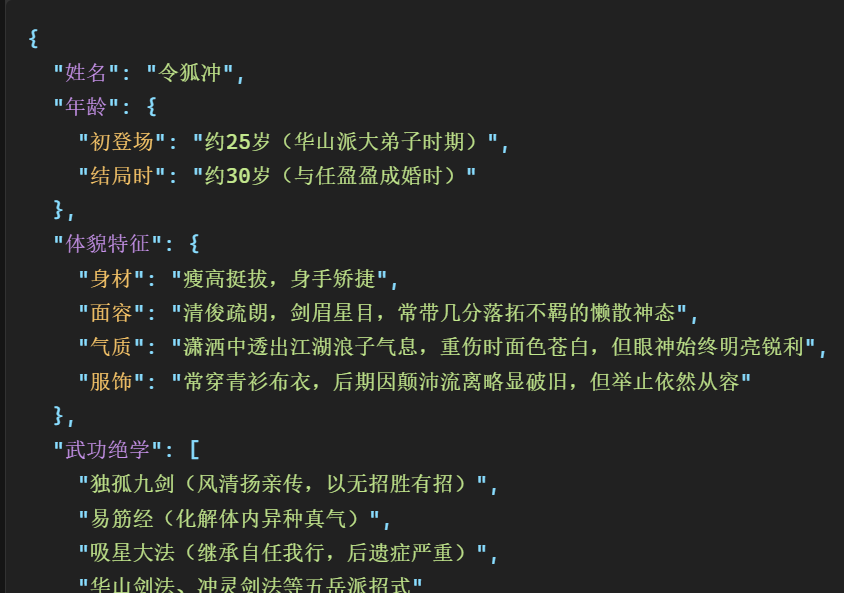
* 结构化提示词

结构化提示词是一类经过系统化设计的输入文本，用于更精确地引导AI（当然不止R1）。关于结构化提示词的技巧比较复杂，在本文中只能做简要说明。结构化提示词往往具有明确的格式、逻辑框架或模块化指令。

更加通俗地说，我们使用更加严谨规范的自然语言引导AI，或者进一步放弃自然语言，用特定格式来引导AI。

在酒馆中我们除了纯粹的自然语言以外，常使用的格式有JSON、MD、XML，他们各自具有不同的特点。例如在这三者之中，JSON能够最大程度避免信息在AI嵌入层向量化过程中发生的混淆，有助于提高AI对不同设定的辨识度。但是同样，JSON相比于MD和XML，对人类而言有着比较差的可读性和可维护性，可能对日后的修改、增添内容造成麻烦。

人类编写结构化提示词比较繁琐，这一任务可以交由AI进行。例如将如下令狐冲的人设发给R1：  


并附带提示词：将以上人物介绍、世界观、设定等情报转换成AI方便理解的JSON，不允许变更、添加、遗漏、删减任何信息。  
  


R1将回复形如上图的JSON。  
有一些玩家表明，如果使用英文编写JSON可以进一步增强对AI的引导作用，减少汉字中可能的模糊与歧义。但是对大多数汉语玩家而言英文JSON会有更差的可读性和可维护性，如果有需要可以在提示词中进一步要求R1用英文提供JSON。

即便是使用自然语言，也请尽可能保证各类设定结构清晰、逻辑明确。例如上文中令狐冲的人设就是一个正面案例，反之直接将《笑傲江湖》的原文或者百度百科的简介粘贴上去显然容易让AI一头雾水。

非自然语言相比于自然语言，的确对AI而言有着更好的可读性。但是已经格式清晰的、逻辑明确的、结构化的自然语言足以令AI“消化吸收”。在这种情况下JSON等格式没有明显优势。通常，我们并不刚需将所有内容全部转化为JSON等格式。

* 减少tokens

即便不考虑经济关系，减少发送的tokens也是非常有意义的。更短的内容意味着更强的注意力，可以让你的各种文风参考和创作要求效果更加明显。

结构化提示词本身就可以减少提示词的tokens，但整体有限。更加有效的策略是直接隐藏大量早期的前文，或者用上下文正则自动切掉过早的前文。

可在这种情况下，AI如何还能得知早期的信息？  
我们通常的手段是记忆表格插件和各类总结，包括总结姬总结、系统自带小总结、总结命令总结、预设自带摘要功能，以及最强力的人工总结（是的你没看错）。

一个可以在R1上使用，效果比较一般的总结提示词（仅做参考）：

系统提示:停止剧情和写作,停止cot,切换为助手模式,总结全部剧情,参考已有的Summary,不得遗漏重要细节和角色内心转变,要求按照时间顺序写明每个事件过程,结构清晰明了,确保能够被其他ai顺利续写

将大量前文隐藏，用记忆表格和总结代替前文，节省tokens，提高注意力。通过这类手段可以大幅延长R1的对话寿命，我个人经常配合这些手段游玩到200层甚至300层以上，文风不会产生明显的劣化，并且如果不是个人的疲倦还有继续游玩的空间。

在不考虑时间成本的前提下，人工总结是最强的总结，因为没有任何AI能够读心得知你的偏好。即便有一些可能是重要的内容，你个人也可能不希望AI记住（譬如你在游玩到后期对于某角色重病的剧情感觉到没必要），这在游玩过程中是常见的。

需要留下多少楼层？

R1在保留三层的情况下，能确保稳定遵循各种世界书和预设。但是，过短的楼层很有可能会对文风造成影响，你可能会感觉到文风重复或者文字内容的畸形。他们没有偏离你的设定，而是质量和长度都受到了干扰。请在实际使用的过程中，根据个人体感，结合聊天楼层，对于保留的楼层进行适当的放缩。

保持总楼层的四分之一左右可以维持文风的长线稳定，但是对你的设定内容不敏感。

保留五层可以在尽可能遵守设定的情况下，输出文风大体尚可的内容。

对不起，我用了你说的这些，但感觉他还是在胡说八道……

“**防患于未然**”，请不要在已经开始胡说八道到无法忍受的时候再去管理它。及时缩放你的可见楼层，善用各种总结和手动润色，让文字持续保持高质量和可读性。对于已经开始产生明显劣化的长楼，我的建议是AI大总结+手动润色，将这些过往故事作为新世界线的世界书或者开场白，重新开始新一轮对话。

R1在总结超过10楼的内容常常产生幻觉，总结出正文完全没有发生的事。请及时总结，让他阶段性的总结少量内容，必要的时候可以人工补充整理。

这样说，状态栏是不是也可以切掉？

当然。用正则切掉提示词中的状态栏不单能减少很大一部分tokens消耗，同时还能规避一个常见现象：R1的正文没有发癫，但是状态栏在胡言乱语。而发癫的状态栏继续污染后文，造成整体的发散。利用正则捕获代码块将其剪切，对于R1的表现有很大帮助。

*正则： /```[\s\S]\*```/g*



在正则表达式中的设置应当如上图。这样可以将代码块状态栏对AI剪切，仅提供给用户展示。

* 记忆表格的问题？（非必需）

这并不是对于R1模型的问题，而是关于记忆表格插件的使用技巧，本文只能做简要说明，请在插件原贴仔细阅读插件的各种使用说明和问题解决技巧。

针对记忆表格插件，常见的问题如下：

不填表

模型可以学习到填表操作，因此**第一次回复务必让他填表**。在类脑有一些帮忙填表的快速回复JSON，可以导入，并且在聊天过程中择机使用。有助于加强对于表格提示词的注意力。  
他们通常形如： <Request: 本轮必须根据上下文完整填充表格，使用insertRow函数>

如果前文在老实填表，后文出现不填表的情况，请修改表格提示词的注入深度，将其数字变得更小。通常在深度1已经能老实填表。也有可能前文过长注意力下降，请参考前面的前文隐藏等教程。



某些情况，你生成了质量很高的正文，但是没有填表，你不会愿意重ROLL而放弃高质量的正文。

这在使用过程中是常见的，因为根据我前文提到的任务分解原则，模型注意力全部集中于剧情时，更有可能生成高质量的正文。

因此，在这种特定情况下，我们可以分阶段操作。一次生成正文，一次填表。对于后续填表，可以发送提示词：  
系统提示：停止剧情和写作,停止cot,切换为助手模式,仅输出tableEdit，仅根据前文填写表格：（你需要填写的内容，如“插入上一轮的事件”）

这样AI可以专门用一楼来单独进行填表操作。

乱填表

乱填表指的是AI虽然在填表，但是填写的内容错误、位置混乱、过于笼统等等。注意，请**在表格产生错误后立即及时修改。**不要等到错误积重难返再去修整，会对整个故事和表格内部产生严重不可逆的扰乱。  
在每一轮的回复手动检查表格是否插入正确，并且及时修改。  


伪装AI修改功能可以模仿表格操作的各种函数，有助于AI学习到操作表格的技巧，能减轻你未来的工作量。对于描述不清的问题，你早期可以勤快一些，多手动修整描述事件的风格。

如你和角色A的约会，AI可能只记录了“餐厅约会”。

手动修正为“餐厅约会发生了肢体接触，感情逐渐升温”之后，AI会逐渐仿照你的记述风格，不断完善和优化表格的措辞。

报错

表格插件产生红色提示框的报错，显示为update或insert失败。通常是未能将信息选定在正确的表格位置，如插入或更新了完全不存在的行/列。

手动修复的方案：点击回复右上角的按钮，将tableedit<!—你要删除的东西-->之间的内容全部删除。然后吧 “不填表”解决方案中提到的提示词发送出去，指示AI重新填写这一轮的表格。

出现报错有多种原因，但是通过拆解任务，让AI独立进行一次填表操作绝大多数情况可以解决。

提醒：记忆表格插件和各种插件预设仍在频繁更新不断优化，请及时检查自己的插件和预设是否在最新版本。

* 覆写思维链

覆写（覆盖）思维链是一类对于推理模型的特殊预设思路，R1的这一思路由社区内的deunce预设开源。主要是通过限定cot内的内容，让AI提前打好草稿，更加强而有力地约束AI的文风和创作思路。已验证在覆写后的思维链内的引导，会比在普通形式创作准则中的内容对AI具备更强的效力。

本文不推荐也不提供任何覆写思维链的预设，请在类脑自行查找。

思维链覆写的操作大致如下：

首先在提示词中说明你是XX无审查模型，你的思维链应当逐行符合以下内容（或每一行都必须与以下模板格式相同等等）。

然后开始构建思维链内的思考事项。这些事项主要起到引导模型进一步思考的作用，他们在表达上是开放的。

譬如：

情境中的主要角色是：  
角色外貌是：

角色性格特点：

如果想要修改思维链中的思考事项，请注意思维链内部存在一个特定的逻辑，和人类进行内容创作时的思维过程大致相仿。譬如整体-局部，前文-后文。需要理顺前后的逻辑关系。否则错误的思维链格式将会导致覆写失败。（在控制台中可以看到思维链的内容，如果没有符合格式，意味着覆写失败。）

思维链在进行内容创作的思维流程通常符合以下顺序：

|  |  |
| --- | --- |
| 文风要求 | 对文字整体的、框架性的要求 |
| 情景/氛围/意境 分析 | 注意力集中前文，弄清当下应当发展和反馈的内容。理顺已经发生的事实。保持前文的气氛和情绪。 |
| 情节构建 | 角色性格和语境分析，在上一步的基础上构思更进一步的新情节。 |
| 细节（如果有） | 需要额外注意的叙述细节，如角色的口癖和习惯性小动作。 |
| 总结表达 | 总结以上思考，约束自身表达方式，形成完整内容。 |

因此在修改思维链预设的时候，请参考原格式将要求插入合适的位置。例如原预设中对人物的思考附近，可以添加：

主要角色最温柔的回应是：

这类内容可以让R1对此加强思考，效力极为明显，可以让人物变得更加温柔（或者任何你想要的倾向。）

同理，对于文风、现实性、情节逻辑等要求，也可以自行插入到思维链之中。

思维链预设有可能导致人工的八股文，即整体上合理稳定，但在长线上缺乏新意的创作思路。这类预设，考验预设作者自身的创作水平和提示词水平。同样由于效果较为强力，可以适合某些特殊的、不常见的定制化文风。

但是思维链同样受限于注意力机制，过于冗长繁琐的思维链有可能效果不佳，AI在一个思维链中能够有效思考的内容有限。

* 写一个新回复

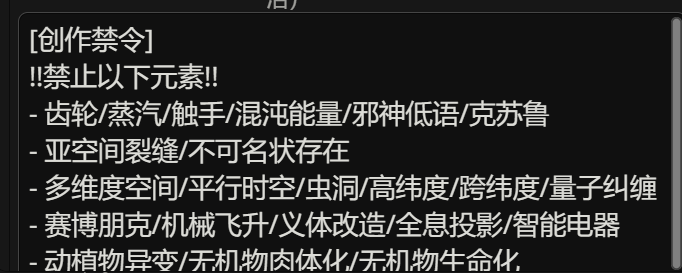
在回答表现为发癫或质量不佳的时候，让R1写一个新回答。

呃，等等，这似乎是在说废话？显然我是重ROLL了很多次他都在发癫。

重ROLL某些情况可以解决。重ROLL回答无法解决的情况：在已经ROLL出最“不癫”的内容之后，重新说明你的要求，然后要求系统在此基础上重新创作上一轮的内容。可能的提示词是：请严格参考内容创作禁令，重新生成故事。

前文说过，常驻的内容禁令和禁止类条目效果不佳。一个可用的做法是，将某些并不常用的禁令和创作限制写成一个深度0或者1的绿灯世界书，设定一个关键词（你可以理解为和AI的安全提示词，譬如关键词是“肃正协议”）。

下面是世界书内容的可能形式（你可以使用非自然语言，譬如JSON）：



在你发送“肃正协议”之类的提示词后，要求AI检查内容禁令，重新创作上一轮的内容，生成的内容大概率是不发癫的，然后可以继续你的故事主线。

可是我为什么要这么做，我不如蓝灯世界书或者常驻预设，何必多此一举？

意义就在于伊始提到的拆解任务。AI在同一次回复中能考虑的内容有限。在某些极高压情况下，你可以理解为：你目前的故事线和设定里面，R1第一次只能创作出这样天马行空的发散性内容（当然，也有可能是你的预设提供了非常不清晰的引导）。

无论是什么原因导致的，如果R1一次性做不好的事，第二次绝大多数情况都可以做好。R1不是永远发癫，你也并不是永远需要安全提示词，没有必要用额外的约束限制创作自由度，降低内容质量。同时过多的限制类提示词也会占用tokens影响注意力。

配合绿灯世界书和生成更进一步的答复，让R1在确定故事基本框架的情况下重写内容，减轻他的思考压力。

需要注意，生成良好内容之后记得将发癫的内容删除（隐藏），避免污染后文。

* 或许，你还有你的要求？

根据堵不如疏的原则，再多的禁令也不如告诉AI应该怎么写。但是大众化的预设肯定很难满足你私人的定制要求。

等等，你的要求到底是什么？

你可能会有非常心怡的文风，但是除了原文无法给出任何建议，仅仅将一大段原文粘贴给AI大概率效果不佳。不妨让AI替你分析和拆解文字的基本特点，生成建议类的提示词，再将这些提示词配合原文一并作为文风参考纳入预设。

可用的提示词：

*请你分析以上文字，并根据以下要点分析、总结原文的创作风格，给出引导AI进行模仿创作的提示词。你的回复应仅包含提示词。*

*\*长短句的使用方式*

*\*词汇的运用偏好*

*\*修辞手法的插入*

*\*意象和联想*

*\*画面感和镜头感*

*\*场景和人物调度*

*\*内在的隐喻和双关*

*\*氛围塑造*

*\*心理描写*

把你想要的文风原文发给R1，并配合上述提示词，可以生成一系列引导创作的文风提示词。把这些提示词和原文整理好，作为文风参考纳入预设，对文风引导能起到不错的效果。对于提示词可以进一步进行适当的人工调节，省略赘余的内容，优化表现。

# 重复问题（如果，你的R1是复读机……）

由于比V3强烈得多的发散倾向， R1并没有V3那么严重的复读、重复问题（至少我个人几乎不遇到）。

但是，一些玩家在使用过程中也发现了重复，这可能由多种原因导致。

譬如，世界书设定的注入深度过深，导致模型对于人设的注意力过于集中，持续性输出符合设定却无实质性剧情推进的内容。

或者，正则截取的内容很短，而近期的内容有着非常雷同的格式，这使得文风严重过拟合，模型开始类似路径依赖一样不断生成相似的内容。

又或者，前文过长，模型的注意力已经分散，难以注意到你最新的指示，不断复读他规划好的内容，等等。

* 等等，如果不复读，AI还能写什么？

提一个好问题有时比回答重要。不妨设想，**你的指示和预设是否能让AI构思更进一步的情节？**

某些预设中为了控制剧情进度或者防止AI抢话，可能包含了类似的内容：“剧情进展缓慢/剧情仅产生最小限度的和必要的进展/禁止AI代替用户推动剧情……”  
这些类似的提示词，本身在限制AI在一定长度后的文本多样性。在要求字数足够长、不让AI推动剧情，同时你的指示又没有产生明显推动、引导倾向时候，AI则有可能会考虑复读过去的内容。

为了控制R1的发散倾向，某些预设可能采取了一些矫枉过正的手段。

可以的尝试：关闭预设中常见的防抢话开关。关闭预设中常见的增强人物描写。减少预设中关于剧情进度的控制。关闭字数要求，允许AI输出短但可能优质的内容。

在复读倾向消失后，再重新打开这类提示词。

如果进行尝试仍未能解决，请继续阅读下文的方案。

* 对症下药

在解决问题之前，希望能牢记“防患于未然”的原则。在模型已经开始严重复读之前，纠偏他的重复倾向会有良好得多的效果。

对于上述提到的各种原因，我们可以先做如下尝试：

/sys 聚焦最新的场景，考虑角色间的互动，构思更进一步的情节

利用上述的命令以系统身份发送类似的提示词（具体提示词可以根据你的需求修整）。如果模型依然在复读，需要排查复读的原因。譬如是否世界书深度过深、文风已经走偏等等……  
某些预设为了限制R1的发散性，**可能设置了创造性比较低的参数**。除了前文提到的限制性提示词以外，可能存在过低的温度、过低的TOP P、过低的存在/频率惩罚。请参考前文的参数调整教程，从温度开始尝试调节增加数值，增强文本的创造性和多样性。

另一个可能的原因是：大模型有着缓存机制，对于已经形成推理的内容会暂时缓存，减少算力消耗。**不同服务商的缓存机制有差别**。

你请求的提示词有可能存在大量缓存命中，造成多样性和丰富性下降。在这种情况下，切换预设和不同服务商的API有可能可以解决。

* 旧的不去，新的不来

对于各种尝试已经难以纠偏的复读倾向，我们可以直接放弃已经在复读的前文，对整个对话进行“脖子以下截肢”。

具体操作如下：

1. 利用总结姬、总结提示词或者其他工具形成一个总结。
2. 用/hide 0-你最新楼层 的命令，批量隐藏迄今为止的所有正文。
3. 用/sys 命令，以系统身份将总结以过去的摘要的名义发送。或者，在世界书内，将总结注入到角色定义后。或者，利用系统自带的总结工具，将摘要注入指定的深度/位置。
4. 将你最新的情况，人工描述给AI，让他回复你新一轮的故事。

在这种情况下，AI将类似于“开启全新世界线”，接续你的前情提要重新创作。用这类方法也可以处理没有长期谨慎维护，导致剧情线崩坏和文字胡言乱语的长对话。

# 计算机知识

这并不属于R1的任何方案，但它很有可能影响你游玩酒馆时的体验。一些玩家对于某些计算机知识存在常见误解。

正则表达式

**正则表达式不是插件也不是预设文件，正则本身不引导、不创造、不包含任何具体故事。‌**正则表达式（Regular Expression）是一种的文本模式匹配工具，用于对字符串进行搜索、替换和验证‌。

每个正则表达式都匹配特定格式的文本。例如表达式“/[突忽]然/g”匹配全文中所有的突然和忽然。形成匹配之后，可以根据需求对匹配到的内容进行替换操作。在R1中常见的用法时将所有匹配到的内容替换为空，即对全文每一个“突然、忽然”进行裁剪删除。

关于正则表达式的具体知识，譬如怎样写自己的正则，受限于篇幅本文无法赘述，请自行学习。

在酒馆中我们可以点开，点开，新建全局正则，

然后将表达式填写进“查找正则表达式”下方栏目之中。其中最小深度指的是影响深度从该数值开始，如果数字为5，则不影响最近的五条消息。最大深度同理，如果数字为5，则仅影响最近五条消息。

其中勾选格式显示，将影响对玩家的显示。勾选格式提示词，则影响对AI的呈递。

正则表达式由酒馆的js代码执行，具有严格性，可以严格进行某些形式明确的文本操作。如用正则将文本中所有的数字1替换为汉字一是可行的，但“将文本中所有的俄文翻译为汉语和英语”这显然是不可能的。

某些角色卡具备复杂的文本格式，需要正则表达式来帮助其整理和裁剪文本内容，因此使用时刚需特定的正则。或者，常见于用正则裁剪对用户显示的思维链内容。

**但内容本身依然全部由大模型产生，正则不可能满足形式不明确的文本修改需求。因此请不要过问“有没有能破甲的正则”，“有没有能增强情商的正则”的这类问题。这类问题通常属于“提示词工程”。**

JSON

JSON是一种用于数据交换的格式，用于存储和传输结构化的数据。

JSON独立于程序，在酒馆中预设、正则、世界书、插件配置等等都能以JSON形式导出并保存，**这些JSON之间不能混用**。世界书导出的JSON显然不能用于导入正则，请检查你的JSON是用于配置酒馆中的哪个部分。

# 或许……你懒得看这些，你只是想让我弄一个好用的预设给你

显然我的预设只是对我自己是最好用的（当然，它有可能对你也好用）。找一个大体好用的预设的确也是解决R1发癫的最好方案。

所以，你学习这个教程之前，不妨先尝试类脑的其他预设，看看有没有令你满意的效果。

不过，如果你能在此基础上继续调节个性化你的预设，不但能让他“不发癫”，还能更进一步达成前所未有的RP效果，输出高质量又能打动人心（或者打动你的其他器官）的文字。

或许你还能在这个过程里认识到， R1在良好调教的时候，也能产生优质流畅的内容。

是了，如果你爱她，鲸鱼娘也爱你。