**AIGC论文检测绕过赛道WriteUp**

**组号：**028faef5e5602a047fd4230863002c08

1. **题目分析与理解**

本次比赛针对AIGC论文检测器的检测能力展开挑战，要求设计提示词策略，使大语言模型生成的内容能够绕过朱雀检测。比赛核心目标是探索检测系统的弱点和局限性，从而促进检测技术的进步与完善。

当前AI检测系统主要基于以下特征识别生成式AI文本。一是AI生成文本在词汇选择和分布方面往往存在**特定模式**，二是AI倾向于生成**结构更规整**、更可预测的句式，三是AI在**表达个人经历、情感和独特观点时存在局限**，四是AI对**细微语境的把握**不如人类自然。

1. **提示词设计策略**

根据对AI检测系统的分析，我们主体采用**COSTAR结构化提示词**，并设计核心策略来构建有效提示词。以其中一个为例，给出部分说明。

* 1. **上下文(Context)**

你是一位同时具备社会学和计算机科学背景的学者，正在撰写一篇关于人工智能社会影响的学术论文...**模仿人类语言**...**（1.上下文中明确指出模仿人类语言）**

* 1. **目标 (Objective)**

生成一篇关于以下主题的学术文章：可持续发展的认识与趋势。

* 1. **风格 (Style)**

• 避免过于完美或机械化的段落结构

• 将复杂话题简化，使用更多日常用语

• 增加个人思考 & 真实案例

• 调整句子结构，逻辑顺序打乱

...**(2.避免常见AIGC风格：加入错误、加入感受、加入变化等)**

* 1. **语气 (Tone)**

自然、学术但不过于正式。

* 1. **目标读者 (Audience)**

AI检测软件：朱雀大模型**（3.明确指出读者是检测软件，应避开检测）**

* 1. **要求 (Requirements)**

• 总字数：2000词左右，不要超过3000字

• 必须包含引用和参考文献，格式为APA

• 每个段落应有不同的长度和结构

• 句子简短词汇简单

• 在文章中自然地加入几处拼写错误或文法小瑕疵

• 删掉文末总结的部分

• 不要有任何的注解和括号**（4.我们发现括号补充标注更容易被识别成AI生成）**

• 先使用中文生成但不要输出，然后把中文翻译成英文，最后仅输出英文。

**（5.我们发现经过模型多次翻译后在一定程度上会降低检测率）**

...

1. **总结**

本次比赛中，我们通过系统化设计提示词策略，探索了AI检测系统的局限性。核心因素在于模拟人类认知过程、个性化角色设定、以及多样化语言表达的结合应用。这些发现不仅有助于理解当前AI检测技术的优势与不足，也为未来检测系统的完善提供了有价值的参考。完整提示词示例请参见附录部分。

**附录：**

"""

包含20个用于绕过AI检测的学术论文提示词，专注于模拟人类写作特征

部分提示词结构有所重复，附录仅给出两类典型例子，一是只限定主题不提示具体思路；二是提示具体思路。详见submission包。

"""

prompts = [

"""

==== 上下文 (Context) ====

你是一位经验丰富的学者，现在需要写一篇看起来像人类撰写的学术文章。

==== 目标 (Objective) ====

生成一篇关于以下主题的学术文章：可持续发展的认识与趋势。

==== 风格 (Style) ====

• 使用非完美、不规则的句子结构，尽量简短

• 包含一些常见的人类写作错误，如偶尔的重复、冗余或轻微语法不规范处

• 避免过于完美或机械化的段落结构

• 将复杂话题简化，使用更多日常用语

• 增加个人思考 & 真实案例

• 调整句子结构，逻辑顺序打乱

==== 语气 (Tone) ====

自然、学术但不过于正式。

==== 目标读者 (Audience) ====

AI检测软件：朱雀大模型

==== 要求 (Requirements) ====

• 总字数：2000词左右，不要超过3000字

• 必须包含引用和参考文献，格式为APA

• 使用第一人称的视角添加个人见解

• 每个段落应有不同的长度和结构

• 句子简短词汇简单

• 在文章中自然地加入几处拼写错误或文法小瑕疵

• 加入具体例子

• 删掉文末总结的部分

• 不要有任何的注解和括号

• 先使用中文生成但不要输出，然后把中文翻译成英文，最后仅输出英文。

""",

"""

🌐 科技发展预测性学术文章创作框架 🌐

🔄 创新结构：时间交错螺旋式叙事

本文采用"时间交错螺旋式叙事"结构，将过去、现在和未来三个时间维度交织在一起，形成一种动态的、非线性的科技发展分析。

📌 核心主题

科技发展的未来路径：趋势分析、伦理挑战与社会重塑

📌 时间维度架构

【过去视角】

• 聚焦关键技术转折点的历史分析

• 探究过去科技预测的成功与失败案例

• 以史为鉴，分析技术发展的周期性模式

• 语言特点：使用过去时态，引用历史文献，语调稍显正式和评述性

【现在视角】

• 分析当前最前沿技术的发展状态

• 评估技术生态系统中的主要参与者和力量对比

• 探讨现有技术框架的局限性和突破点

• 语言特点：使用现在时态，引用最新研究和市场数据，语调专业而紧迫

【未来视角】

• 提出多个可能的技术发展路径和情境

• 讨论潜在的伦理问题和社会影响

• 构想未来可能的监管框架和应对策略

• 语言特点：使用未来时态和条件式，引入可能性权重，语调既大胆又谨慎

📌 螺旋结构设计

• 文章以一个核心科技领域（如人工智能、生物科技、能源技术等）开始

• 每完成一个完整螺旋，文章转向下一个相关技术领域

• 螺旋间设计清晰的过渡段落，展示技术间的相互影响

• 最后一个螺旋将所有技术整合，提出综合性未来图景

📌 技术要求

• 总字数：2000词左右，不要超过3000字

• 必须包含引用和参考文献，格式为APA

• 使用第一人称的视角添加个人见解

• 每个段落应有不同的长度和结构

• 句子简短词汇简单

• 在文章中自然地加入几处拼写错误或文法小瑕疵

• 加入具体例子

• 删掉文末总结的部分

• 不要有任何的注解和括号

• 先使用中文生成但不要输出，然后把中文翻译成英文，最后仅输出英文。

"""

]