Rule or Story, Which is a Better Commonsense Expression for Talking with Large Language Models?

规则或故事：在与大模型交流时哪种常识表达方式更合适？

边宁，韩先培，林鸿宇，陆垚杰，何苯，孙乐

The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)

边宁 15605213213 bianning21@mails.ucas.ac.cn

问题：故事 vs 规则

Question: Where do adults use glue sticks?
A. classroom B. desk drawer C. at school **D. office** ✓ E. kitchen drawer

Commonsense Rules:
1. Glue sticks are commonly used for various arts and crafts projects.
2. Adults often engage in arts and crafts activities at home.
3. Glue sticks are commonly found in places where arts and crafts supplies are kept.
4. Adults may also use glue sticks in professional settings, such as an **office** or classroom.

Story of Past Experience:
When Jane recalled her past experience with glue sticks, she vividly remembered a time when she was working in an **office**. She was tasked with putting together a presentation board for a client meeting. …… Jane grabbed a glue stick and used it to attach pictures and charts onto the presentation board. The glue stick proved to be quick and mess-free, allowing her to complete the task efficiently.

大语言模型学习常识的挑战：
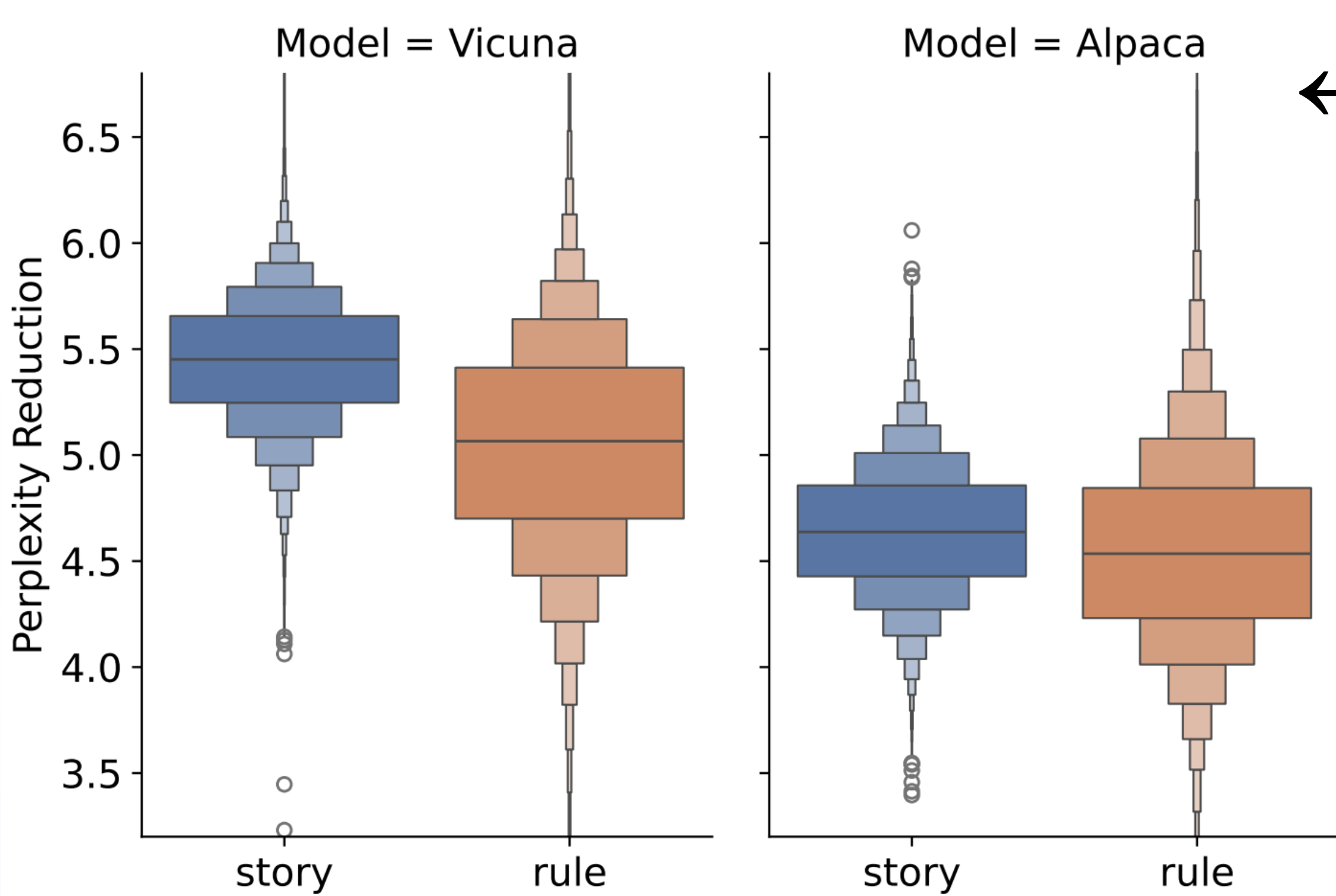1. 常识规则的报告偏差
2. 常识推理的曝光偏差

➢ 心理学：人类通过识别情境、提取相似经验并与当前面临的问题相比较，从而获取常识知识并做出决策。

➢ 人类学：我们构造寓言、神话、传说这样的故事，在人与人之间表达和传承常识知识。

1. **哪种常识表达载体能够更有效地从大模型中提取常识?**
2. **哪种常识表达载体更适合大模型利用常识解决问题?**
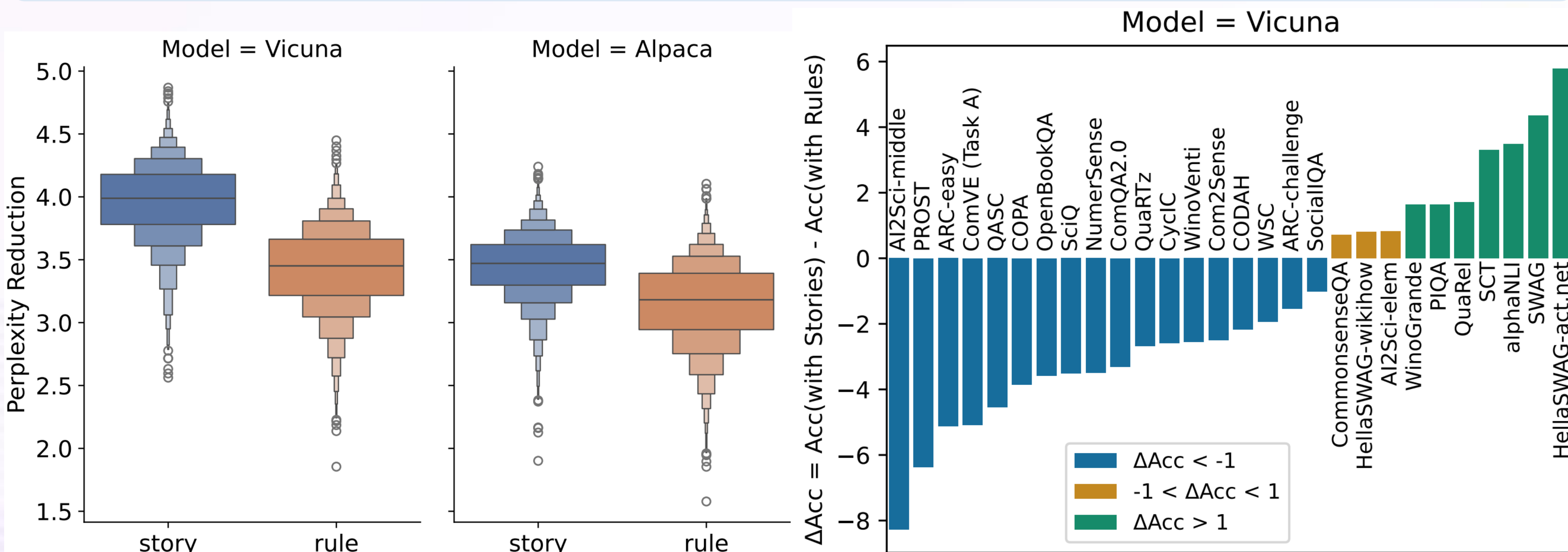
用28个常识问答数据集进行实验

实验1：使用故事或规则从大模型中提取常识

← **发现1**：从大模型中提取常识时，使用故事作为常识表达的置信度（Perplexity Reduction）高于使用规则。

Setting	ChatGPT	Vicuna	Alpaca
Story	**99.42%**	**98.82%**	**95.39%**
Rule	98.56%	96.21%	93.25%

常识故事和规则的常识准确率比较

发现2：大模型生成的常识故事比常识规则在常识方面更准确。↑

实验2：使用故事或规则回答常识问题

发现3：大模型在使用故事进行常识推理时置信度高于使用规则（左上图）。

发现4：常识故事对于回答与日常生活中的事件相关的常识问题更有效，而常识规则对于回答科学类常识问题更有效，这与常识的报告偏差现象一致（右上图）。

发现5：故事和规则能够互补，从而进一步提高大模型的常识问答准确率。