



ReadingQizMaker: A Human–NLP Collaborative System that Supports Instructors to Design High–Quality Reading Qiz Qestions

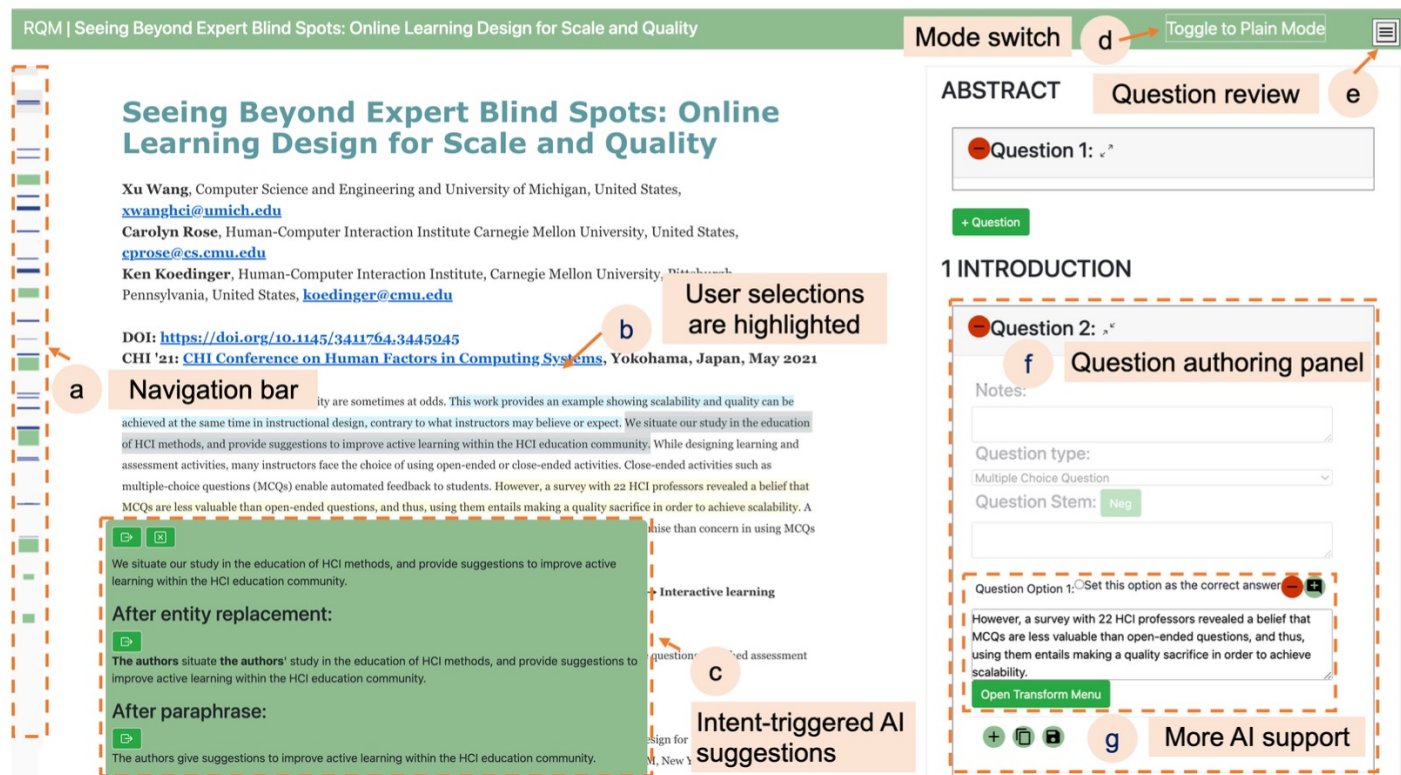
阅读报告

201250125 刘承杰



ReadingQizMaker: 一个人工智能与自然语言处理协作系统，支持教师设计高质量的阅读测验问题

- 导航栏 (a)、论文面板 (中间) 和问题创作面板 (右侧)
- 用户可以使用导航栏导航论文
- 绿色块表示表格和图形
- 用户可以在论文面板中选择内容，将其传输到问题创作面板以创建问题
- 根据用户对文本的选择提供人工智能建议 (例如，改述) (c)
- 用户的选择被突出显示 (b)
- 在问题创作面板上，用户可以编辑问题主干和选项 (f)，并预览更多的人工智能建议 (g)
- 用户可以在基于部分的模式和普通状态之间切换 (d)，并查看他们创建的所有问题 (e)



The screenshot displays the ReadingQizMaker interface. At the top, a green header bar contains the text "RQM | Seeing Beyond Expert Blind Spots: Online Learning Design for Scale and Quality". Below this, the main content area shows a paper titled "Seeing Beyond Expert Blind Spots: Online Learning Design for Scale and Quality" by Xu Wang, Carolyn Rose, and Ken Koedinger. The paper text is partially highlighted in blue, with a label "b" pointing to it. A label "a" points to a vertical navigation bar on the left side of the paper. A label "c" points to a green box containing AI suggestions for paraphrasing the text. On the right side, there is a "Question review" panel with a label "e" pointing to it. Below this, there is a "Question 2" panel with a label "f" pointing to the "Question type" dropdown menu. A label "g" points to a "More AI support" button at the bottom right. A label "d" points to a "Mode switch" button at the top right. A label "User selections are highlighted" points to the highlighted text in the paper.



- ReadingQuizMaker是一个协作系统，旨在帮助教师设计高质量的阅读测验问题。
- 它通过自然语言处理技术，支持教师在创建问题的过程中获得过程性的支持，包括选择何时使用NLP模型、选择模型的输入和编辑输出。
- 在评估研究中，教师们发现ReadingQuizMaker生成的问题与他们之前设计的测验问题相当。教师们赞扬ReadingQuizMaker易于使用，并认为NLP建议是令人满意和有帮助的。
- 与自动生成问题的基线条件相比，教师们更喜欢ReadingQuizMaker提供的人工智能协作方法。
- 我们的研究结果为支持教育的大型语言模型的使用提供了重要建议。
- 我们认为，在高风险任务（如教育内容创作）中，允许用户提供输入并给予用户足够的控制权比完全自动化的方法更可取。



- **大学阅读任务的低合规性：**只有20%–30%的本科生会为课程进行阅读，需要更好的实践来帮助学生阅读，通过降低阅读任务的难度，为阅读技能不足的学生提供支持和练习，并为学生的阅读过程提供反馈以提高参与度
- **支持阅读实践的策略：**学生无法获得关于他们对内容理解的反馈、设计高质量且引人思考的阅读测验问题可能会很困难
- **支持积极阅读的界面：**专注于开发技术来支持高等教育背景下的阅读理解
- **用于教育目的的问题生成技术：**引入了一种人工智能与人类协作的方法，其中在设计多项选择问题时，人工智能为人类教师提供过程导向支持。目标是使问题创作过程更为稳健和灵活，并产生更高质量的问题。
- **用于教育的人工智能系统：**研究了人工智能与人类协作的方法在支持教师创建测验问题方面的能力
- **人工智能系统设计准则：**通过与11名大学教授进行深入的形成性研究，并经过多轮试验测试，以确保人工智能与人类的互动对用户来说是直观且可取的。



对11位大学教师进行了一项形成性研究，以了解他们创建问题的自然工作流程，目的是理解教师手写问题时面临的独特挑战，并总结开发支持该过程的以用户为中心的系统的设计要求

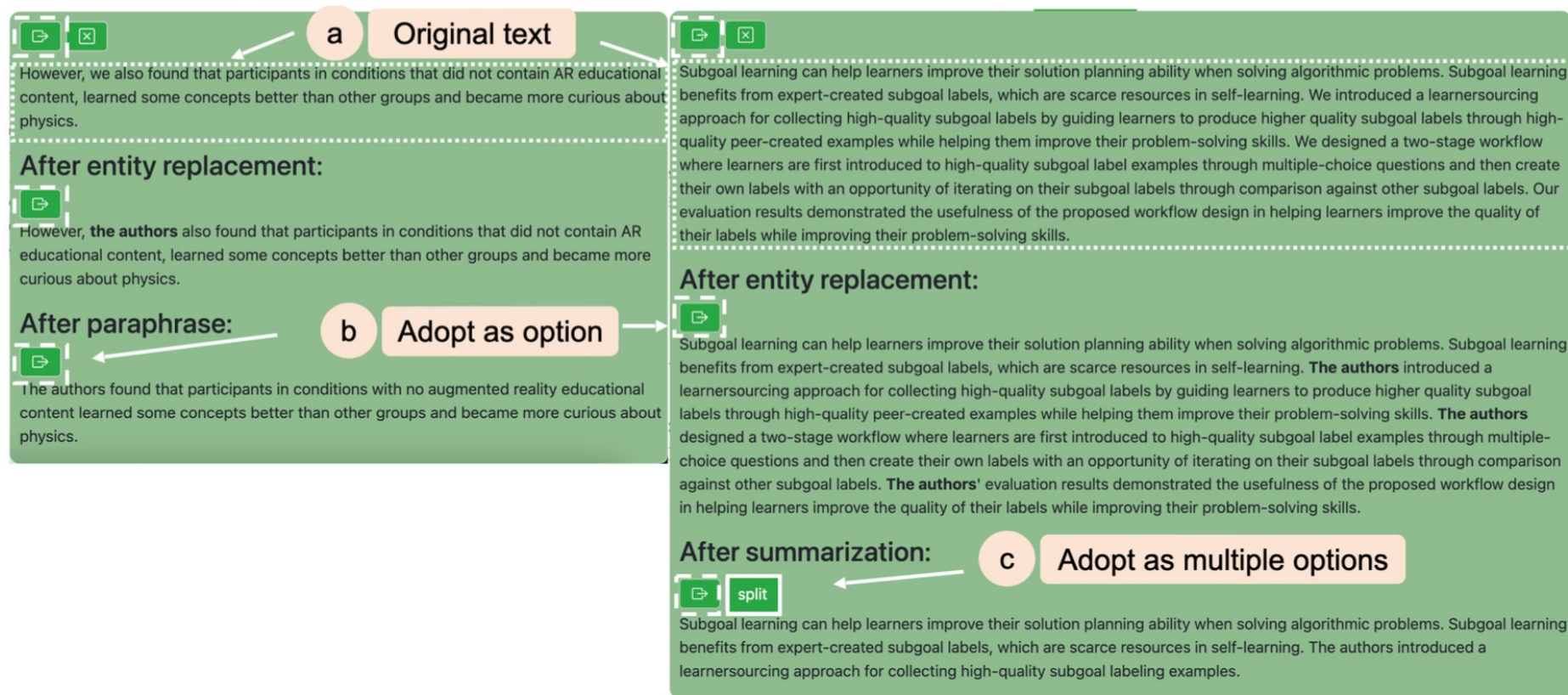
- 来自7所不同大学的11位教师参与了研究（男性6位，女性5位）。这些教师的教学经验范围从2年到40年不等，涉及计算机科学、信息科学、数据科学、教育、发展心理学和政治科学等学科。
- 采访通过Zoom进行，每次持续50到75分钟。参与者获得了50美元的礼品卡。
- 首先要求参与者分享他们如何处理阅读作业。大部分时间都花在让参与者根据他们选择的阅读文本设计问题上。
- 在会话期间，参与者能够设计3到10个多项选择题（一个问题干和四个选项）。然后，我们要求参与者反思他们是如何得出每个问题干和选项的，并分享他们在整个过程中遇到的挑战。最后，研究人员要求参与者想象有一个智能系统可以在测验设计过程中提供支持
- 将采访记录进行了转录，并使用亲和图分析数据



- 支持教师创建有说服力的干扰项。许多参与者表示在生成干扰项方面有困难，大多数参与者表示，他们希望在创建有意义的干扰项方面得到支持。
- 提供过程导向的支持，使教师能够融入他们的专业知识。教师不喜欢传统的端到端AI方法来创建问题。教师希望在需要时灵活决策并融入他们的专业知识，例如课程的背景、学生的背景知识等。
- 问题创建需要快速且与教师当前的工作流程整合。大多数参与者希望缩短他们在问题创建上的时间。
- 使教师能够轻松为问题编写反馈。反馈对教师非常重要，为问题提供反馈的机制，尤其是对于不正确的答案，是至关重要的。
- 在与AI交互时给教师一种控制感。当如果AI提供建议，那将是有效的，但他们更愿意保持控制并在创建问题时嵌入自己的知识。
- 给予教师在决策方面的灵活性 我们观察到不同的教师采用了不同的创建问题策略。教师对不同的问题类型有偏好。教师还希望问题库具有不同难度水平。



- 与任何具有HTML源的文章兼容。
- 导航栏支持用户对内容覆盖范围的阅读和导航到图表和表格
- 提供NLP建议的即时预览
- 适应用户的自然工作流程并使用户能够为选项编写反馈
- NLP工具箱
- Review and Output



ReadingQuizMaker 在用户的文本选择基础上立即提供AI建议，包括 1) 实体替换，将第一人称代词替换为第三人称代词；2) 对单个句子进行释义（左）；3) 对段落或长文本进行摘要（右）。用户可以选择采纳任何AI建议作为问题选项。用户还可以选择将摘要结果拆分为多个选项。

Question 2:

Notes:

Notes field for planning

Question type: Multiple Choice Question

Image:

Customer: Speed Inc.
Recommendation: Sell at \$100 per account.
Justification: The AI system considered the following information:
 [X] Gupta gains [X] Comparison pricing with other customers [X] Cost \$55 (recommended price)

For this customer, 3 members of your team received pricing recommendations in past sales. However, 1 out 3 have sold at the recommended price. Click to see more details.

Action: Project Recommendation
Comment: Long term profitable customer, earn revenue from a different vertical, selling at cost price to maintain relationship.
 On 3, 2019

Outcome: No Sale

Action: Accredit Recommendation
Comment: Recommended price aligned with profit margins, customer felt the price was fair.
 On 14, 2019

Outcome: Sale

Action: Project Recommendation
Comment: Covid 19 pandemic, cannot lose long term profitable customer, offered 10% below cost price.
 May 5, 2020

Outcome: Sale

Question Stem: Neg

Delete option

Question Option 1: Set this option as the correct answer

Feedback: Correct! Mentioned in the caption

Add/Edit feedback

Open Transform Menu

Add an empty option

Copy the last option

Save the question

Question Stem:

research **a** **Keyword Search**

Which of the following is NOT correct about the research contribution of this paper?

Which of the following are research questions asked in the paper? Select all that apply.

Why this is an important research question?

Which of the following are research methods used in the study?

Which of the following is NOT correct about what the paper does?

(Show More)

b **Stem suggestions based on section**

Question Stem:

Stem suggestions for image questions **c**

See above the Figure of the system design, which of the following is NOT a component in the system?

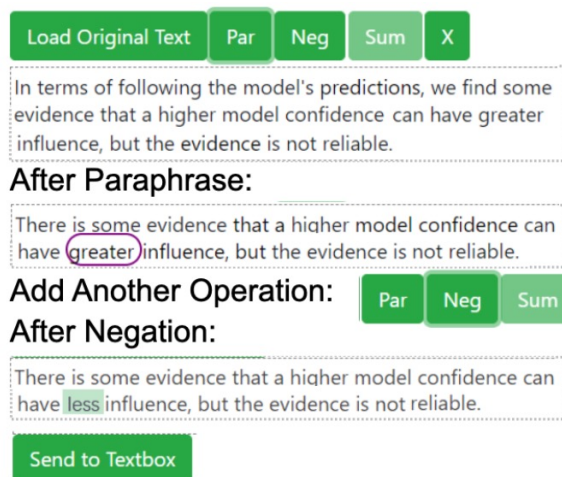
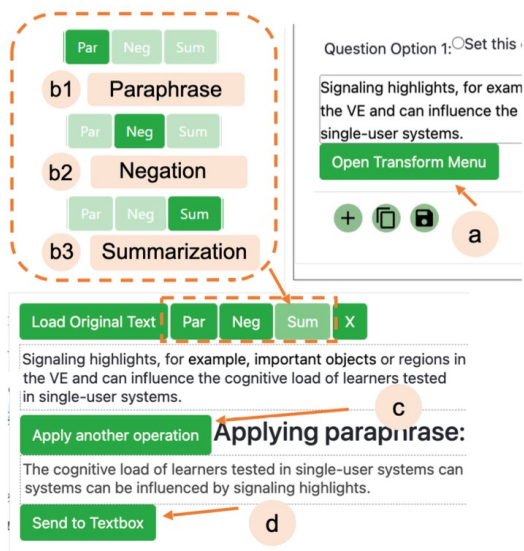
See the Figure above, which of the following is correct?

The Figure shows the study procedure, which of the following is correct?

(Show More)

ReadingQuiz-Maker中的问题创作面板是可折叠的。系统支持3种问题类型 (a) 。用户可以添加高层次规划的注释 (b) , 添加/删除选项 (e, c) , 为每个选项添加/删除反馈 (d) , 复制先前的选项 (f) , 并保存 (g) 。

ReadingQuizMaker帮助用户提出问题的主题。系统根据关键词提供主题建议（a）。系统根据用户所在的部分（例如，引言，相关工作），问题类型（单选题，多选题，开放式问题）以及用户是否选择了一个图（b, c）提供不同的建议。



转换菜单可以通过点击每个选项下的按钮 (a) 打开。该菜单将选项视为原始文本，用户可以应用释义 (b1)、否定 (b2) 或总结 (b3) 转换。他们可以选择在结果上应用另一操作 (c) 或使用AI转换的结果 (d)

通过链接NLP操作生成一个干扰项的示例。用户将释义操作与否定操作链接在一起。在否定操作中，用户选择要否定的单词“greater”。

Output Questions

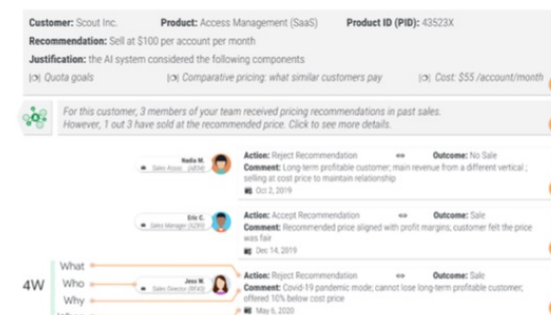
Question 1:

Which of the following is NOT correct about the research contribution of this paper?

Question 2: Figure for system design

See above the Figure of the system design, which of the following is NOT a component in the system?

Image:



A. a high-level summary of how many teammates in the past had received the recommendation

B. History actions from all users including receiving, accepting and rejecting recommendations

问题审查面板提供了所有已创建问题的概览



进行了一项经过IRB批准的评估研究，旨在了解ReadingQuizMaker的可用性和实用性。我们还对比了用户对ReadingQuizMaker作为人工智能协作系统与自动方法的反应，关注以下研究问题：

- RQ1: ReadingQuizMaker是否可用？教师能否使用ReadingQuizMaker创建令其满意的问题？
- RQ2: 教师如何看待人工智能建议？它们的质量是否令人满意？是否分散注意力？教师是否认为人工智能建议是有用的还是令人不满的？
- RQ3: 教师如何比较由ReadingQuizMaker提供的人工智能协作方法与自动问题生成方法？
- RQ4: 用户面临什么挑战，以及为开发教育领域的人工智能协作系统有哪些设计启示？

研究采用了混合方法，包括定量和定性数据的收集和分析。研究的参与者是12名教师，他们被要求使用ReadingQuizMaker创建问题，并回答有关他们的使用经验和反馈的问题。研究使用了多种数据收集方法，包括问卷调查、访谈、日志记录和系统生成的数据。研究的数据分析包括定量和定性方法，包括描述性统计、主题分析和内容分析。通过这些方法，研究评估了ReadingQuizMaker的可用性、有用性和效果，并提供了有关教师如何使用ReadingQuizMaker创建问题的见解。



- 对教师在创建阅读测验问题时面临的挑战进行了形成性研究
 - 介绍了带有NLP支持的ReadingQuizMaker
 - 对ReadingQuizMaker的可用性和实用性进行了评估
-
- ReadingQuizMaker支持教师设计高质量的阅读问题
 - 适应教师的工作流程并提供NLP支持
 - 评估显示了教育设计中人-AI协作方法的潜力



南京大學
NANJING UNIVERSITY

谢谢大家

THANK YOU

