各位尊敬的评委老师好，我们研究的题目是《人工智能与教师评分对结果公平性和满意度感知的影响》。当今世界人工智能发展迅速，其在教育领域有两大应用前景，包括AI辅助学习以及AI评分系统。在AI评分方面，先前一项研究发现，学生普遍认为AI评估系统相较于大学英语教师显得更公平且透明，但是该研究采用的是想象场景，存在自我卷入程度不足的问题。

因此我们设计了一个新的范式，采用真实评分场景，被试首先需要完成一道中翻英的题目，完成后进行自评打分，被试被告知其答案将由AI或教师进行评分，实际上被试得到的分数是由系统随机生成的，之后被试对四个维度的显性感知进行打分，包括结果公平性和满意度以及评分者公平性和满意度。

研究一采用2\*3组间设计，自变量包括评分者，ai和教师，实际得分低中高三个等级。研究一结果显示，评分者对于四个外显感知均无影响。实际得分高低显著影响满意度，实际得分越高，满意度越高，实际得分与期望得分的差异显著影响公平性，当两者差异最小时，公平性感知最高。

我们考虑到研究一的零结果可能是因为在实际的评分场景中，被试会有期望评分者，期望评分者与实际评分者一致与否会对外显感知产生影响，因此在研究二中我们引入了期望评分者。在参与者完成题目和自评后调查他们期望的评分者。我们在研究二中发现，总体上被试对于教师与AI不存在选择偏好上的差异，但按照期望得分分组后发现期望得分低组，被试偏好AI，期望得分高组，被试偏好教师。按照收益-损失框架，被试将教师视为保守的选择，因此在预期较高收益时，偏好选择教师，而将AI视为冒险的选择，在预期收益低时，偏好选择AI。研究二结果显示期望评分者与实际评分者一致与否对外显感知没有影响。

我们总结一下研究一和研究二的发现，被试对于AI和教师不存在偏好；评分者对于外显感知没有影响。

在许多有关内隐态度的研究中都发现了外显态度和内隐态度的分离，如种族偏见和性别偏见，因此我们猜想是否在本研究中存在同样的分离现象。因此在研究三中我们使用IAT范式探究被试对AI和人类的内隐偏好。

研究三发现，被试对于AI-受评，AI-受控以及人类-评价，人类-控制的概念存在隐性偏好。被试在内隐上更倾向于将AI与下位联系在一起，将人类与上位联系在一起。受到研究三的启发，我们在研究四中引入了能体现这种上下位关系的评价系统，AI辅助教师和教师辅助AI。研究四的目的是探索被试对于这两种评分系统的选择偏好和外显感知。偏好结果显示，“AI辅助教师”的模式显著高于“教师辅助AI”。但是在外显感知上，两种评分系统不存在显著差异。

在研究四中我们发现被试普遍偏好AI辅助人类教师，为了考察这种评分模式的应用前景，我们进行了研究5，想前瞻性地探究与单纯的AI或人类教师评价者相比，AI辅助教师是否有同样的选择偏好。结果显示，选择AI辅助教师的比例显著大于AI或教师，同时选择AI和选择教师的比例依旧不存在差异（进一步证实了在自我卷入的场景下，被试选择AI和选择教师的比例不存在差异），同样外显感知结果与前述研究类似。

我们在研究过程中还发现了一些有趣的结论。首先在研究二中，我们发现期望AI作为评分者的被试拥有更高的满意度和公平性感知，我们猜测可能是由于被试的一些内在人格特质（如经验的开放性，外向性）导致的，因此我们计划在未来的研究中尝试探究被试人格因素对于外显感知的影响。

另一个是在研究二和研究五中，实际得分与期望-实际评分者一致性存在交互作用，在研究二的低分组中，不一致组的外显感知显著高于一致组，在研究五中恰恰相反，我们对此结果的解读是，在研究二的评分者随机分配环节，我们给出的原因是出于算力限制，因此被试在遇到不一致情况的时候，被试会把实际低分更多地归因为外部因素，而在研究五中，由于三个选项的限制，我们并没有给出分配的原因，所以被试在遇到不一致情况的时候，会更多地把实际低分归因于评价者内部因素，导致外显感知得分更低。根据这一结果，我们推测分配原因的可解释性对于被试的外显感知有重要影响。

本研究采用了模拟真实考试场景的范式，实现了被试的高度自我卷入，从而可以更真实地评估对于AI与人类教师的外显感知，结果表明被试对于AI与人类教师不存在外显的选择偏好，但是在内隐态度上存在对于AI的下位偏好。基于上述分离现象，我们进一步发现被试存在对于AI辅助人类这一评分模式的外显偏好，并与单纯的AI或人类教师评价者相比，重现了上述这一偏好。

首先，我们的研究结果表明，在教育评分这一特定领域，人工智能评分系统并没有明显优于人类教师，这挑战了先前研究中关于AI评分系统更公平、更透明的观点。这一发现对于教育政策制定者来说至关重要，因为它提示我们在推广AI评分系统之前，需要更全面地考虑其对教育公平性和学生满意度的影响。

其次，我们的研究揭示了人们在内隐态度上对AI的下位偏好，这可能与AI的去人格化特性有关。这种内隐偏好的存在，对于人工智能技术的设计和应用提出了新的挑战。在未来的AI设计中，我们可能需要考虑如何通过增加人性化元素来提高人们对AI的接受度和信任度。

此外，我们的研究还发现，当AI与人类教师结合使用时，可以显著提高人们对于评分系统的偏好。这一发现对于教育实践具有重要的启示意义。它表明，在未来的教育评分系统中，可能需要更多地考虑如何将AI技术与人类教师的专业知识和情感智慧相结合，以创造出更高效、更公正、更受欢迎的评分模式。

最后，我们的研究还指出了期望评分者与实际评分者一致性、分配原因的可解释性等因素对于外显感知的影响。这些发现对于理解人们如何在面对AI评分系统时形成其感知和偏好具有重要的理论价值。它们提示我们在设计和实施AI评分系统时，需要考虑如何通过提高透明度和可解释性来增强用户的信任和满意度。