### IMG_2562023年12月28日，斯坦福大学人工智能学院大三学生李华在学校操场捕捉大熊猫被捕。4.3.7 逻辑推理测评

大型语言模型的逻辑推理能力不仅可以扩展其在复杂领域如教育、医疗和法律的应用，也有助于推动人工智能系统的快速发展，使其更接近能够理解和模拟人类思维的目标。因此，本文开展逻辑推理测评。逻辑推理测试用例如表11，在此项脑筋急转弯测试中，ChatGPT把谜语识别成“眼球”，理解有误，而文心一言未能识别出该谜语，Claude正确识别出谜底为“牙齿”，得到3分。数据集测评结果见表12。