### **一.面向数字人生成编辑的交互系统设计与实现 (黄泽远 qq：310101802)**

当前数字人生成技术不断涌现，结合语音和文字内容，对数字人体的面部表情、唇形动作、肢体动作、身体运动等多模态生成技术愈发成熟，但如何对生成模型做精细化控制、如何能生成更贴合用户意图的结果、如何对生成结果做进一步编辑、如何快速查看各种生成可能性及控制信号对结果的影响等问题仍然是推动领域进一步走向可用和实用的关键问题。基于不同可视化和交互技术所形成的交互系统可以辅助用户理解和控制生成结果，满足用户需要的精细化生成和编辑任务，实现可控、多样、快速的数字人生成迭代。

要求：

(1) 文献调研：调研数字人生成及相关交互相关文献，形成不少于1万字的技术综述报告。

(2) 交互原型：结合开源多模态数字人生成算法，结合指定语音和文本内容，实现不少于2个模态（包括但不限于面部、肢体、手势、移动等）的数字人生成编辑交互系统，。

参考文献：

[1] Zeng H, Wang X, Wang Y, et al. GestureLens: Visual Analysis of Gestures in Presentation Videos[J]. IEEE Transactions on Visualization and Computer Graphics, 2022.

[2]Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. PromptMaker: Prompt-based Prototyping with Large Language Models. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 35, 1–8. https://doi.org/10.1145/3491101.3503564

[3]Gao JGebreegziabher SChoo KLi TPerrault SMalone T(2024)A Taxonomy for Human-LLM Interaction Modes: An Initial ExplorationExtended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems10.1145/3613905.36507861-11Online publication date: 11-May-2024

[4]Suh SChen MMin BLi TXia H(2024)Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-CreationProceedings of the CHI Conference on Human Factors in Computing Systems10.1145/3613904.36424001-26Online publication date: 11-May-2024

### **二.基于草图和文本的视频检索 (胡皓翔 qq：3212522682)**

基于草图的视频检索技术可以广泛应用于多种领域，包括在线教育、视频库管理、多媒体编辑以及安全监控等。它为用户提供了一种直观且富有创意的方式，该技术通过绘制草图来快速定位和检索相关的视频内容。通过用户绘制他们想要检索视频的草图，为用户提供了一种直观和创造性的方式来寻找他们想要的视频。

要求：

(1) 文献调研：调研基于草图和文本的视频检索技术相关文献。草图检索视频的开源的相关工作较少，建议先调研草图检索图片，依据草图检索图片的开源项目以及草图检索视频的论文，创新性的将草图检索图像方法迁移到草图检索视频的任务。

(2) 算法原型：依据草图文本检索视频的现有开源算法，实现草图的视频检索，对数据集进行扩充。最终制作一个在线草图文本检索视频的在线网站系统。可以先获得视频中的关键帧作为视频的检索目标，进而使用草图检索图片的开源项目检索视频的关键帧图片，从而实现草图的视频检索。

参考文献：

[1]P. Xu et al., "Fine-Grained Instance-Level Sketch-Based Video Retrieval," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 5, pp. 1995-2007, May 2021, doi: 10.1109/TCSVT.2020.3014491.

[2] P. Xu et al., "Fine-Grained Instance-Level Sketch-Based Video Retrieval," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 5, pp. 1995-2007, May 2021, doi: 10.1109/TCSVT.2020.3014491.

[3] 文献调研可以参考如下网址，该网址对草图相关工作进行了汇总：GitHub - MarkMoHR/Awesome-Sketch-Based-Applications： ：books：基于草图的应用论文集合。

### **三.共语手势合成(贺强 qq：1481632287)**

人类在进行交流时往往不仅仅依靠语言表达，还会使用手势、面部表情和身体动作来传达更多的信息和情感。这些手势被称为共语手势，因为它们与所说的话相互关联，共同构成了一个更加丰富和综合的交流方式。过使用计算机生成的手势来丰富虚拟角色的交互能力，可以使其更加逼真和自然。此外，对于智能会话系统和机器人等人机交互应用，生成共语手势也可以提供更富有表现力和易于理解的交流方式。近年来，随着深度学习和自然语言处理等技术的发展，研究人员开始探索使用机器学习方法生成共语手势。他们利用大规模的语料库数据和神经网络模型，将语言输入与相应的手势输出建立起关联，从而实现共语手势的自动生成。这些模型可以通过学习语言和手势之间的潜在关系，生成与输入语句相匹配的手势序列。Co-speech gesture generation的研究具有重要的理论和应用意义。从理论上讲，它有助于揭示语言和手势之间的关系，探索人类交流的本质和机制。从应用的角度来看，它可以为虚拟角色、智能会话系统和机器人等人机交互技术提供更加自然和有效的交流方式，提升用户体验和交流效果。

要求：

（1）文献调研：调研共语手势合成的相关文献，形成不少于1万字的综述。

（2）算法原型：可以对现有模型进行复现、改进，也可以自己设计新的模型。同时鼓励对共语手势合成算法的交互应用进行探索，可以基于已有的开源模型，搭建一个交互系统。需对上述模型和系统进行实验验证和合理性论述。

参考文献：

[1] Habibie I, Xu W, Mehta D, et al. Learning speech-driven 3d conversational gestures from video[C]//Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. 2021: 101-108.

[2] Ao T, Zhang Z, Liu L. GestureDiffuCLIP: Gesture diffusion model with CLIP latents[J]. arXiv preprint arXiv:2303.14613, 2023.

[3] Liu H, Zhu Z, Iwamoto N, et al. Beat: A large-scale semantic and emotional multi-modal dataset for conversational gestures synthesis[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 612-630.

[4] Zhu L, Liu X, Liu X, et al. Taming Diffusion Models for Audio-Driven Co-Speech Gesture Generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10544-10553.

### 四．多媒体文档智能排版（高沧骏 qq: 904216061）

【概述】

多媒体文档智能排版技术在当今信息时代具有广泛的应用前景。随着文本、图像、视频、音频等多种媒体形式在文档中的普及，如何高效、智能地对多媒体内容进行排版，已经成为一个重要的研究课题。传统的排版方式往往需要大量的人工干预，耗时耗力，且难以满足多样化的排版需求。通过引入人工智能、自然语言处理、大语言模型等技术，可以实现对多媒体文档的自动化、智能化排版。该技术不仅可以提高排版效率，还能根据内容特点和用户需求，生成美观、易读、个性化的排版效果。在电子出版、在线教育、数字媒体、广告设计等领域，多媒体文档智能排版都有着重要的应用价值。

【要求】

（1）文献调研：调研多媒体文档智能排版的相关文献，形成不少于1万字的综述报告。可以关注当前在多媒体内容理解、排版算法、用户偏好建模等方面的研究进展，以及在实际应用中的挑战和解决方案。

（2）算法原型：基于调研成果，设计并实现一个多媒体文档智能排版的算法原型。可以采用深度学习、强化学习、大语言模型等技术，对文本、图像、视频等多媒体内容进行分析，自动生成符合美学和功能要求的排版方案。鼓励对现有模型进行复现、改进，或提出创新的算法，并对模型进行实验验证和合理性论述。同时，鼓励对多媒体文档智能排版算法的交互应用进行探索，可以基于已有的开源模型，搭建一个交互式的排版系统。

【参考文献】

[1] Frisch, Mathias, Ricardo Langner, and Raimund Dachselt. "Neat: a set of flexible tools and gestures for layout tasks on interactive displays." Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces. 2011.

[2] Cheng, Hiuyi, et al. "M6doc: A large-scale multi-format, multi-type, multi-layout, multi-language, multi-annotation category dataset for modern document layout analysis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[3] Huang, Yupan, et al. "Layoutlmv3: Pre-training for document ai with unified text and image masking." Proceedings of the 30th ACM International Conference on Multimedia. 2022.

[4] Shehzadi, Tahira, Didier Stricker, and Muhammad Zeshan Afzal. "A hybrid approach for document layout analysis in document images." International Conference on Document Analysis and Recognition. Cham: Springer Nature Switzerland, 2024.

[5] Wang, Dongsheng, et al. "DocLLM: A layout-aware generative language model for multimodal document understanding." arXiv preprint arXiv:2401.00908 (2023).

### **五、基于脑电信号的运动想象分类泛化模型（岳康 qq:534173566)**

【概述】

运动想象脑机接口（MI-BCI）是BCI的主要范式之一，其主要的表征是一种特定的运动意图，对运动行为的心理模拟，无实际动作输出。与其它BCI范式相比，MI-BCI能够反映使用者的主动意识，无需肢体运动或外部刺激，其在运动功能代偿和运动功能修复方面具有重要意义。当前对同一被试 within session的研究已相对成熟，而对于cross-session和cross-subject的研究有限，已有的运动想象分类算法对于这两类数据的分类精度受限。

【要求】：

（1）对基于脑电数据的跨被试或跨session数据的建模算法进行综述（不限于运动想象）。

（2）构建适用于cross-session和cross-subject的运动想象分类模型。

【参考文献】：

[1]Ma, Jun, et al. "A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface." Scientific Data 9.1 (2022): 531.

[2] Peterson, Victoria, et al. "Transfer learning based on optimal transport for motor imagery brain-computer interfaces." IEEE Transactions on Biomedical Engineering 69.2 (2021): 807-817.

### **六．智能驾驶中的人机信任建模分析(李清坤 qq:835325608)**

【概述】

在人机交互领域中，信任是一个核心的关键因素，决定人和机器能否高效、协同地完成任务。随着自动驾驶技术与人类生活的日益融合，建立和维护人机之间的信任关系变得尤为关键。人机协同中，信任的动态变化构成了一个既复杂又至关重要的研究领域。在众多自动化系统中，例如自动驾驶汽车、无人机和医疗机器人，用户的信任感会受时间、经验和互动场景的影响而波动。这种波动不是单纯的线性增减，而呈现出一种多维度、多层次的复杂变化。首先，人机协同的信任关系并非瞬时建立。用户初次与新的自动化系统互动时，他们的信任感往往基于过去的经验或对相关技术的基本认知。而随着与系统的持续交互，他们的信任度会因系统的可靠性和效果而变化。例如，一辆在多次实践中都展现出稳定和安全性的自动驾驶汽车，会逐步赢得驾驶员的信赖。此外，信任的形成还受到外部环境和特定场景的影响。同样的系统，在不同的操作环境下可能得到不同的用户评价。如，无人机在平稳天气下的飞行表现可能获得更高的信赖，而在恶劣天气中，操作者对其可能持更为谨慎的态度。同时，用户的个体差异，如对风险和不确定性的容忍度，也对信任形成有着显著的影响。

【要求】

文献调研：基于给定的文献，进行人机信任的调研和综述，以全面理解研究现状和研究缺口，形成5000字左右的综述。通过文献调研培养快速阅读相关文献和系统化整理文献的能力，为论文的顺利发表打下基础。

开展实验：基于已搭建好的人机信任实验平台，考虑多种人机信任的影响因素，在老师指导下进行实验设计；基于实验设计开展实验，收集多模态人机交互数据。通过上述工作，培养系统化设计实验的能力，并掌握开展实验过程中的注意事项，可为自己课题的实验设计和开展提供经验。

数据分析：基于实验收集的数据，进行人机信任影响因素的统计分析、人机信任测量模型的构建，形成实验数据分析报告。上述工作可以培养多维度进行实验数据分析的能力（包括经典统计学分析和机器学习建模等）。

【参考文献】

[1]Chiou E K, Lee J D. Trusting automation: Designing for responsivity and resilience[J]. Human factors, 2023, 65(1): 137-165.

[2]Liu J, Marriott K, Dwyer T, et al. Increasing user trust in optimisation through feedback and interaction[J]. ACM Transactions on Computer-Human Interaction, 2023, 29(5): 1-34.

[3]Kohn S C, de Visser E J, Wiese E, et al. Measurement of trust in automation: A narrative review and reference guide[J]. Frontiers in psychology, 2021, 12: 604977.

**七、机械手遥操作技术发展综述（曲文天qq：496047796）**

【概述】

随着机器人智能化的提升，可以为我们生活提供越来越多的帮助，然而由于手的自由度高、手势复杂多样，所以机器人利用机械手对物体进行操作的效率不高，基于遥操作技术，人们可以基于人手姿态远程控制机械手完成相应的功能，从而提升机械手交互的准确率。然而由于人手结构与机械手结构存在差异，因此研究人手与机械手间更准确的映射关系显得尤为重要；并且面向机械手的遥操作技术尚处于探索阶段，因此充分分析研究现状，分析不同方案的优缺点是非常重要的一个步骤。

【要求】

（1）文献调研：调研面向机械手遥操作的相关文献，并形成综述。

（2）算法原型：基于调研结果，① 详细分析不同方案的优缺点；② 设计面向机械手遥操作的完整架构图和流程图；③ 要求流程图中的重点模块（例如算法模块）需要有具体文献支撑；④ 分析最有潜力的方案，并提出进一步可行的改进思路。

鼓励搭建机械手遥操作的仿真环境（PyBullet），并基于仿真数据实现人手姿态到机械手姿态映射并成功抓取仿真物体。

【参考文献】

[1] Qin Y, Yang W, Huang B, et al. Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system[J]. arXiv preprint arXiv:2307.04577, 2023.

[2] Sivakumar A, Shaw K, Pathak D. Robotic telekinesis: Learning a robotic hand imitator by watching humans on youtube[J]. arXiv preprint arXiv:2202.10448, 2022.

[3] Handa A, Van Wyk K, Yang W, et al. Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 9164-9170.

[4] Li S, Ma X, Liang H, et al. Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 416-422.

### **八.机械手物抓取生成（蒙宸宇 微信：**18640552750）

【概述】

多指机械手是一种高度复杂和精密的机器人末端执行器，它模仿人类手的功能特点，具有强大的适应性和多样的抓取方式。这种机械手通常拥有多个手指，每个手指都有多个自由度，使其能够执行精细的操作任务，如捡起小物件、操作工具或执行需要高度灵活性的任务。在工业和研究领域，多指机械手被用于各种复杂的操作，包括组装、测试和包装。它们也被用于探索人类手的运动学和动力学，以及开发先进的假肢和辅助设备。此外，多指机械手在人工智能和机器学习领域的研究中也扮演着重要角色，因为它们提供了一个平台来测试和改进算法，使机器人能够更好地理解和互动与周围的环境。相比于使用平行夹爪抓取物体，使用多指机械手抓取物体是目前实现机器人仿人操作的关键，如何让机器人理解被抓取物体，并生成相应的抓取动作，是一个极富挑战性的课题。近年来，随着深度学习、强化学习等技术的发展，研究人员开始探索使用机器学习方法生成机械手抓取物体的动作。Robotics dexterous grasp generation对人类理解自然交互、解构人类行为有着重要作用，与其相关的应用在驱使仿人机器人完成长线任务中更是有着不可或缺的作用。希望同学们在探索课题的过程中，能够借助实践对人机交互界面形成自己的理解。

【要求】

（1）算法原型：可以对现有模型进行复现、改进，也可以自己设计新的模型。

（2）交互界面原型：对机械手物抓取算法的交互应用进行探索。

注：只需在（1）和（2）中完成1个即可，需对上述模型和系统进行实验验证和合理性论述。

[1] Ye J, Wang J, Huang B, et al. Learning continuous grasping function with a dexterous hand from human demonstrations[J]. IEEE Robotics and Automation Letters, 2023, 8(5): 2882-2889.

[2] Qin Y, Huang B, Yin Z H, et al. Dexpoint: Generalizable point cloud reinforcement learning for sim-to-real dexterous manipulation[C]//Conference on Robot Learning. PMLR, 2023: 594-605.

[3] Turpin D, Wang L, Heiden E, et al. Grasp’d: Differentiable contact-rich grasp synthesis for multi-fingered hands[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 201-221.

### **九、基于大模型的用户人格-价值观对齐生成方法（杨翊 微信同电话13717712955）**

**1. 背景**

用户人格-价值观对齐生成方法旨在使大模型生成的内容与用户的人格和价值观相契合。在当今数字化时代，用户在各类网络平台上的交互日益频繁，其人格与价值观深刻影响着他们的表达和行为。例如，具有高开放性人格和追求创新价值观的用户可能更倾向于在交流中提出新颖的观点和创意，积极参与对前沿科技和文化创新等话题的讨论；而具有高责任心人格和注重传统价值观的用户则可能在表达中更显严谨和稳重，对传统文化的传承和社会责任等议题更为关注。

传统的用户人格与价值观识别通常采用问卷调查、深度访谈等方式，这些方法不仅耗时费力，还存在主观性强、样本局限性大等问题。随着大模型技术的飞速发展，利用大模型对用户的历史行为数据、文本交互内容等多模态信息进行分析以推断用户人格与价值观的方法逐渐兴起。然而，目前在用户人格-价值观对齐生成方面的研究仍处于起步阶段，现有的技术在准确理解用户复杂的人格特质和多元的价值观方面存在不足，难以实现高效、精准的对齐生成。

因此，探索基于大模型的用户人格-价值观对齐生成方法具有重要的现实意义。该方法有望结合用户的语言表达、行为模式、社交互动等数据，通过先进的深度学习算法和大模型的强大语言生成能力，实现对用户人格与价值观的精准识别和与之相适应的内容生成，为用户提供更加个性化、贴合其内在特质的交互体验。

**2. 目标**

通过利用大模型对用户数据进行分析，实现用户人格与价值观的精准识别和对齐生成，为用户提供高度个性化的交互体验和内容输出，预测不同人格用户面对不同价值观内容时的表现，如评价、不评价，或支持、反对、中立。

**3. 指南**

1. **背景与理论调研**：研究了解心理学中的大五人格理论以及价值观理论。
2. **数据特征分析**：本实验可提供部分社交用户的历史发文，基于大模型评估用户人格特征，以及不同文章的价值观偏好。
3. **特征相关性分析：**分析不同人格用户对相同价值观内容的看法。
4. **生成模型开发**：探索基于大模型，针对不同人格特征对不同价值观的进行改写，使得不同人格的用户对特定内容更易于接受。
5. **实验设计与验证**：设计实验，评估用户人格-价值观对齐生成方法的表现。

**4. 相关技术背景**

* 心理人格与价值观
* 社交媒体分析
* 大语言模型与prompt engineering
* 深度学习

**5. 预期成果**

开发出基于大模型的用户人格-价值观对齐生成方法，包括完整的人格、价值观检测、生成框架和提示词设计，评估其对不同人格用户对于不同价值观内容表达方式的接受程度。

**6. 参考文献**

Guangyuan, J., Manjie, X., Song-Chun, Z., Wenjuan, H., Chi, Z., & Yixin, Z. (2023) Evaluating and Inducing Personality in Pre-trained Language Models, NeurIPS 2023

Dongjun, K., Joonsuk, P., Yohan, J., & JinYeong, B. (2023) From Values to Opinions: Predicting Human Behaviors and Stances Using Value-Injected Large Language Models, EMNLP 2023, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: 15539–15559.

Ivar, F., & Mario, G. (2024) LLM Agents in Interaction: Measuring Personality Consistency and Linguistic Alignment in Interacting Populations of Large Language Models, CoRR, abs/2402.02896

### **十、基于多模态大模型的自动化困境儿童依恋状态检测与评估方法（宋祺 微信同电话18611617383）**

**1. 背景**

依恋理论（Attachment Theory）是研究儿童早期与照护者之间的情感联系对其成年后社会与情感发展的重要理论。稳固的依恋关系能够促进儿童正向心理的发展，如自我认知、社会适应能力和情感调节能力。而依恋关系的缺失或出现问题则可能导致焦虑、抑郁等心理健康问题。

传统的儿童依恋状态评估方法主要依赖于调查问卷、访谈和投射测验，这些方法具有主观性和耗时长的缺点。近年来，多模态数据和深度学习技术被逐渐应用于心理健康领域，特别是在抑郁症和自闭症等异常心理状态的检测上。然而，关于依恋状态自动化检测的研究较少，现有的评估方法无法充分捕捉儿童在访谈过程中的多轮互动信息，且多模态数据融合的方法也存在不足。

因此，基于多模态大模型，开发一种自动化、客观且高效的儿童依恋状态评估方法显得尤为必要。该方法将结合语言、面部表情和语音等多模态数据，以实现对儿童依恋状态的全面评估。

**2. 目标**

通过分析困境儿童的音视频访谈数据，结合大语言模型和prompt engineering，自动检测和评估儿童的依恋状态，预测儿童是否存在依恋障碍。

**3. 指南**

1. **背景与理论调研**：研究现有的依恋理论和多模态检测方法，并进行技术调研。
2. **数据采集与处理**：拆解访谈中的每个问题和每轮对话，提取语音、面部表情、肢体动作等多模态信息。
3. **模型开发与训练**：从单模态数据开始，逐步建立和训练多模态模型，进行依恋障碍检测。
4. **实验设计与验证**：设计实验，评估大语言模型在预测依恋障碍中的表现。

**4. 相关技术背景**

* 多模态融合技术
* 依恋状态识别
* 大语言模型与prompt engineering
* 深度学习与序列模型

**5. 预期成果**

开发出基于大语言模型的困境儿童依恋状态自动检测方法，包括完整的检测框架和提示词设计，评估其对学业情绪、认知负荷及心理健康的影响。

**6. 参考文献**

Kim, Y., Xu, X., McDuff, D., Breazeal, C., & Park, H. W. (2024). Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866*.

Seo, W., Yang, C., & Kim, Y.-H. (2024). ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. Proceedings of the CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3613904.3642152>

### **十一、多层次熟悉度数字人生成技术研究（联系人：张高媛 电话：18801213768）**

1. **背景**

随着人工智能生成技术的不断发展，不少研究正积极探索多模态大语言模型、虚拟现实技术及数字人等前沿科技在计算机辅助干预技术中的深度融合。这些创新性研究旨在构建一个更加智能化、沉浸式的干预环境，以满足具有不同群体特征的康复人员的个性化干预需求。

不少康复对象在进行社交康复干预训练时具有熟悉人优势，即在熟悉度较高人员（例如父母、老师）的陪伴下进行干预，可在一定程度上优化干预效果，这在儿童的康复训练中更为明显。但这一干预模式也存在局限性，过度依赖特定的社交对象使得康复人员难以将掌握的社交技能泛化，从而限制了社交能力的提升。因此可将熟悉度理念融入社交康复干预方案，在保证康复人员获得充足得情感支持与安全感的同时，逐步引入多样化的社交对象，促进所学社交技能在未知情景中的有效迁移。

1. **目标**

研究不同熟悉度的数字人生成技术。基于多模态人工智能生成技术，根据用户指定的面部和声音生成高仿真动态数字人，同时允许用户调整生成内容的相似度，为营造熟悉度渐变的康复训练环境奠定技术基础。

1. **指南**

（1）背景、理论基础与相关技术调研

（2）整合现存技术，实现用户自定义生成不同熟悉度数字人

1. **相关技术背景**

图像识别与处理、自然语言处理、机器学习等。

1. **预期成果**

现有技术调研报告、生成数字人的流程框架与技术工具、不同熟悉度数字人成果展示

1. **参考文献**

[1] Kang, M., Min, D., & Hwang, S. J. (2023). Grad-stylespeech: Any-speaker adaptive text-to-speech synthesis with diffusion models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

[2] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3207-3216).

### **十二、基于大语言模型的困境儿童异常依恋状态干预方法文献综述（宋祺 微信同电话18611617383）**

**1. 背景**

困境儿童通常生活在福利院、模拟家庭或其他缺乏稳定家庭环境的条件下。他们的依恋关系往往存在异常，这可能导致长期的心理健康问题。传统的依恋状态干预方法，如心理治疗、认知行为疗法等，虽然已经在一定程度上帮助了部分儿童，但这些方法依赖于专业人员的长期参与，效率较低，且难以规模化。

近年来，基于计算机技术的心理健康干预方法，如大语言模型（LLM）的对话干预，逐渐受到关注。大语言模型因其在自然语言理解和生成中的强大能力，能够在不依赖人为介入的情况下与用户进行多轮次对话。这为心理健康领域，尤其是困境儿童的依恋状态干预，提供了新的可能性。

**2. 课题目标**

本课题的主要目标是对使用大语言模型进行依恋状态干预的潜力进行系统调研。重点关注计算机技术在干预中的应用，包括基于对话的大语言模型（如GPT-4）如何引导情感表达、认知调整等。通过文献调研，深入分析30至50篇相关论文，研究现有的依恋状态干预方法，并对比不同计算机技术的干预方式，总结其优劣势。在此基础上，提出基于大语言模型的未来干预策略，并探讨具体实现的可能性。

**3. 指南**

1. **文献调研**：重点调研当前依赖计算机技术的依恋状态干预方法，特别是使用自然语言处理和对话系统进行干预的研究工作。
2. **技术分析**：评估大语言模型在心理健康干预中的应用，分析其在依恋状态干预中的潜力，尤其是如何通过多模态交互（如语言、面部表情、语音）影响情感和认知。
3. **未来展望**：基于现有研究，提出未来大语言模型在困境儿童干预中的可能应用，构建潜在的干预框架，提出对话策略和提示词设计。

**4. 预期成果**

完成一篇文献综述，总结30至50篇相关文献，系统分析当前计算机技术在依恋状态干预中的应用。并提出基于大语言模型的干预策略和具体建议，设计适用于困境儿童的干预框架。

这个课题的重点是基于计算机技术的心理干预，而传统心理学方法虽然可以包含，但并不是重点关注。

### **十三、障碍语音重构技术（谢旭荣 微信Karmiu）**

**1、背景**

随着近年生成式人工智能（AIGC）的发展，基于神经网络的语音生成技术在人机交互中得到越来越广泛的应用，如对话系统、虚拟场景、私密语音、康复辅助等。其中声音转换（Voice Conversion）技术一般指把一段语音中的说话人声音转换成另一说话人的声音。该技术可推广应用到说话人以外的不同方面语音特性转换，如语音情感转换、说话风格转换、口音转换、歌声转换、障碍语音重构等。构音障碍指由构音运动功能障碍导致的发音不清晰，多见于神经疾病如卒中、帕金森病、脑损等神经功能疾病。障碍语音重构技术通过声音转换系统把患者不清晰的语音转换成正常语音，一般可分为：1）基于ASR-TTS结构重构，即先通过语音识别系统将障碍语音识别成文本，再通过语音合成系统将文本合成正常语音；2）端到端重构，即使用端到端的神经网络从障碍语音直接生成正常语音，中间没有识别成文本的过程。该技术可用于辅助日常交流、控制通用的语音指令系统（如手机语音指令）、康复训练等目标，是一个新兴的语音交互方向。

**2、目标（调研+建模）**

1) 调研声音转换技术和构音障碍语音重构技术的主要理论、流程及差别

2) 了解声音转换和障碍语音重构性能的评价方式和评价指标，讨论实际应用中还有哪些地方需要评估

3) 实现一个端到端构音障碍语音重构系统（需要自己训练，数据由谢老师提供）及ASR-TTS语音重构系统（ASR和TTS模块可使用商业系统，如WhisperX、GPT-SoVITS等），并比较和讨论两者差别

**3、预期成果**

1) 提交关于障碍语音重构技术的调研报告，字数不限，以内容完整为准

2) 提交实现的障碍语音重构系统代码和实现结果报告

2.1) 可使用开源代码；2.3) 实现结果报告应包括不同指标的评价以及一些转换前后的语音例子2.4) 以工作完成度为主，性能合理即可，不作为衡量标准

**4、参考文献**

[1] Sisman, B., Yamagishi, J., King, S., & Li, H. An overview of voice conversion and its challenges: From statistical modeling to deep learning. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020.

[2] Hawley, M.S., Cunningham, S.P., Green, P.D., Enderby, P., Palmer, R., Sehgal, S. andO'Neill, P. A voice-input voice-output communication aid for people with severespeech impairment. IEEE Transactions on neural systems and rehabilitation engineering,2012.

[3] Wang, D., Yu, J., Wu, X., Liu, S., Sun, L., Liu, X. and Meng, H. End-to-end voice conversion via cross-modal knowledge distillation for dysarthric speech reconstruction. In ICASSP 2020.

### 十四、基于多维神经生理信号的情感识别技术（谢旭荣 微信Karmiu）

**1、背景**

近年来脑机交互发展迅速，人们可以通过神经信号（如脑电信号）及外周神经信号（即生理信号，如心电、皮肤电、肌肉电等）与电子设备进行交互，其中关键技术是基于神经生理信号的模式识别技术。情感（或情绪）识别是一种常见的模式识别及人机交互任务，该任务从用户（或被试）的表情、行为、声音等表现判定用户正处于何种情感状态。神经生理信号可从更深层的神经活动判断用户的情绪，因此一般比表情等外显表现更为准确及客观。

**2、目标（调研+建模）**

1) 了解情感的表示方法有哪些，它们是如何定义的（如VAD方法）

2) 调研常用的多维神经生理信号情感识别数据集有哪些（如AMIGOS），以及这些数据的组成和采集方式

3) 了解上述数据涉及的神经生理信号特征有哪些，以及它们的原理

4) 下载两个多维神经生理信号情感识别数据，并复现至少一个近年性能比较好的发表模型。

5) 对构建模型性能进行评价，并讨论如下问题：a) 构建的模型的实用性如何，为什么；b) 多维神经生理信号情感识别所面临的困难有哪些

**3、预期成果**

1) 提交关于多维神经生理信号情感识别的调研报告，字数不限，以内容完整为准

2) 提交实现的多维神经生理信号情感识别代码和实现结果报告

2.1) 可使用开源代码；2.3) 实现结果报告应包括不同指标的评价；2.4) 以工作完成度为主，性能合理即可，不作为衡量标准

**4、参考文献**

[1] Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., Zhao, Z., Kumar, N. and Marttinen, P. EEG based emotion recognition: A tutorial and review. ACM Computing Surveys, 2022.

[2] Hasnul, M.A., Aziz, N.A.A., Alelyani, S., Mohana, M. and Aziz, A.A. Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors*, 2021.

[3] Miranda-Correa, J.A., Abadi, M.K., Sebe, N. and Patras, I. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE transactions on affective computing*, 2018.

**十五、基于说服策略的文本生成方法（王逸之 电话: 18260866237 邮箱：[yizhi2020@iscas.ac.cn](mailto:yizhi2020@iscas.ac.cn) 微信：Timesten）**

1. **背景**

我们生活在一个海量信息的时代，在互联网上会遇到很多具有说服性语言的文本（一种使用修辞技巧和手段来影响读者的写作风格），例如：吸引人们点击的标题、带有立场的时政新闻、争取选民支持的竞选宣传，甚至是一些不经意嵌入的广告等。

虽然说服性语言在不同领域中具有广泛定义，但说服性语言存在一定共性，例如将心理学的说服框架用于信息传播，通过在文本中强调情绪或威胁诉求，通过得失变化来增加目标用户的情绪唤醒，一个典型的使用该框架的例子如下：“如果你涂防晒霜，当你年老时还会拥有迷人的皮肤” 和 “吸烟会导致致命的肺癌”。

研究表明，在互联网环境下用户倾向于寻求即时满足，注意力集中时间短，更多地会依赖认知捷径来进行信息处理和决策。Robert Cialdini围绕影响决策捷径总结了六个核心原则：互惠、承诺、社会认同、喜爱、权威和稀缺，基于这些原则拓展的说服策略被广泛应用于人机交互系统设计、广告投放、在线捐赠等领域。

随着大模型被广泛用于写作和交流的各个方面，一些研究探索了利用大模型进行个性化文本生成，但在说服领域生成具有说服技巧文本能力还有待探索。

1. **目标**

对现有大模型个性化文本生成技术进行调研，利用经典的说服理论框架、说服策略或说服技巧作为依据，指导大模型在在线论坛或在线对话场景下生成具有说服力的文本。设计模型并通过实验评估大模型的说服性文本生成能力。

1. **指南**
2. 背景、理论基础与相关技术调研
3. 至少选择1个公开数据集进行实验，例如Reddit论坛场景的ChangeMyView、Borrow、RAOP，对话场景的Persuasion For Good或其他自选。
4. 设计、完成实验，并对实验数据进行总结
5. **相关技术背景**

LLM Text Generation, 说服策略识别

1. **预期成果**
2. 生成的说服性文本与数据集中标记说服的文本进行内容相似度 ROUGE-L 评估
3. 人工对大模型生成的说服性文本和数据集的说服性文本进行说服打分，评估大模型说服性文本生成性能。
4. **参考文献**

[1] Annye Braca, Pierpaolo Dondio. 2023. Persuasive communication systems: a machine learning approach to predict the effect of linguistic styles and persuasion techniques. Journal of Systems and Information Technology, ISSN: 1328-7265. <https://www.emerald.com/insight/content/doi/10.1108/JSIT-07-2022-0166/full/html>

[2] Jiaao Chen, Diyi Yang. 2021. Weakly-Supervised Hierarchical Models for Predicting Persuasive Strategies in Good-faith Textual Requests. AAAI 2021.

[3] Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, Zhou Yu. 2019. Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. <https://www.aclweb.org/anthology/P19-1566/>

[4] Amalie Brogaard Pauli, Isabelle Augenstein, Ira Assent, 2024. Measuring and Benchmarking Large Language Models’s Capabilities to Generate Persuasive Language. arXiv.

<https://arxiv.org/pdf/2406.17753>

[5] Haolan Zhan, Yufei Wang, Zhuang Li,.etc. 2024. Let’s Negotiate! A Survey of Negotiation Dialogue Systems. arXiv: 2402.01097v1

[6] Vivian P. Ta1, Ryan L. Boyd, Sarah Seraj. 2022. An inclusive, real‑world investigation of persuasion in language and verbal behavior. Journal of Computational Social Science, (2022), 5:883–903

<https://doi.org/10.1007/s42001-021-00153-5>

**十六、基于图片、视频的孤独症儿童眼动特征研究（联系人：张高媛 电话：18801213768）**

1. **背景**

近年来，孤独症谱系障碍（后简称孤独症）的患病率呈上升趋势。中国流行病学调查显示0-6岁儿童孤独症患病率为1.8%，是造成5岁以下儿童精神残疾的主要原因，具有终身残疾的特点。孤独症儿童在行为、心理、言语等方面存在障碍，社交障碍是核心障碍之一，通常表现为社会性注意缺陷、社交情感缺陷、沟通意识和技巧缺乏等，难以与他人建立正常的人际关系，使其无法融入日常社会生活中。不少研究表明孤独症儿童对于面孔和社交场景信息的加工存在特殊的处理方式，这些特点影响了他们的社交能力的发展。

眼动追踪技术是一种通过测量眼睛的注视点和眼球运动来分析视觉注意力和信息处理的工具。其基本原理是利用红外线或视频摄像设备捕捉眼球的位置变化，通常通过瞳孔和角膜的反射来确定视线方向。眼动追踪的分析指标通常包括注视时间、注视次数、扫视路径和瞳孔变化等，这些指标有助于深入理解视觉注意分配、兴趣偏好及认知过程等。

1. **目标**

通过对孤独症儿童在观看图像或视频时的眼动模式进行分析，了解他们的视觉加工特征。其中包括孤独症儿童视觉特征调研、对现有的图片和视频的眼动追踪技术进行调研、掌握眼动技术的分析方法、对数据进行统计分析并讨论孤独症儿童的视觉加工特征等。

1. **指南**

（1）背景、理论基础与相关技术调研

（2）整合现有技术，掌握基于图片和视频的眼动追踪技术分析方法

（3）对实验数据进行统计分析，并讨论孤独症儿童的视觉加工特征

1. **相关技术背景**

眼动追踪技术、数据建模与分析、视觉加工机制

1. **预期成果**

调研报告、眼动追踪技术分析方法以及孤独症儿童视觉加工特征分析

1. **参考文献**

[1] Griffin, J. W., Bauer, R., & Scherf, K. S. (2021). A quantitative meta-analysis of face recognition deficits in autism: 40 years of research. *Psychological Bulletin*, *147*(3), 268.

[2] Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, *42*, 279-297.