综合AI技术（CV+NLP）在会议室场景中的部署与应用

智能会议助理项目（暂定名）计划书

一、智能会议项目研究目标与成果预期

1、项目背景介绍

结合CAD实验室对AI与大数据的科研实力，团队科研技术水平在算法、模型、数据等方面优势明显。为了给实验室研究生同学以及研究教师提供更真实的实验场景与研发动力，同时在科研进展中进行大量日常数据的积累，项目组在考察了多个场景后，认为在我们身边大量发生的是学术交流。这种学术交流在企业中也通过会议的形式正在广泛使用。如果我们能够将最新的科研成果与技术，应用在这样的高频使用场景中，不仅可以增加实验室人员对技术、产品、体验的关注，也为科研的方向提供了独特的场景和落地的环境。

同时，学术交流与各类会议的场景是具备相当的复杂度的。智能化过程会涉及到自然语言处理、图像处理、声音处理、手势识别、人体跟踪等多种技术与应用。众多最新的算法与训练的模型可以在这样复杂的环境中验证效果，也有机会让研究人员更关注算法实际效果与实现的性能，为实验室带来“实战”精神。是一种产学研结合的实践尝试。当今技术发展，需要考虑算法与终端、与用户体验真正的结合，才能有生命力。我们的同学们也需要一个真实的环境，了解终端硬件对算法的影响，了解使用场景对模型的重要性。

在进行智能技术实践的过程中，我们也希望增加了一种体验式的实验室宣传方式。通过对实验室同学们的技术分享智能的生成分享的视频，并在分享的视频中标注“由智能技术记录”和“北京邮电大学CAD实验室”等文本，非常有利于扩大实验室的影响力，学术性的分享为广大学生所喜爱。视频的传播也有机会帮助同学们扩展其在学术界和产业界的影响。

综合上述原因，项目筹备组拟定该智能会议助理项目计划书，以期为实验室打造一个“综合AI技术（CV+NLP）在会议室场景中的部署与应用”的实战环境，激发实验室科技实践与创新热情，训练研究生对实际应用与算法实际性能的关注，扩展实验室与研究生同学的综合影响力。也期望有一天智能会议助理可以成为明星应用，吸引全校与业界的关注，扩大战果。

2、项目整体目标与成果预期。

整体项目目标为：用2年左右的时间，打造完整用户体验的智能会议助理功能。对“完整”的定义，包括智能技术在会议流程上的完整参与（会议前准备、会议中主持、会议后纪要等）；也包括在从语音到图像， 到实时动态视觉处理等完整的AI应用技术链；也包括在从硬件终端到主机，再到用户web端功能完整的交付与数据处理，最终实现用户的无感知体验提升，最大化智能会议助理的交互流畅性与通用功能的智能自动性。

在完成上述目标的同时，在项目落地过程中，预期可以产生多套创新应用与系统，包括不限于智能会议语音机器人主持系统、智能课件生成系统、智能会议纪要系统、智能会议签到系统、智能会议当前发言人识别与显示系统等。预期可以在CV、NLP等智能领域，产生大于10篇的技术论文与研究资料。预期实现实验室曝光度超过万人次，以及其他扩大实验室影响力的成果。

二、智能会议项目实施计划

1、项目阶段设置

项目拟分为四个阶段进行实施：a）原型验证阶段、b）项目一阶段、c）项目二阶段、d）项目收尾与完善阶段。

2、作为起步的原型验证阶段：实现以会议内容分享为目的的学术会议讲解过程的自动记录与自动合成功能。具体而言，拟定采用1.5至2个月的时间，在2019年8月31日前，实现以下功能：

i、学术分享活动中，演讲者的人像追踪与视频记录；

ii、学术分享活动中，演讲者的桌面录屏与鼠标操作视频记录；

iii、自动生成演讲者半身像记录视频；

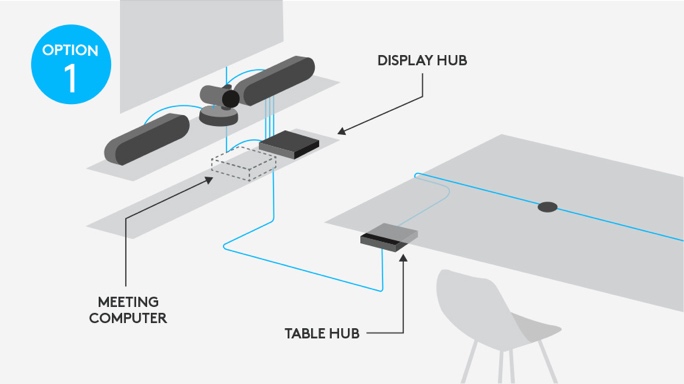
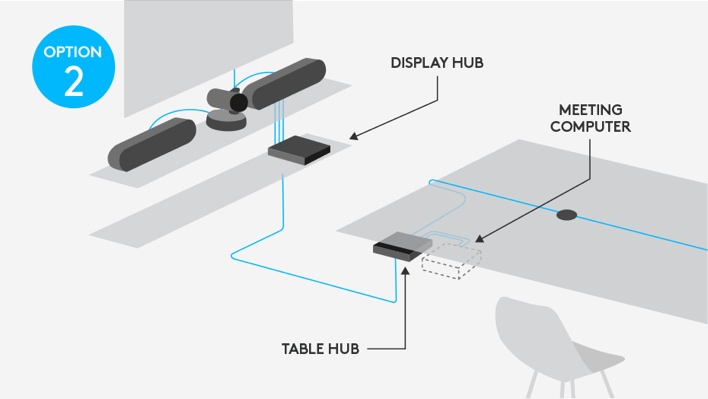
iv、自动生成由演讲者半身像、桌面录屏、演讲题目标识、实验室名称与网站链接等信息组成的学术分享视频。

所需人员：目前已有多名研究生自愿加入，并进行了初步分工。

所需资源：项目拟在新科研楼821房间试用，需要授权在房间内顶部挂载带有云台功能的高清摄像头一个，侧部挂载带有云台功能的高清摄像头二个，移动式带有云台功能与麦克风阵列的高清摄像头一个（分别用于演讲者跟随、参会者人脸识别、参会者姿态捕获等功能）。放置一台用于音视频处理的高性能台式机。并需要采购上述设备。在AI模型与数据计算方面，需要使用实验室相关GPU与TPU计算资源用于模型训练与优化。

设备选型：高清摄像头硬件方面，选择罗技Rally系列超高清会议摄像头系统。该系统支持较大面积的会议室，有多种固定方案，配件齐全，可以配置麦克风阵列，也可以配置每个桌位的麦克风板，并有简洁的遥控器，符合工业设计、美观时尚、硬件系统性能指标高等要求。同时，罗技的设备对于网络会议体系Zoom等也进行了良好的适配，可以为会议室提供Zoom视频会议、网络直播等功能的基础设施，极大的方便了国际国内视频会议的举办。（台式机性能需求较为标准，此处暂略去选型。）

硬件预算：Logitech Rally系统套件价格在2500美元，增加单独的3个摄像头，每个约2000美元。合计四套摄像头总价在8500美元以内（约合RMB6万）。另需一台高性能台式机，预算RMB3万。硬件预算合计9万元人民币。系统组图与Rally产品如下图所示。（底部两个图双击可以播放动图演示。）

3、项目一阶段：实现以会议组织为目的的学术会议签到、流程提示、发言人聚焦以及智能屏显等智能会议助理功能。具体而言，拟定采用6至9个月的时间，在2020年5月31日前，实现以下功能：

i、学术分享活动中，对进入会议室的人员进行自动签到；

ii、学术分享活动中，对发言人进行实时聚焦，在大屏上通过子窗口形式，动态显示当前发言人头像、职衔、姓名等提示信息；

iii、学术分享活动中，通过预先录入的会议流程和时间要求，智能进行会议的流程的串导，使用自动语音与现场进行交互，实现会议时间提示、会议流程执行等智能会议助理功能；

iv、自动生成发言人的文字记录，并自动形成完整版的会议记录。

所需人员：以原型阶段团队为核心，扩展相应技能同学或工程师。

所需资源：项目拟在新科研楼821房间试用，需要授权使用。其余部分待定，暂时使用原型阶段硬件即可。

设备选型：待定，截止确定日期2019年9月30日。

硬件预算：待定，截止确定日期2019年9月30日。

4、项目二阶段：实现整体会议管理系统的主要功能，比如会议组织、会前通知、会场管理、会后总结等，拟定采用6至9个月的时间，在2021年春节前，实现以下功能：

i、会议过程的相对完整的流程、内容、组织形式等后台管理系统；

ii、会议中数据（语音、视频、图像、文本）管理功能；

iii、会议开放性内容的互联网分发与推广系统；

iv、提供会议内容检索与查询功能，以及其他未尽述之功能。

所需人员：以原型阶段团队为核心，扩展相应技能同学或工程师。

所需资源：待定，暂时使用原型阶段硬件即可。

设备选型：待定，截止确定日期2020年6月30日。

硬件预算：待定，截止确定日期2020年6月30日。

5、项目收尾与完善阶段：本阶段为机动阶段，对于通过前三个阶段的实施未能完善的功能，通过本阶段进行完善和调整，提升整体系统的交互体验与功能完善性、系统稳定性。具体而言，拟定采用6个月左右的时间，在2021年8月31日前，实现以下功能：

i、完善系统整体稳定性，提升交互体验与功能完善。

所需人员：以原型阶段团队为核心，扩展相应技能同学或工程师。

所需资源：待定，暂时使用原型阶段硬件即可。

设备选型：待定，截止确定日期2021年3月31日。

硬件预算：待定，截止确定日期2021年3月31日。

三、智能会议项目责任人：AMC Team、李卓桓