DTS311TC FINAL YEAR PROJECT

Player-Aware Intelligent Monitoring and Operations Navigator

Proposal Report

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Engineering

Student Name	:	Taimingwang Liu
Student ID	:	2037690
Supervisor	:	Xihan Bian

School of AI and Advanced Computing
Xi'an Jiaotong-Liverpool University

November 2025

Abstract

Apply the font of Times New Roman to the paragraphs of the abstract using font size of 12. An abstract is usually one to three paragraphs long with a length of 150 to 350 words.

Contents

1 Introduction

1.1 Problem Setting & Motivation

面向玩家的伴随式（*companion-style*）实时助手，目标是在不打断游戏流程的前提下，持续提供事件提示（*event spotting*）、策略建议（*tactical guidance*）与语音回路（*voice loop*）的低时延体验。近年的*LLM*×游戏与*GUI* 智能体研究快速推进，但落地层面仍存在两类缺口：一是动作接口不统一，大量工程依赖场景定制；二是评测协议不一致，导致复现实验与横向对比困难。围绕这些缺口，已有工作正从统一协议/消融与脚手架/污染控制两侧收敛可复现方法学，为产品化路径提供操作依据[1]–[3]。同时，近期综述将LLM 游戏智能体抽象为记忆—推理—*I/O*三件套，并强调I/O 质量与允许动作集合的一致性是交互成败的关键，这与"结构化输出/约束解码"的工程路径相吻合[4]。

1.2 Scope & Working Definitions

本项目选择以**GUI**（**GCC**）为统一动作接口（*screen-in, keyboard/mouse-out* 的人类同态通道），并以结构化输出（**structured output**）**+** 约束解码（**constrained decoding**）作为默认的动作生成与落地路径。已有研究表明：在不依赖应用专用*API* 的前提下，通过"规划—技能整理（macro/skill）—自反思—记忆"的管线可跑通长链路任务；在真实平台上，"截图 → 结构化动作"的端到端导航具有可复现性，且合法动作映射（*legal move constraint*）能够显著降低无效动作（invalid actions）[5]–[7]。内部组织采用**MCP-style** 编排（模块/技能/工具的注册与路由），支撑消融与替换；评测遵循统一协议**+**脚手架 的做法，并引入机会导向的OAS/RT/APO 三项工作定义以刻画伴随式体验[1], [2]。为避免名词漂移，本文将记忆**—**推理**—I/O**作为术语锚点：*planning*+*reflection* 对齐*reasoning*，*skills/macros* 属于*I/O* 侧的动作产出形态，*memory* 独立；输出端默认遵循"语义 → 允许动作"映射或在允许集合上建模概率的合规策略[4]。

1.3 Design Principles & System Preview

Design principles. 本文遵循四项原则：结构化输出（降低无效/便于审计）、协议一致（可复现/可消融）、低耦合编排（MCP-style，便于插拔skills/tools）、低时延（面向voice loop 与事件提示）。

System preview. 系统流为：*screen/audio* → 轻量*VLM* 感知→ *agentic*（planning/memory/reflection）→ *MCP-style* 技能/工具路由（含OCR/检索/计算等*tool use*）→ *GUI* 执行（kb/mouse）→ *safety*（确认/回滚/急停）。为降低端到端时延，部署层面将结合工具增强型*MLLM* 的分担思路与端侧推理的量化/缓存策略作为工程抓手[8], [9]。

1.4 Key Challenges

落到真实玩家场景，难点不是单一模型分数，而是稳、准、快、可控的整体体验。下面按可测的挑战项列出。

长链路稳定（**long-horizon stability, GCC**）。只走GUI（GCC）时，误点与偏移会沿交互链放大，导致“走着走着就跑偏”。可用*pass@k*、回滚率（rollback）与**APO**（attempts per opportunity）来量；“*planning* + *skills*（*macro*）+ *reflection* + *memory*”的组合能缓解漂移，但并非万灵药[5]。

视觉定位与记忆（**vision-centric grounding & memory**）。在仅视觉/连续空间设定下，定位/追踪/计数、时机控制、长期视觉记忆是当前模型短板，可按**OAS**（opportunity-normalized success）分机会类型统计，如“可拾取物/时间点/路径节点”等[10]。

无效动作与幻觉（**invalid actions & think–action mismatch**）。自由文本到动作容易“想得对、点错位”。结构化输出+ 合法动作映射能够压低*Invalid%*，并可用*Brier/MAE* 看校准；训练侧以“格式/类型/坐标/内容”四粒度奖励对齐执行细节[6], [7]。在实现层面，还可采用“语义 → 允许动作”映射或在允许动作集合上建模概率的输出策略，从源头抑制越界与错位[4]。

OOD 与协议一致（**OOD & protocol consistency**）。环境一换、版本一更，成绩就难对比。需要过程生成（procedural generation）做分布外（OOD）评测，固定*post-processing* 与提示脚手架（scaffold）控变量，并在统一协议与*leaderboard/battle arena* 下报告*macro/micro* 指标[1], [2], [7]。

时延与交互体验（**latency & UX**）。伴随式助手要“当下就回应”。关键是**RT**（reaction time per opportunity）与*voice RTT*（语音往返）。工程上依赖量化/剪枝/*KV* 缓存/流式解码与端侧/端云协同压时延，搭配单航班（single-flight）与可打断（barge-in）策略[9]。

安全与健壮性（**safety & robustness**）。高风险动作必须“可确认、可回退、可追溯”。做法包括权限白名单、双确认、影子执行（*shadow execution*）与回滚/急停，并保存日志/审计定位*think–action mismatch*[11]–[13].

1.5 Project Objectives & Expected Deliverables

Objectives. (i) 实现一个以**GCC** 为主的伴随式实时助手原型，覆盖事件提示、策略建议与语音回路；(ii) 采用**MCP-style** 编排组织*skills/macros*、*planning*、*memory*、*reflection*，在不依赖应用专用API 的前提下实现可插拔与可审计；(iii) 明确一套小而可复现的评测要素（任务脚本与指标族），关注*pass@k/TTC/Invalid%/macro–micro* 以及*OAS/RT/APO* 等体验相关量[1], [2].

Expected deliverables. (a) 系统原型：屏幕采集与轻量感知、agentic 模块、MCP-style 技能总线、GUI 执行器与基础安全护栏；(b) 评测脚本与配置：可复现实验的任务脚本、指标计算与日志审计工具（含模块开关用于对照）；(c) 使用文档与演示：安装/运行说明、配置模板与演示视频。

1.6 Assumptions & Out-of-Scope

Working assumptions. 默认采用"*single-flight* + *event-triggered* + *frame-window*（*3–5* 帧）+ *text-first*"的工程姿态以降低时延与方差；评测记录*post-processing* 与环境版本以保持协议一致。

Out-of-scope. 不开展逐个应用/游戏的专用**API** 适配；不在本报告中承诺大规模端到端训练与数据采集；不依赖平台级增强权限（如A11y/私有DOM 钩子）；**VLA** 直出动作仅作为评测对照而非默认路径[7].

1.7 能力边界（Capability Gaps）

以视觉为中心的自由形式游戏显示，多模态模型在若干关键能力上仍与人类存在差距：定位/追踪/计数、历史依赖与锚定偏置、时机控制、视觉记忆、文本识别与空间理解以及高层时序推理等维度普遍薄弱。采用Elo 风格相对强度排名与"模型/策略"的管线分离，从评测组织上揭示了这些系统性短板与不稳定性[10]。更通用的LLM×游戏评测同样显示，交互稳定性与污染控制会显著影响结果分离度，提示伴随式场景需要机会归一化与反应时等细粒度度量以及脚手架约束[2]。

1.8 术语与范围对齐（Glossary & Scope Alignment）

GCC（General Computer Control）：*screen-in* + *keyboard/mouse-out* 的人类同态动作接口；本文默认执行通道[11]. **LAM（Large Action Models）**：以结构化动作为一等产出的模型族；本文作为对照范式引用[11]. **VLM vs VLA**：文本/*JSON* 输出经映射进入动作空间vs 直接动作向量/分布；评测统一采用合法动作映射+ 约束解码口径[12]. **Scaffold vs Orchestration（MCP-style）**：前者为评测期稳定交互"脚手架"，后者为模块/工具注册与路由；二者互补[13]. 指标口径：*pass@k*、*TTC*、*Invalid%*、*macro/micro* 并报；机会导向的**OAS/RT/APO** 为伴随式场景核心补充[14]. 记忆—推理—I/O（**memory–reasoning–I/O**）：本文的内部工作划分与术语锚点；*planning*+*reflection*→*reasoning*，*skills/macros*→*I/O*，*memory* 独立；输出端遵循"语义 → 允许动作"或在允许集合上建模概率的合规策略[4].

2 Literature Review

2.1 接口可行性（Interface Feasibility: GUI/GCC）

要把“伴随式实时助手”落到真实游戏，首先要回答：不接专用**API**、仅用**GUI**（**GCC, screen-in & keyboard/mouse-out**）是否可行？代表性工作证明，在不依赖应用专用接口的条件下，通过“信息采集—规划—技能整理（macro/skill）—自反思—记忆”的管线，可以在桌面/游戏场景跑通长链路任务，从而确立了人类同态接口（*human-homomorphic interface*）的可迁移性与可复现性[5]。

进一步地，端到端实践展示了“截图 → 结构化动作（*structured output*）”的可落地路径：以受约束的动作模式替代自由文本，使执行更稳且便于审计（auditability）；同时，评测框架通过感知/记忆脚手架（scaffolds）稳定交互回路，把“游戏→评测”流程工程化[2], [6]。对比研究还提出合法动作映射与约束解码（**constrained decoding**）以显著降低无效动作（invalid actions），为后文的评测协议埋下方法学钩子[7]。既然“能做”已被验证，接下来就需要统一“怎么比”。

2.2 评测协议与实用要素（Protocols & Practicalities）

在统一协议方面，已有研究通过*MCP-style* 编排解耦代理与多游戏环境，统一配置与日志，在同一协议下做可复现的消融以比较*planning / reflection / memory / skills* 等模块，并配套*leaderboard* 与*battle arena* 维持跨任务可比性[1]。为减少提示方差与避免训练—测试污染，评测侧引入Gym-style 接口与感知/记忆脚手架（harness），记录*post-processing* 并固定环境版本，报告在多模型下具有良好分离度的结果[2]。

在“无效动作治理”上，动作显式化是共同结论：将输出映射到合法离散动作空间并采用结构化输出/约束解码联合抑制*Invalid%* 与坐标偏差；评估层面同时使用*Micro/Macro Precision/Recall/F1* 与*Brier/MAE* 兼顾分类与校准[7]。与此一致，近期综述提出两条工程路径：其一将自由文本语义映射到最近的允许动作；其二直接在允许动作集合上建模概率，两者均可与本文“合法动作映射/约束解码”的实现口径对齐[4]。面向结构化动作的端到端导航，还可在RL-finetune 中将奖励拆分为格式/动作类型/坐标/内容四粒度，将“合规性与细粒度正确性”纳入可学习信号，形成工程可落地的闭环[6]。

为了与“伴随式实时体验”对齐，本文在客观指标（*success/pass@k*、*time-to-completion, TTC*、*misclick/rollback*、*latency*——含*voice RTT*）与*macro/micro* 并报之外，引入三项机会导向工作定义：**OAS**（*opportunity-normalized success*，成功次数÷可操作机会数）、**RT**（*reaction time per opportunity*，机会出现到首个有效动作/提示的时间）与**APO**（*attempts per opportunity*，每次机会的平均尝试/回滚次数）[1], [2], [6], [7]。

2.3 Agentic 模块（Planning / Memory / Reflection / Skills）

围绕长链路稳定性，研究通常将代理拆解为可组合的四要素：*planning*（任务分解与策略选择）、*memory*（短长时与用户偏好）、*self-reflection*（纠错与风格一致）与*skills/macros*（原子到复合技能）。该组合有助于缓解错误累积与状态漂移，支撑*GCC* 通道上的可迁移闭环[5]；在统一协议下，可复现的模块消融提供了边际效应与搭配选择的证据，使“用/不用、强/弱”不再停留于经验判断[1]。训练与数据侧通常与结构化动作相匹配：通过*RL-finetune* 与反馈式调优（如*GRPO/RFT*），并以“格式/类型/坐标/内容”四粒度奖励，收紧输出空间同时提升细粒度正确性[6]。作为收敛视角，近期综述将上述组件归并为记忆—推理—*I/O*三件套：*planning+reflection* 对齐*reasoning*，*skills/macros* 属于*I/O* 的动作产出形态，*memory* 独立，便于在实现与评测时统一术语与接口[4]。

训练与数据侧的配套通常与结构化动作相匹配：通过*RL-finetune* 与反馈式调优（如*GRPO/RFT*），并以“格式/类型/坐标/内容”四粒度奖励，收紧输出空间同时提升细粒度正确性，从而在“反思—记忆—技能”的外圈外，再加一层可学习的约束[6]。此外，接口位形也影响模块产物：*VLM* 倾向输出文本/JSON，经由映射进入动作空间；*VLA* 则直接产出动作向量/分布，二者在无效率、校准与OOD 行为上的取舍需要在同一协议中对照评测[7]。要让这些模块更稳/更强，下一步问题就是：如何学习与调优。

2.4 学习范式（Learning Paradigms）

指令化与条件化提供了“看懂—再行动”的可解释路径：*R2-Play* 将多模态游戏指令（*multimodal game instructions, MGI*）并入*Decision Transformer*，用“游戏描述—轨迹—操作引导（含关键元素位置）”三段式模板共享跨任务知识，在多任务与泛化上报告相对优势，可作为提示/指导结构的模板参考[15]。对结构化动作任务而言，*RL-finetune* 配合四粒度奖励把“合法性+定位/文本输入”压进可学习信号，直接对齐 2.2 节中的“无效动作治理”目标[6]。

为在资源与稳态间取舍，工具增强型*MLLM* 的综述总结了外部工具（OCR/检索/计算/专家模型等）在*tool use / MCP-style* 编排下对延迟与鲁棒性的影响与边界，提示通过检索与轻量工具链分担大模型负载；而接口对照研究将*VLM*+映射 与*VLA*直出动作并置，指出两类路径在无效率、置信校准与OOD 表现上的差异，需要结合具体协议与数据分布权衡[7], [8]。即便“能做—能比—能学”已形成闭环，模型仍有系统性短板，必须在评测中直面。

3 Project Plan

3.1 Proposed Solution / Methodology

3.2 Experimental Design

3.3 Expected Results

3.4 Progress Analysis and Gantt Chart

3.4.1 Risk & Ethics

4 Conclusion

References

[1] D. Park *et al.*, "Orak: A foundational benchmark for training and evaluating llm agents on diverse video games," 2025, arXiv:2506.03610. arXiv: 2506.03610.

[2] L. Hu *et al.*, "Lmgame-bench: How good are llms at playing games?, 2025a," *URL https://arxiv. org/abs/2505.15146*,

[3] X. Xu *et al.*, "A survey on game playing agents and large models: Methods, applications, and challenges. arxiv pre-print," *arXiv preprint arXiv:2403.10249*, 2024.

[4] S. Hu *et al.*, "A survey on large language model-based game agents," *arXiv preprint arXiv:2404.02039*, 2024.

[5] W. Tan *et al.*, "Cradle: Empowering foundation agents towards general computer control," *arXiv preprint arXiv:2403.03186*, 2024.

[6] Z. Gu *et al.*, "Ui-venus technical report: Building high-performance ui agents with rft," *arXiv preprint arXiv:2508.10833*, 2025.

[7] P. Guruprasad, Y. Wang, S. Chowdhury, H. Sikka, and P. P. Liang, "Benchmarking vision, language, & action models in procedurally generated, open ended action environments," *arXiv preprint arXiv:2505.05540*, 2025.

[8] W. An, J. Nie, Y. Wu, F. Tian, S. Lu, and Q. Zheng, "Empowering multimodal llms with external tools: A comprehensive survey," *arXiv preprint arXiv:2508.10955*, 2025.

[9] J. Xu *et al.*, "On-device language models: A comprehensive review," *arXiv preprint arXiv:2409.00088*, 2024.

[10] X. Zheng *et al.*, "V-mage: A game evaluation framework for assessing vision-centric capabilities in multimodal large language models," *arXiv preprint arXiv:2504.06148*, 2025.

[11] C. Zhang *et al.*, "Large language model-brained gui agents: A survey," *arXiv preprint arXiv:2411.18279*, 2024.

[12] F. Tang *et al.*, "A survey on (m) llm-based gui agents," *arXiv preprint arXiv:2504.13865*, 2025.

[13] X. Hu *et al.*, *Os agents: A survey on mllm-based agents for computer, phone and browser use*, 2024.

[14] Z. Durante *et al.*, "Agent ai: Surveying the horizons of multimodal interaction," *arXiv preprint arXiv:2401.03568*, 2024.

[15] Y. Jin *et al.*, "Read to play (r2-play): Decision transformer with multimodal game instruction," *arXiv preprint arXiv:2402.04154*, 2024.

Appendix A. Title of Appendix A

A.1 Appendix Heading 1

Text of the appendix goes here

A.2 Appendix Heading 2

Text of the appendix goes here

A.3 Appendix Table and Figure Captions

In appendices, table and figure caption labels and numbers are typed in manually (e.g., Table A1, Table A2, etc.). These do not get generated into the lists that appear after the Table of Contents.

Appendix B. Title of Appendix B

Text of the appendix goes here if there is only a single heading.