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

> 说明：交代问题空间（游戏伴随式助手，companion-style game agent）、范围（scope）、研究缺口（gap）、本文立场与贡献（positioning & contributions）。涉及了2.1-2.4和2.6-2.7。

1.1 Problem Setting & Motivation

TODO: 背景：实时游戏场景（real-time gaming），玩家需要事件提示（event spotting）、策略建议（tactical guidance）、低延迟语音交互（voice loop）。动机：现有VLM/VLA在桌面/游戏的落地与稳定性存在鸿沟。

1.2 Scope & Working Definitions

TODO: 给出工作定义（working definitions）：多模态（multimodal）、动作接口（action interface: GUI-only vs API/MCP），伴随式助手（companion-style），短时托管（autopilot/macros），评测术语（success rate, latency, advice adoption）。

1.3 Key Challenges

TODO: 长链路稳定性（long-horizon stability）、UI变化鲁棒（robustness）、延迟预算（latency budget）、权限安全（permissions/rollback）、跨游戏迁移（generality/portability）。

1.4 Our Positioning & Contributions

TODO: 工程立场：GUI-first＋机会主义API/MCP；引入skills/macros、planning/memory/reflection 语音链路；提出面向伴随式助手的评测协议（advice adoption, voice RTT, macro success）。可列1-3条要点。

1.5 Design Principles & System Preview

TODO: 一句话系统图预告：screen/audio→VLM→LLM/agentic modules→(GUI kb/mouse | API/MCP)→safety guard（permissions, rollback, kill-switch）。把详图留到方法章节。

1.6 Summary of Findings (Optional)

TODO: 一句话总结文献趋势：从GUI-only通用性到API/MCP确定性，从对话式感知到任务化（taskification）与技能化（skills）。可选，若版面紧张可删。

2 Literature Review

> 说明：采用taxonomy组织而非按时间；每小节末给1-2句“与本文关系”。Cradle放到2.2（GUI-only/GCC）而非开头。涉及了2.1-2.9。

2.1 Perception: Modalities & Grounding

> 说明：输入模态与定位。涉及2.3-2.4。**TODO:** 视觉为主（screen/video）+可选音频（audio）；VLM能力：检测/描述/grounding；提一嘴VLA（如果动作权重内生）与传统VLM+tool的差别。与本文：我们选轻量VLM，优先本地（on-device）与流式ASR/TTS。

2.2 Action Interfaces: GUI-only (GCC) vs API/MCP

> 说明：动作接口对比的核心小节，放Cradle。涉及2.6-2.7。**TODO:** 定义GCC：screen-in, keyboard/mouse-out。代表作：**Cradle**（GUI-only，skill curation/registry, reflection/memory）。优点：通用性（generality）、可迁移性（portability）；缺点：确定性/延迟。API/MCP：确定性高、速度快、但依赖适配。本文策略：GUI-first + API/MCP作为加速通道，GUI兜底。附一句对比表指引。

2.3 Agentic Modules: Planning, Memory, Reflection, Skills

> 说明：机制视角。涉及2.1, 2.2, 2.4。**TODO:** 规划（planning）、记忆（memory, 用户偏好/历史）、反思（self-reflection, 纠错/风格一致）、技能库（skills/macros, 原子→复合）。说明这些机制如何提升长链路成功率与体验一致性。与本文：直接采纳skills+reflection+memory组合。

2.4 Learning Paradigms: Zero-shot, RAG, Finetune, IL/RL, Distillation

> 说明：训练与推理范式。涉及2.1, 2.3。**TODO:** 列常见范式及成本/收益：零样本与提示工程、检索增强（RAG for UI schema/FAQ）、轻量微调（LoRA）、模仿/强化（IL/RL）、蒸馏到小模型。与本文：优先零样本+RAG，必要时小规模LoRA以稳UI。

2.5 Benchmarks & Datasets (OS-like, Games, Desktop)

> 说明：基准版图。涉及2.6。**TODO:** 按类型分：桌面/操作系统类（如OSWorld系）、游戏/模拟器类（如ALE等）与自建任务脚本。指出覆盖能力与缺口：缺少“伴随式建议/语音互动”的评测。与本文：定义我们的小型、可复现实验设置与演示脚本。

2.6 Evaluation Protocols & Metrics

> 说明：强烈关联本文贡献。涉及2.2, 2.7。**TODO:** 客观：success rate, time-to-completion, no-misclick/rollback rate, latency（voice RTT, frame→hint时间）；主观：advice adoption, user satisfaction。与本文：将新增advice adoption与macro success作核心指标。

2.7 Deployment & Real-time Considerations

> 说明：工程现实。涉及2.2, 2.6。**TODO:** 本地/云混合、量化（INT4/FP8）、流式解码、语音中断（barge-in）、资源占用与帧率影响。与本文：给出延迟预算（如≤500ms提示、≤ 1.5s语音回路）。

2.8 Safety, Permissions & Robustness

> 说明：安全边界。涉及2.2。**TODO:** 权限模型（whitelist, scope）、操作确认、影子模式（shadow mode）先预测后执行、回滚/急停。与本文：作为系统必要模块。

2.9 Synthesis: Trends, Gaps & Our Niche

> 说明：综述收束到本文位置。关联全篇。**TODO:** 趋势：GUI-only通用→API/MCP混合确定性；机制：从对话到任务化/技能化；缺口：缺少“伴随式建议+语音”的统一评测与低延迟实现。本文niche：针对实时游戏的companion-style助手，提供可复现小型协议与演示。

3 Project Plan

3.1 3.1 Proposed Solution / Methodology

This section contains the methodology, technical design for the project.

3.2 Experimental Design

This section contains the methodology, technical and experimental design for the project.

3.3 Expected Results

This section contains the expected results.

3.4 Progress Analysis and Gantt Chart

This section contains the progress analysis and Gantt chart.

4 Conclusion

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.