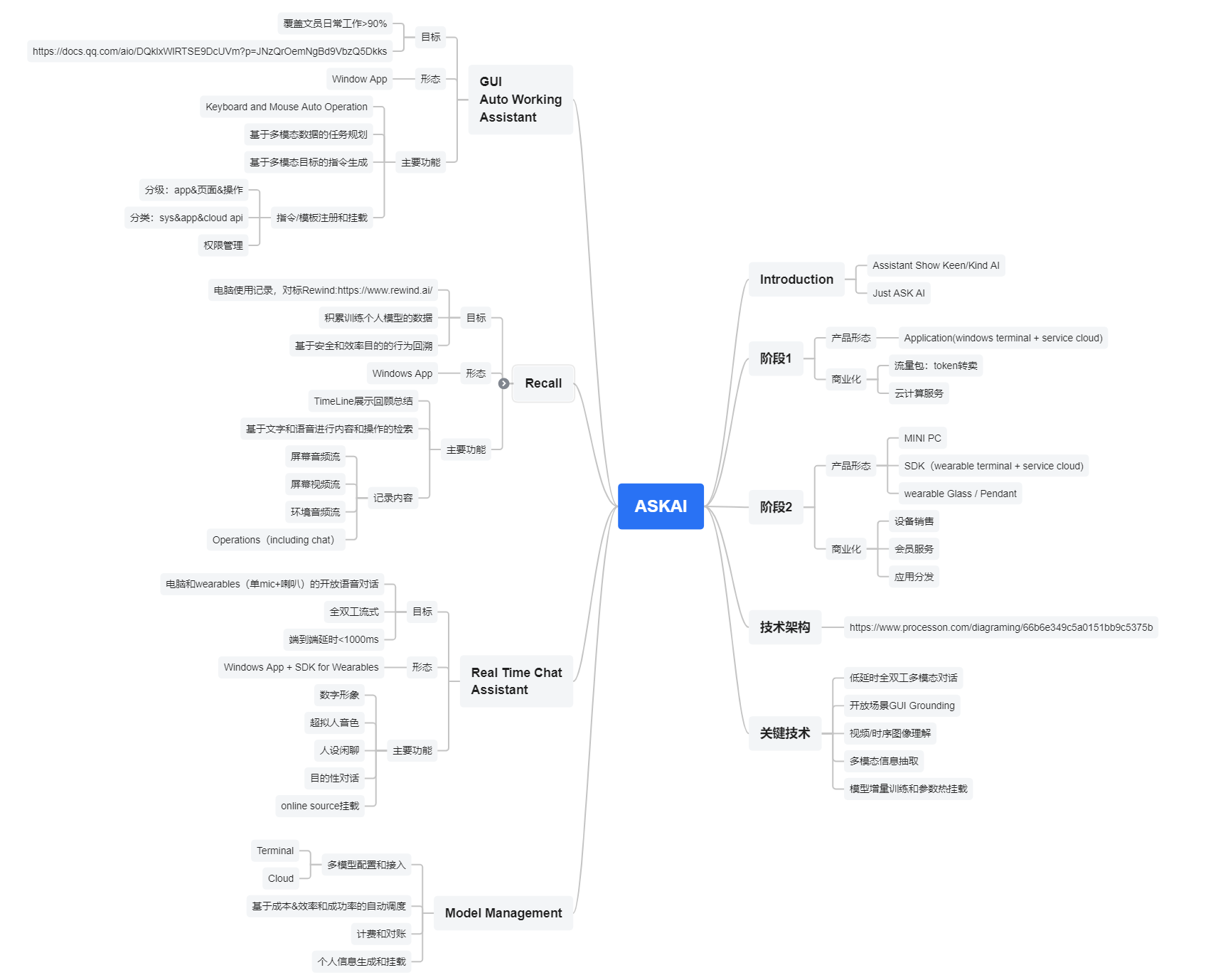
**# 修改记录**

|  |  |  |
| --- | --- | --- |
| 时间 | 修改人 | 修改内容 |
| 2024/8/12 | 李开 | 文档v1.0，创建 |
|  |  |  |

**# Guidline**



**# Instroduction**

1. 目标

Personal Assistant Shows Kind AI

Daily work，Just Ask ASKAI

1. 人群：
2. 一二线城市办公室白领，年龄25~50，月薪>20K
3. 由职业者，程序员&设计师等职位，年龄30~50，平均月收入>20K
4. 场景：
5. 远程非实时沟通协调
6. 非专业文案撰写
7. 可以在PC上完成的重复&事务型工作
8. 实时非专业信息获取和总结
9. 形态：
10. 阶段1：

Assistant：Windows Application

Caller：Phone Application

1. 阶段2：

Assistant：MINI PC（Assistant App inside）

Caller：

- Wareable hardware（Glasses / Pendant）

- SDK：Wait for intergrating

**# Recall**

1. Record

以时间为轴，记录如下内容：

|  |  |  |  |
| --- | --- | --- | --- |
| 内容 | 获取方式 | 存储方式 | 存储逻辑 |
| 屏幕视频流 | - 屏幕录制 | - Local File、  - 格式和压缩方式待定 | - 默认开启记录，可设置  - 默认屏幕分辨率，可设置  - 默认每段1h，可设置  - 默认总记录72h，可设置缓存时间和缓存存储，超期/超量则直接丢掉最前 |
| 关键帧 | - 按时序以n frame/s对视频帧进行拆帧，其中n可设置  - 对拆帧后图像做差分，结果<阈值则丢掉，结果>阈值则获取该帧为关键帧  - 对关键帧进行内容提取：OCR和多模态理解 | - 关键帧图像：Local File  - 关键帧的内容：DB存储原始数据和向量化后的tensor | - 按时序处理和存储关键帧  - 按顺序，以较低优先级（涉及模型调度）对关键帧进行OCR和多模态理解  - 处理后存储（id, image\_URL, timestamp, OCR\_RS, summary\_RS）  - 处理后对图像进行压缩处理（考虑直接变成tensor存储）  - 多模态理解summary\_RS:  (summary, apps)  - 默认一直记录，存储空间满后则按时间顺序，对最早期数据进行丢弃处理 |
| 设备操作 | - 鼠标/键盘/触控等外界设备行为进行录制 | DB | - 默认开启记录，可设置  - (id, timestamp, action\_type, action\_id, action\_description, focus(app,page), \*\*kwargs) |
| API Call | - 获取非系统级的System API调用 | DB | - 默认开启记录，可设置  - (id, timestamp, action\_type, action\_id, action\_description, \*\*kwargs) |
| 屏幕音频流 | - 抓取输入到播放设备的音频流数据  - 对音频流数据进行分类，将其中的dialog分类进行ASR获取Text | - 音频流：Local File  - 音频Text：DB | - 默认开启记录，可设置  - 默认采样率为16000，可设置  - 默认每段最长1h，可设置  - 连续10s无音频流时停止录制，直至有新的音频流开启，并重新启动新文件进行存储  - (id, timestamp, type, text) |
| 环境音频流 | - 通过mic抓取到环境音频流数据  - 对音频流数据进行分类，将其中的对话分类进行ASR获取Text | - 音频流：Local File  - Type and Text：DB | - 默认不开启，可设置  - 默认采样率未16000，可设置  - 默认每段最长1h，可设置  - 连续10s无环境音停止录制，直至有新的音频流开启，并重新启动新文件进行存储  - (id, timestamp, type, text) |

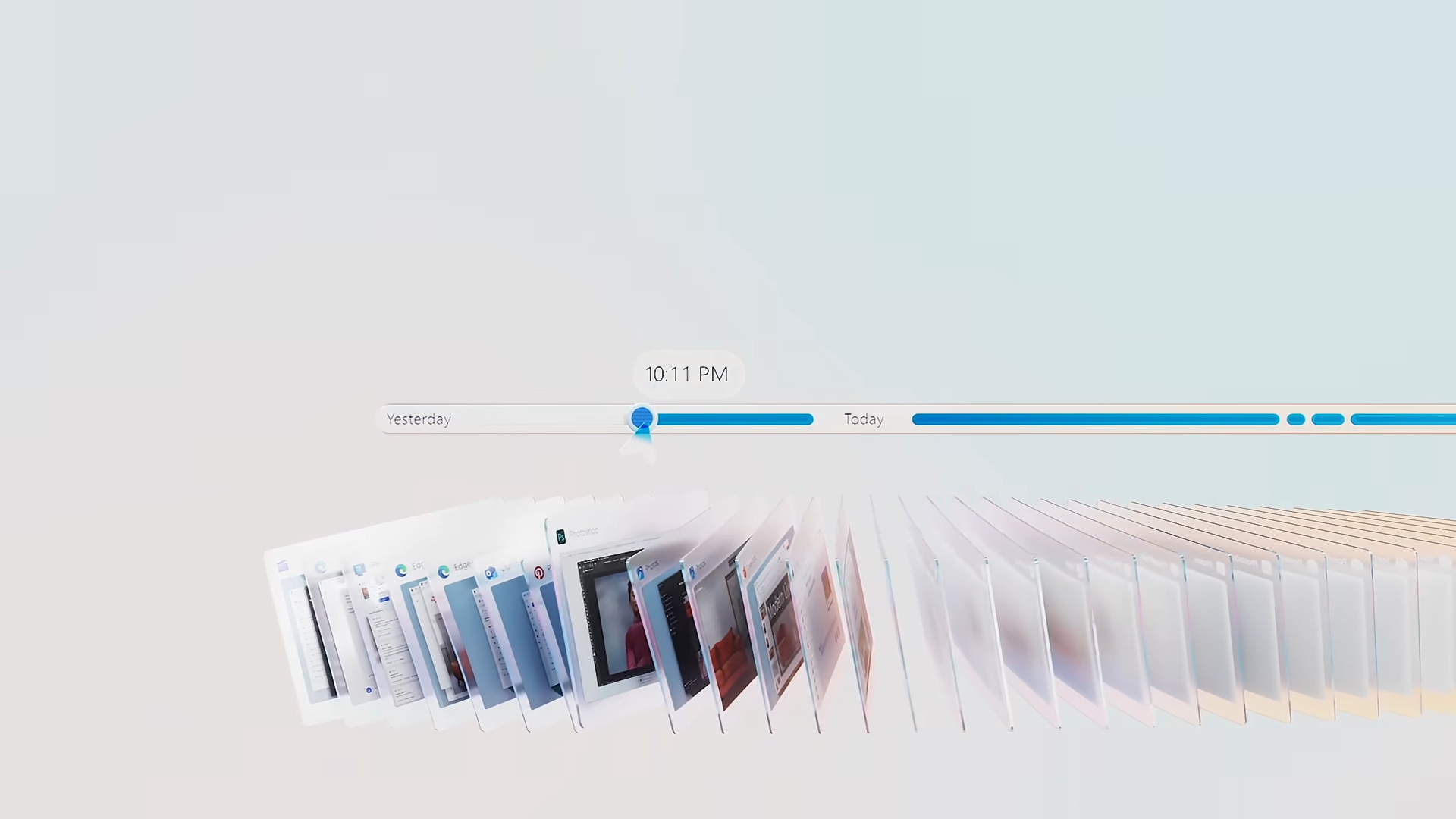
1. Timeline

- 对关键帧数据，在时间轴上做展示，焦点概览展示时间、截图和summary

- 默认焦点为当前最新记录时间，以当前焦点时间为轴，向前后各展示12小时轴数据

- 可以左右滑动焦点，或鼠标滚轮上下，向前后挪动焦点位置

- 可以通过点击Timeline上的timestamp，或输入Datetime，快速定位焦点



- 点击焦点概览，查看详细，除概览信息，还包括OCR结果、正在使用 的app、前后20s音频流（if have）、前后20s视频流（if have）、前后20s设备操作和System API列表（if have）

1. Search

- 由Timeline搜索入口进行输入，并提供搜索API给到对话和记忆使用

- 基于多模态数据(image, text)，提取对应的搜索指令（timescope，keyword）

- 基于搜索指令，在向量数据库中进行检索，返回对应关键帧信息，检索范围包括：

- timescope搜索如下列：

Table关键帧（timestamp）

Table设备操作（timestamp）

Table System\_API\_Call（timestamp）

Table屏幕音频流（timestamp）

- keyword搜索如下列：

Table关键帧(OCR\_RS, summary\_RS)

Table System\_API\_Call(API\_name, API\_description)

Table 屏幕音频流(type, Text)

- 如返回多条搜索结果，则展示结果列表，并基于模板对多条结果生成概括总结

1. Support for Router && Memory && Real Time Chat

- Router意图：

- recall timeline

- recall search

- Memory调用：

- recall search

- Chat调用：

- Table 环境音频流

1. Data Support for Training Personal Model
2. Data format for GUI Grounding

{

    "current\_shot":"image url/image tensor",

    "task\_instruction":"task instructions",

    "bbox":[

        "(left\_up\_x,left\_up\_y,right\_down\_x,right\_down\_y)",

        "(left\_up\_x,left\_up\_y,right\_down\_x,right\_down\_y)",

        "……"

    ]

}

1. Data Format for Operation Generating

{

    "current\_shot":"image url/image tensor",

    "task\_instruction":"task instructions",

    "task\_objective":"task objective",

    "last\_operations":[

        {

            "action\_type":"km-输入设备/api-app接口/template-模板/success-完成",

            "action\_id":"km\_id/api\_url/template\_id/",

            "bbox":"(left\_up\_x,left\_up\_y,right\_down\_x,right\_down\_y)",

            "……":"……"

        },

        {

            "action\_type":"km-输入设备/api-app接口/template-模板/success-完成",

            "action\_id":"km\_id/api\_url/template\_id/",

            "bbox":"(left\_up\_x,left\_up\_y,right\_down\_x,right\_down\_y)",

            "……":"……"

        },

        {

            "……":"……"

        }

    ],

    "next\_operation":{

        "action\_type":"km-输入设备/api-app接口/template-模板/success-完成",

        "action\_id":"km\_id/api\_url/template\_id/",

        "bbox":"(left\_up\_x,left\_up\_y,right\_down\_x,right\_down\_y)",

        "……":"……"

    }

}

1. Data Format for Task Planning

{

    "current\_shot":"image url/image tensor",

    "task\_instruction":"task instructions",

    "task\_objective":"task objective",

    "task\_plan":[

        {

            "app":"appid / api",

            "subtask\_instrction":"subtask instructions",

            "subtask\_objective":"subtask objective",

            "calling\_url":"endpoint and parameters"

        },

        {

            "app":"appid / api",

            "subtask\_instrction":"subtask instructions",

            "subtask\_objective":"subtask objective",

            "calling\_url":"endpoint and parameters"

        },

        {

            "……":"……"

        }

    ]

}

1. 收集和清晰方法

- 记录真实人员操作

- 对Table 关键帧、Table 设备操作、Table API Call三个表数据做关联处理

- 过程中通过语言指令补充任务信息 / 过程后通过文本补充任务信息

# Real Time Chat Assistant

1. Router

# GUI Auto Working Assistant

1. Task Planning
2. Online Skills Mount
3. Applications Mount（Including System Apps）
4. Model Call
5. Operation Generating
6. App Skills Mount
7. Model Call
8. Operation Auto
9. GUI Grounding and KM Operations Call
10. Application API Call
11. Template
12. GUI Task Manager
13. Auto Learning

# Model Manager