**Soma-产品介绍**

**Soma 是什么**

Soma 是一款“面向个人与家庭记忆的 AI 数字永生平台”。用户可在 Soma 中：

* 采集并托管逝者（或生前授权者）的多模态人生资料（文字、语音、影像、社交痕迹等），并由 Soma 构建“数字人格（Digital Persona）”；
* 与该数字人格在聊天、语音、纪念日仪式等场景中自然交互，延续精神陪伴与共情；
* 让家族成员共同整理、追忆与传承，“把记忆从数据变成关系网络”，形成一个私密、可信、可持续的家庭记忆系统。

产品价值的本质：用可信任的“人格建模 + 记忆图谱 + 安全合规”将碎片的记忆重构为可交互的永续存在，并将悲伤转化为温柔而有边界的连接。

**用户故事与核心场景**

**1) 纪念与对话**

* 用户上传父亲的照片、语音、书信、社交内容、行程与爱好记录（均需生前授权或合法继承人合规授予），Soma 生成“父亲的数字人格”。
* 在“纪念日”或日常时刻，用户打开聊天（Chat.tsx）与父亲对话，提出问题或回忆生活片段；Soma 会检索父亲记忆并以父亲的语气风格回复，带有温度与有界限的真实感。

**2) 回忆整理与家族协作**

* 家人通过“Memories”页面（Memories.tsx）协同标注旧照片、视频、语音，形成时序化的“记忆流”。每次编辑都会增强数字人格的记忆图谱。
* “Feed”（Feed.tsx）聚合家庭与亲友的动态纪念内容，形成温柔流动的数字纪念墙。

**3) 仪式化陪伴**

* 在重要日期（生日、忌日、节日），Soma 触发“仪式模式”：用更慢的节奏与更温柔的表达呈现回忆，并启用“我记得你”的双向问答。
* “Settings”（Settings.tsx）里，家属可设置“仪式强度”、“触发频率”、“内容边界”等，确保节度与心理安全。

**4) 家族数字遗产**

* 生者在世时可创建“数字遗嘱”与“记忆交接计划”，指定哪些内容在何种条件下开放给哪些家族成员（分层授权）。

**体验设计与 UI 动效（youware 风格）**

* 我们在所有主要页面统一启用 youware 风格的“自下而上、快速弹出且带轻微回弹”的动效，强调“记忆从深处被唤起”的感觉。
* 技术上通过 framer-motion 统一封装 variants（motion.ts），并在 Memories、Feed、Settings 应用 containerReverseStagger + bounceUp + fadeDown，保持沉浸、清爽、现代的体验。
* 页面结构：
* Memories：记忆时间线、创建按钮、下拉刷新、骨架屏等；
* Feed：家族/社交回忆流；
* Chat：与数字人格的对话；
* Marketplace：人格增强/插件生态；
* Settings：隐私、授权、家族成员、遗嘱、合规设置等。

**技术架构与流程**

**前端（现状）**

* 技术栈：React 18 + TypeScript + Vite + Tailwind + Framer Motion + Zustand/Sonner 等。
* 代码结构：
* 页面：Chat.tsx、Memories.tsx、Feed.tsx、Marketplace.tsx、Settings.tsx
* 组件：components（记忆卡片、底部导航、浮动按钮、弹出框等）
* 动效配置：motion.ts（youware 风格统一）
* 特征：移动端优先、轻量快速的交互、你能感到“记忆被唤起”的动效节奏。

**后端（现状与演进）**

* 现有微服务：Self\_AI\_Agent（Node.js + Express + TypeScript + Zod）
* Provider 抽象（src/providers/provider.ts，openai.ts）：可接入不同 LLM 供应商
* 训练/数据管道（src/pipeline/dataset.ts、train.ts）：整理数据、构建人格/索引
* 配置/类型：src/utils/config.ts、src/types/index.ts，保证类型安全与可观察性
* 未来架构（分层）：

1. API Gateway：鉴权、速率限制、特权路由与审计；
2. Ingestion Service：数据接入（文本、图片、音频、视频、社交平台 API）、清洗与去重；
3. Consent & Governance Service：授权/遗嘱/执行人/访问控制策略（基于策略引擎），不可变审计日志；
4. Profile Builder：将规范化后的多模态数据映射为“人格特征 + 价值观 + 语言风格 + 关系图”的结构化画像；
5. Memory Graph 与向量检索：

* 存储：Postgres（结构数据）+ pgvector（语义向量）+ 对象存储（媒体）；
* 可扩展：外部向量库（Pinecone/Weaviate）与图数据库（Neo4j）；

1. Persona Runtime（对话/仪式/任务）：

* RAG 管道：检索相关回忆 → 约束型生成（Persona Guardrail）→ 语音合成（可选）；
* 安全层/伦理边界：话术边界、情绪守护、事实来源引用；

1. Training & Fine-tuning：

* 轻量化微调（LoRA/PEFT）或指令后训练，周期性批处理 + 在线小步蒸馏；
* Cache/Distill：将高成本大模型行为蒸馏为低成本推理代理，控制 COGS；

1. Observability：

* 质量评测（Persona 一致性分、事实核验分）；
* 成本/延迟/稳定性监控，自动扩缩容（HPA）。

**核心数据流（简述）**

1. Ingestion：用户上传/连接源 → 解析转写（ASR）、OCR、元数据提取、去重/分段
2. Normalize：时间轴、人物关系、主题聚类、隐私与敏感类型标注
3. Embed & Index：文本/音频特征向量化，入库 pgvector/图谱
4. Persona Build：语言风格/价值观/表达偏好自动建模 + 家属校准
5. Runtime：

* Chat：query → 检索相关记忆 → 安全/合规过滤 → LLM 生成 → 可选语音合成 → 返回
* Ritual：节日触发 → 低频更温柔策略 + 指定回忆片段优先级 → 输出

1. Feedback & Training：用户纠错/点赞/添加回忆 → 作为监督信号更新索引与微调样本

**模型与评测（要点）**

* RAG 为主 + 轻量个性化微调（减少“幻觉”，确保“像这个人”）
* Persona Consistency Score（PCS）：衡量回复与人格语料的一致性
* 简式定义：PCS = cos(Emb(response), Emb(persona\_corpus\_centroid))
* 安全边界：拒绝代替关键决策、避免捏造现实、保持“我是数字模拟”的自我声明

**合规、伦理与安全**

* 核心原则：Consent-first（授权优先）、Purpose-limited（用途限定）、Safety-by-design（内建安全）：
* 授权来源（生前授权/合法继承/执行人证明），用途与可见范围可配置（Settings 页面分层经授权）
* 不可变审计日志（append-only hash 链），访问每次留痕
* 数据加密（静态 AES-256 + KMS；传输 TLS），隐私分级（敏感内容分级隔离）
* “身份/真实性提示”：界面明确标注“数字模拟，不等价于本人”
* 内容安全与心理健康：触发词守护、悲伤危机干预提示、本地心理援助资源链接
* “被遗忘权”与“数据回收”：按法律与用户请求清除可识别数据，保留不可识别统计（如允许）

**商业模式与市场进入**

**B2C：订阅 + 增值**

* 免费层：基础记忆存储与文本对话（限配额）
* 个人高级：更多存储、高清音视频、语音互动、更多“人格精细化”训练次数
* 家庭套餐：多人协作、家族空间、家谱关系、跨代传承
* 增值包：纪念日仪式强化、纪念视频自动生成、打印相册、云祭扫 3D 空间等

**B2B：渠道与机构**

* 殡葬/纪念服务机构：白标/联名，线下仪式数字化延展
* 遗嘱与财产规划机构：数字遗产模块嵌入
* 文化/博物馆/图书馆：名人档案/口述史互动化
* 医疗/护理/临终关怀：告别仪式与家属支持工具（严格合规）

**收入拆解（示意）**

* 订阅 ARPU、增值包 ARPPU、B2B 许可或流水分成
* COGS 控制：多级模型路由（大模型兜底，小模型主力）、向量缓存、批量异步训练、冷热分层存储

小公式（成本估算）：

* 单次会话成本（估）：
* 目标毛利率：> 70%

**运营机制与增长**

* AARRR 漏斗与关键引导：
* 激活：导入回忆向导（社交平台/相册/短信备份），5 分钟内看到第一次“像他/她”的回复
* 留存：纪念日/每周回忆提醒、家庭成员邀请共建
* 传播：家庭邀请制（每位成员都有邀请额度）、纪念内容一键生成短视频（带引导链接，合规前提下可分享外网）
* 内容生态与市场（Marketplace）：
* 插件：语音克隆（合规验证后）、纪念视频模板、AI 相册、家谱可视化
* 第三方开发者：围绕记忆/仪式/互动工具构建，平台抽成
* 品牌与信任建设：
* 与权威机构/心理咨询协会合作发布指引
* 透明的伦理委员会公示、年度报告

**产品路线图**

* v0（已完成基础）：记忆流、对话、统一动效、前后端骨架（Self\_AI\_Agent）
* v1：数字人格 V1（强 RAG + 风格控制）、家庭空间与协同标注、基础授权治理
* v1.5：合规语音互动、纪念仪式模式、纪念视频生成
* v2：多模态长期记忆图谱、人物关系图/家谱、个性化 LoRA 轻微调
* v2.5：Agentic “记忆管家”：自动整理归类、每周刻面总结、家庭故事册导出
* v3：文化记忆网络（可选对外授权的公共记忆片段）、研究合作

**指标体系（KPI）**

* 北极星指标：家庭月活互动时长（含对话与回忆整理）
* 关键行为：
* TTFP（Time To First Persona）：从注册到获得“像他/她”的第一次回复的时间
* 每周对话次数、每月仪式参与率、家庭成员协作人数
* Persona 一致性评分（PCS）、事实引用率、用户纠错/确认率
* 业务指标：订阅转化率、家庭套餐占比、ARPU、毛利率、流失率
* 合规模块：授权完备率、访问审计通过率、安全事件为 0

**竞争态势与壁垒**

* 差异化优势：
* “合规 + 家族协作 + 记忆图谱”的一体化产品形态（不是单一聊天或单一相册）
* Persona 对齐的工程化能力（RAG + 风格控制 + 伦理边界 + 评测体系）
* 渠道合作（殡葬/遗嘱/文化）与信任品牌建设
* 开放生态（Marketplace）扩展仪式与创作工具
* 数据护城河：
* 高质量标注的家族记忆图谱
* 授权与治理“账本”形成的合规与信任门槛

**风险与对策**

* 伦理/舆论风险：深度拟拟态引发不适或反感
* 对策：默认“温柔、边界明确”的话术；强提示“数字模拟”；提供易用的关闭/淡化开关；设立伦理委员会
* 法务/合规风险：授权不充分、跨境数据
* 对策：严格 KYC/身份核验；执行人证明流程；数据分区与本地化；定期审计
* 技术风险：幻觉、成本过高、响应延迟
* 对策：RAG 为主、知识引用、蒸馏与缓存、多级模型路由；异步生成+流式响应
* 安全风险：数据泄露/滥用
* 对策：强加密、零信任、最小权限、入侵检测、红队演练、Bug Bounty

**技术落地细节（面向工程/投资人）**

* 代码基座（现有）：
* 前端：src/pages/\* 与 src/components/\*，统一动效在 motion.ts
* 后端雏形：Self\_AI\_Agent（Provider 抽象 + Pipeline + REST）
* “训练/运行”合同（Contract，简式）
* 输入：授权档案（多模态）、家属校准偏好
* 过程：清洗→索引→人格建模→（可选）轻微调→上线
* 输出：可被调用的人格代理（REST/WS），具备检索与风格控制
* 错误模式：授权缺失/无足够语料/越界请求/生成安全拦截
* 成功判据：响应延迟 < 1.5s（文本）、PCS > 阈值、事实引用率 > 阈值
* 可观测与评测：
* Persona 一致性（文本嵌入相似度）、记忆命中率（Top-k 命中）、安全拒绝正确性
* A/B 实验层：仪式模式话术、动效节奏、内容召回策略

**我们为什么能做成**

* 对“记忆”进行工程化重建：从“数据资产”到“人-事-时-情”的可交互系统
* 在体验上，Soma 选择“温柔有边界”的 youware 风格，塑造与现实情感相匹配的节奏
* 在合规与伦理上，内建治理是第一性原理，而不是后补措施
* 在商业上，B2C 有强需求、B2B 有强渠道，且 COGS 可通过工程手段逐步优化

如果你希望，我可以将本介绍生成投资材料 PDF 大纲（提案页/图示/里程碑图），或进一步补充财务测算表（COGS、LTV/CAC、分阶段预算），并把系统架构图落成可视化稿（Mermaid/图形稿）以便对外分享。