

42.9%

✓ SOTA on Benchmark

WPSR (Quality) **34.6**
MATCR (Stability) **48.8%**

Level-2 (Intermediate) **40.0%**
唯一的 Level-2 高可用模型

Mobile-Agent-e (Gemini2.5-pro)

22.9

↗ Framework Gain: 使用相同 Gemini2.5-pro 后端, LightManus 相比 Mobile-Agent-e 提升了 74%.

Model	Level-1	Level-2	Level-3	WPSR
LM + Claude-3.5-Sonnet	60.0	40.0	20.0	34.6
LM + Gemini2.5-pro	73.3	20.0	10.0	29.9
LM + GPT-o3	60.0	30.0	0.0	22.0
Mobile-Agent-e	46.7	10.0	0.0	23.2

7



3. Jarvis 设计：双流感知与去噪机制

① 视觉语义流 (Visual Stream)

利用 VLM (Qwen2.5-VL) 的强多模态能力, 通过 **截图 (Screenshot)** 理解非结构化信息 (如图标含义、图片内容、布局逻辑), 弥补 XML 的语义缺失。

</> 视口结构流 (Viewport-Aware Stream)

解析 UI Automator XML, 但不仅仅是解析:

- 视口过滤:** 剔除 bounds 在屏幕可视区域外的无效节点, 消除模型幻觉。
- 信息压缩:** 截断过长文本, 仅保留 clickable 等关键属性, 降低 Token 负载。

