JISUMEN MailHalo 使用说明书（1.0.6）

版本：1.0.6 • 适用系统：Windows 10 / 11（64 位） • 配套软件：Foxmail 7.x • 主程序：MailHalo.exe

MailHalo（内部代号 FoxPet）是贴在 Foxmail 窗口右侧的 AI 助手侧栏。它只读取你正在查看的邮件内容并发送给你配置的模型服务；不会自动发送邮件，不会自动保存或关闭写信窗口，不修改、不注入 Foxmail。

1.0.6 的核心变化是翻译流式预览：整封邮件翻译时，译文边翻译边显示，不用等全部块翻译完。本文按 1.0.6 的实际行为编写；关闭流式开关时，程序行为与 1.0.5 完全一致。

1. 安装与首次配置

1.1 安装

- 免安装：拿到官方离线包 mailhalo-{{PACKAGE_NAME}}.zip（包名与 SHA256 见发布页），解压到你有写权限的目录（例如 D:\MailHalo），不要放在受保护的 Program Files 下。
- 主程序是解压目录里的 MailHalo.exe，双击运行；同一时间只允许一个实例。
- 显示器物理分辨率必须不低于 1920 × 1080，否则程序会提示并退出。

1.2 首次配置

第一次启动会要求设置管理员密码（只存本机 %LOCALAPPDATA%\JISUMEN\FoxPet\account.dat，存哈希不存明文），登录后进入主界面。然后配置模型服务：

配置项	说明	示例
api_base	模型服务的 OpenAI 兼容地址	https://api.deepseek.com
model	服务端提供的模型名	deepseek-flash
api_key	你的密钥，保存后加密存放	申请自服务商

配置写在程序目录的 foxpet.json 里。配置文件必须与 MailHalo.exe 在同一目录，程序只从 exe 所在目录读取；找不到时会用内置默认值并在日志中记录。首次运行会自动生成默认配置，把 api_key 换成你自己的即可。

修改配置后重启 MailHalo 生效。配置是否按预期生效，看日志里的 config: translation 行（见第 6 节）。

1.3 翻译策略的三层优先级

翻译相关参数（分块大小、校验、流式开关等）按下面的优先级生效，高层覆盖低层：

第 1 层（最高）：`low_resource` —— 低资源语言独立一套，命中即整体接管，不看 `profile`

第 2 层： active_profile —— 指定的 profile 覆盖顶层同名字段

第 3 层（最低）：顶层字段 —— 兼容旧配置的默认值

- 顶层字段是默认值；active_profile 指向 profiles 里某个 profile 时，该 profile 的字段覆盖顶层同名字段（没写的字段继续用顶层的）。
- 邮件语言命中 low_resource.languages 列表时，直接使用 low_resource 段的 api_base / model / translation，不经过 profile——两套是独立的。
- 判断当前哪一层生效：日志 config: loaded 行末尾的 profile= 显示激活的 profile；翻译低资源语言时会出现 translate: route low_resource=xx model=... 行。

示例片段（节选自 foxpet.example.json，展示三层各自的位置）：

```
{
  "api_base": "https://api.deepseek.com",
  "model": "deepseek-flash",
  "translation": {
    "chunk_chars": 3000,
    "stream_enabled": false,
    "stream_flush_ms": 75
  },
  "active_profile": "deepseek-public",
  "profiles": {
    "deepseek-public": {
      "translation": {
        "chunk_chars": 3000,
        "validate_output": true,
        "stream_enabled": false,
        "stream_flush_ms": 75,
        "max_retry_depth": 1
      }
    }
  },
  "low_resource": {
    "languages": ["zu", "sw", "am", "bn", "ta", "th", "he", "el", "vi"],
    "api_base": "https://api.deepseek.com",
    "model": "deepseek-flash",
```

```
    "translation": {
      "chunk_chars": 3000,
      "validate_output": true,
      "stream_enabled": false,
      "stream_flush_ms": 75
    }
  }
}
```

想只给低资源语言开流式：把 low_resource.translation.stream_enabled 设为 true，profile 与顶层保持 false。

2. 翻译流式预览（1.0.6 新增）

2.1 开关

在配置文件的 translation 段：

字段	类型	说明
stream_enabled	布尔	流式预览总开关，默认 false（关闭）
stream_flush_ms	整数	预览刷新合并窗口，单位毫秒；配置值超出 50–100 时程序自动夹取到 50–100，缺省 75

stream_flush_ms 的含义：预览帧最多每这么久上屏一次，把频繁的小增量合并成一次刷新，避免界面抖动。设 50 更跟手、刷新更密；设 100 更省。块翻译完成（验收通过）的那一帧不受这个窗口节流，验收结果立即上屏。

2.2 预览的内容：已验收前缀 + 未校验尾段

预览不是"模型吐什么就显示什么"，它由两部分拼成：

- 已验收前缀：从第一块起连续通过质量校验的块拼接的文本。这部分与最终译文一致，可以放心阅读。
- 当前块的未校验尾段：正在翻译的那一块已经流出的文字，尚未通过校验，可能被后续重试整段替换或撤销。

因此看到的内容会呈现"前面稳定、末尾变化"的形态：末尾几十到几百字可能滚动、回退或整段重写，这是设计行为，不是故障。预览不写缓存、不计入完成进度，随时可撤销；只有验收通过的块才会进入最终译文和缓存。

2.3 预览显示在哪里

预览跟随你当前的视图：

- 浮层：贴在 Foxmail 阅读区旁的译文浮层。没开工作台时，预览直接增量追加到浮层里，逐段长出来。
- 工作台：对照阅读工作台打开时，浮层不显示，预览走工作台自己的增量通道，追加在译文栏末尾；段落对齐等索引在翻译全部完成后重建。

也就是说：工作台开着，预览就进工作台；只看浮层，预览就进浮层，不需要任何额外设置。

2.4 关闭开关时的行为

stream_enabled: false（默认）时，不创建预览通道、不发流式请求，翻译流程与 1.0.5 完全一致：按块翻译，块完成后按 1.0.5 的渐进回显逻辑显示，最终结果一次性定稿。升级到 1.0.6 但没开这个开关的用户，感知不到任何变化。

2.5 已知边界

- 流式预览只对「整封邮件翻译 + validate_output 打开」生效。validate_output 关闭时，即使 stream_enabled 打开也不走预览（质量校验是预览验收机制的前提）。
- 选中翻译不走流式：对选中文字的翻译仍按原流程一次出结果。
- 服务端忽略 stream 参数时（返回普通 JSON 而非 SSE 事件流），解码器判协议失败，上层按失败处理并自动重试，不会把半截内容当成功显示。
- 低资源语言路由是否走流式，由 low_resource.translation.stream_enabled 独立决定，与主路由开关无关。

3. 日常使用

- 看邮件：在 Foxmail 选中邮件，侧栏点「邮件对照翻译」；译文出现在浮层或工作台。
- 流式预览体验（开关打开后）：点翻译后数秒内浮层/工作台开始逐段出现译文，先出的部分是已验收前缀，末尾在变化的是当前块；全部块验收完成后译文定稿。
- 取消翻译：翻译进行中再次操作或切换邮件即取消；已显示的预览随之失效，不会残留半截译文当结果。
- AI 回复 / 待办提取 / 润色：与 1.0.5 一致，详见 1.0.5 说明书的对应章节，这些功能不涉及流式预览。

4. 代理与 VPN

MailHalo 直接连你配置的 api_base，不经过任何本地代理服务。如果你电脑上挂着代理 / VPN（Clash、v2rayN 等客户端），请配置分流规则，让以下两个目标直连：

目标	规则	原因
MailHalo.exe	进程名直连	程序自身流量走代理会叠加延迟与失败重试
deepseek.com（及你的	域名直连	国内可直连的模型服务，绕道

api_base 域名）		境外节点显著变慢

示例（Clash 系规则语法）：

```
rules:
  - PROCESS-NAME,MailHalo.exe,DIRECT
  - DOMAIN-SUFFIX,deepseek.com,DIRECT
```

特别注意：开启 TUN 模式（全局虚拟网卡接管）会显著变慢。 TUN 模式下所有流量都被虚拟网卡截获，进程级直连规则可能失效，流式预览的逐块到达间隔会被拉长，表现为预览长时间不动、突然一大段涌出。排查网络问题时，先关 TUN 再看。

5. 故障排查

5.1 打开日志

发行版默认不写日志。复现问题时，用命令行带 -debug 启动：

```
MailHalo.exe -debug
```

日志写入：

```
%LOCALAPPDATA%\JISUMEN\FoxPet\foxpet.log
```

（在文件资源管理器地址栏粘贴 %LOCALAPPDATA%\JISUMEN\FoxPet 回车即可找到。）

5.2 四条关键日志行

排查流式翻译问题时，按出现顺序看这四行：

日志行开头	出现时机	看什么
config: translation	启动加载配置后	stream=1/0（开关是否生效）、flush_ms=（实际夹取后的值）、lr_stream= / lr_flush_ms=（低资源层）
model: stream	每次流式请求结束	end= 结束原因、chars= 收到字数、first_ms= 首字延迟、finish= 结束标志
overlay: preview	每次预览上屏	chars= 本次上屏字数、show_ms= 渲染耗时（应为毫秒级）
translate: accepted chunk	每块验收通过	chunks_done=x/y 整体进度、flush_ms= 本轮节流窗口
perf: stream	一次流式翻译收尾	first_delta_ms= 模型首字、first_visible_ms= 首帧可见耗时、frames= 预览帧数、show_ms_p50/max= 每帧端到端上屏耗

		时、accept_chunks= 验收块数

速查：

- config: translation 里 stream=0 但你已改配置 → 配置文件不在 exe 目录，或字段写在了错误的层（见 1.3 节三层优先级）。
- model: stream end= 后跟非正常结束码、chars=0 → 服务端可能忽略 stream 返回了 JSON（协议失败），或网络中断；结合 finish= 判断。
- 完全没有 overlay: preview 行 → 开关没生效，或这封邮件是选中翻译（选中翻译不走流式）。
- 预览长时间不动 → 检查是否开了 TUN 模式 / 代理分流（第 4 节）。
- 想量化速度 → 看 perf: stream 行：first_visible_ms 就是"用户看到第一段译文"的毫秒数，show_ms_p50 是每帧渲染开销（正常应 ≤5ms）。

5.3 常见问题

Q：改了 stream_enabled: true 但没有预览。

依次确认：① 配置文件与 MailHalo.exe 同目录；② 字段写在 active_profile 指向的 profile 里（profile 会覆盖顶层）；③ 这封邮件是整封翻译且 validate_output 为 true；④ 重启程序后看 config: translation 行的 stream= 值。

Q：预览末尾文字反复变化。

正常现象，是当前块的未校验尾段（见 2.2 节）。已验收前缀部分不会变。

Q：翻译中途切换了邮件。

旧翻译被取消，预览失效；新邮件重新翻译，日志里 epoch 变化。不会把旧邮件的译文挂到新邮件上。

6. 数据与隐私

内容	位置
运行日志（-debug 时）	%LOCALAPPDATA%\JISUMEN\FoxPet\foxpet.log
账号	%LOCALAPPDATA%\JISUMEN\FoxPet\account.dat
配置	程序目录 foxpet.json（API Key 加密存放）

程序只读取你正在 Foxmail 中查看的邮件内容，并发送给你配置的模型服务处理。数据边界要说清楚：使用云端 API 时，正文会离开本机、离开内网；改用本地或内网模型服务，正文仍会离开本机进入该服务，只是不出内网。除此之外不向其它服务器上传数据；日志不记录 API Key 与密码。

反馈问题时请提供：现象、操作步骤、-debug 模式下复现时的 foxpet.log，以及截图/录屏（涉及界面时）。

本说明书对应 MailHalo v1.0.6（2026-11-14）。功能以实际安装版本为准。