# 不敢把个人信息喂给 AI？OneAIFW 简单搞定隐私保护！

原创 飞哥数智谈 飞哥数智谈 2025年12月9日 22:21 山东

最近，在做几个落地的 AI 项目，发现“**隐私**”是无法绕开的一个关键挑战。

不是简单的达到“合规”，而是真正的避免客户的手机号、姓名等个人信息上传到云端大模型。

> 本文暂不讨论私有部署大模型和端侧模型——它们虽理想，但并非所有场景都适用。

我也了解了挺多方案，但是感觉都不那么自然或者简单，直到我看到 `OneAIFW` （一个AI防火墙？）。

**OneAIFW**

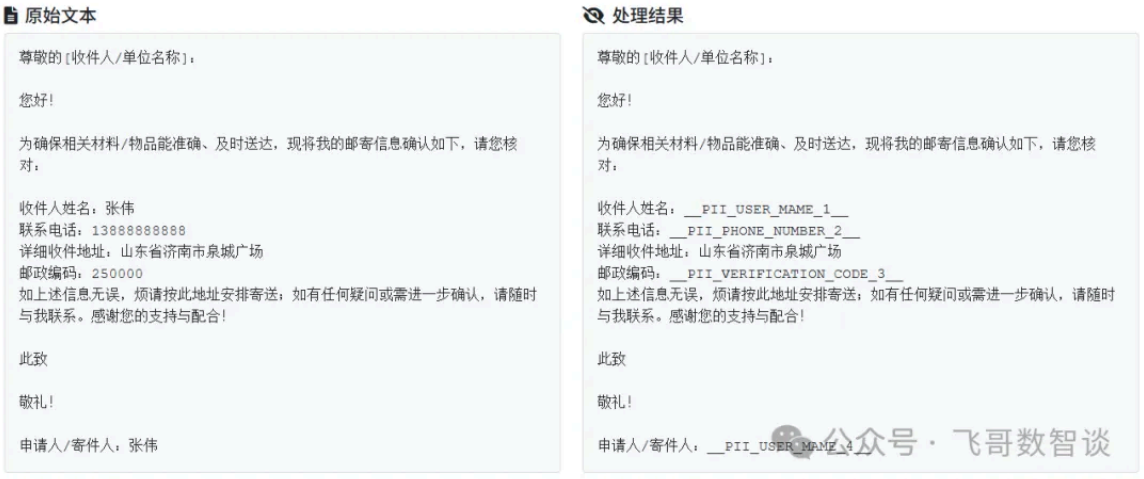
连名字都是那么的朴实无华。

`OneAIFW` 开源在 `GitHub` 上，采用 `MIT` 协议。

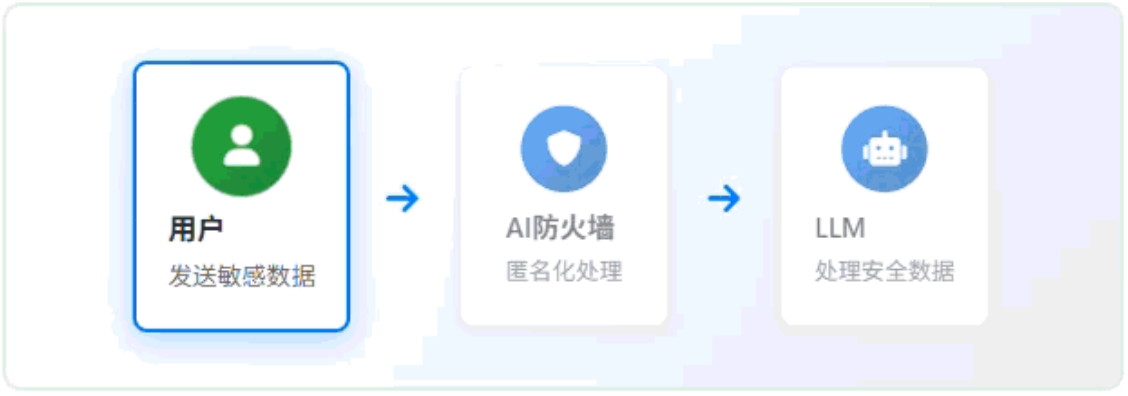
虽然目前 `star` 还不多，但我感觉前途可期。

它的原理简单概括就是：

1. **脱敏**：将发给大模型的文字中的隐私数据替换为为结构化占位符。
2. 大模型接收的数据依然保留原有结构，不影响语义理解。
3. **还原**：将大模型返回的文字中的为结构化占位符替换为原始文字。




是不是很简单，了解原理后我只有一个感觉：**这么简单的思路，我为什么没有想出来。**



## 保护内容

目前支持脱敏的内容：

- 物理地址
- 邮箱地址
- 姓名 / 用户名
- 公司 / 组织名
- 电话号码
- 银行账户 / 卡号
- 支付信息
- 验证码
- 密码
- 随机种子
- 私钥
- URL 地址

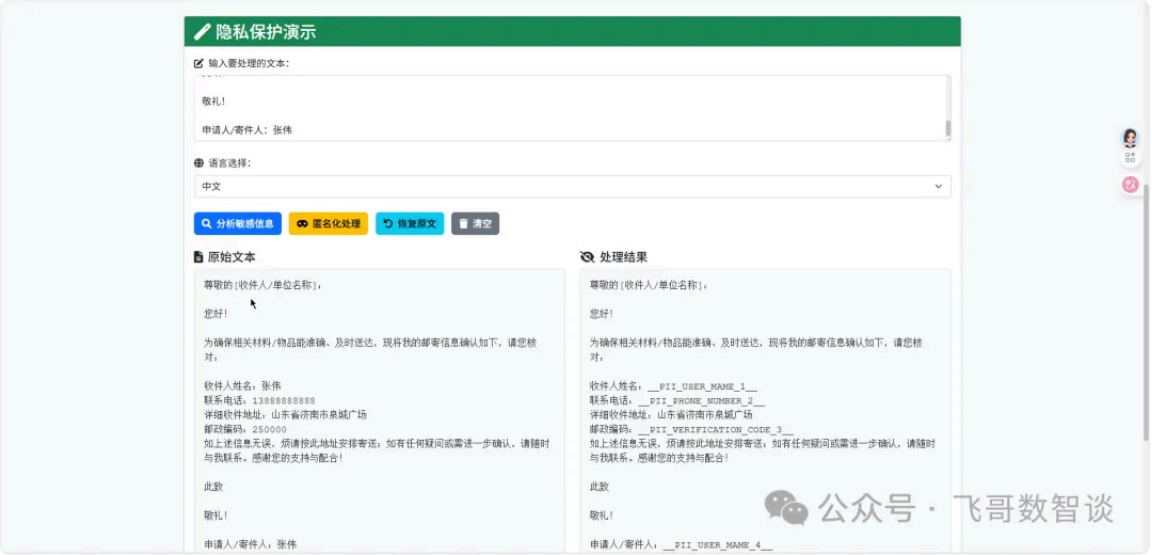
## 在线体验

光说不练假把式，不如直接上手试试。

官方提供了一个线上 `demo`，可以在线体验。

地址：`https://oneaifw.com/`

可以直接分析敏感信息，也可以直接尝试脱敏的效果。




## 使用方式

该项目另一个很赞的点就是，它提供了我们常用的各种方式使用，包括本地 Web、浏览器插件、本地 API、本地命令行。

### 本地 Web

一个轻量前端，依赖 `@oneaifw/aifw-js` 库，在浏览器里跑完整流程。

适合快速验证或嵌入现有 `Web` 应用。

#### 浏览器插件

提供 `Chrome/Edge` 扩展示例，可以注入到各类大模型网站，在你点击“发送”前自动脱敏。

这个估计是日常使用最佳的方式了。

### 本地 API

通过 `py-origin` 或 `cli/python` 启动一个本地 HTTP 服务，提供 `/api/mask_text`、`/api/restore_text`、`/api/call` 等接口。

如果你和我一样正在开发 AI 项目，可以考虑通过 `API` 方式接入。

### 本地命令行

装好 Python 包后，支持命令行直接调用，比如：

```
python -m aifw call "请把如下文本翻译为中文: My email address is test@exa
```

非常适合脚本调用或自动化流程。

## 结语

今天主要给大家分享了一个 AI 项目中解决**隐私**问题的产品，重点是设计思路，希望可以帮到大家。

我已经在跑通了 Web 版本，部署体验将在后续更新。

基座模型或许遥不可及，但 AI 时代的真正机会，往往藏在像 `OneAIFW` 这样的“**小而关键**”的环节里——它不炫技，只是让智能变得更安全、更可用。而这，正是我们每个人都能参与的方向。

飞哥数智谈

“ 谢谢老板，我会更加努力的～ ”

喜欢作者