AI 产品经理作品集

直聘打招呼助手

给求职者的工具：粘贴一段岗位 JD，用已保存的简历生成 3 条可直接发送的打招呼话术，以及逐条可采纳的简历改写建议；改完的简历能**按原文件的排版**导出。 所有内容必须能在简历原文找到依据——**编造即失败。**

在线体验：zhipin.xinhao02.ccwu.cc | PDF 版 | 单人项目（产品判断 + 实现 + 评测） | 数据截至 2026/09/21

（67 秒主流程录屏在 HTML 版与在线地址里：zhipin.xinhao02.ccwu.cc）

67 秒主流程（实测 67.24 秒）：保存简历 → 生成话术与匹配分析 → 复制（自动入历史）→ 按 JD 改简历 → 存为定制简历。

1.7 s
完整生成耗时（8 组样本实测，线上含公网往返 2.9 s）

0 例
编造（8 组样本 × 2 轮，数字与技能全部可溯源）

92.44%
测试覆盖率（阈值 80% 写在配置里）

557 项
自动化测试（单测 301 + 文档一致性 234 + 端到端 22）

证据边界

上面四个数字都是实测的。但**用户侧证据目前只有我自己的场景**——这一条决定了它们能证明什么、不能证明什么。

证据类型	状态
工程与评测数据	**实测**：557 项测试、覆盖率、限流压测、模型选型、规则基线对照，全部可复现
自访谈	1 份，受访者是我本人——属于自身场景的加强版，**不计入用户证据**
投递纵向记录	本人 2026 年 8–9 月 BOSS 直聘 30 天沟通列表里可见的 **17 条**记录：有回复率 **52.9%**（下限）、有效回复率 **11.8%**（2/17），其中至少 4 条是对方先联系。同样是自身场景——样本 1 人、未区分投递方向，**不是用户调研**
公开讨论材料	牛客网公开讨论检索 **16 条**（2026/09）：其中「已读不回 / 投了没人回」这类表达 10 条，方向覆盖嵌入式、Python、C++、前端、全栈、Java。属被动观察——样本自选、无法追问发布者，**不是访谈**
第三方用户证据	**0**：没有第三方访谈、问卷；公开讨论材料已整理，但它属于被动观察，**不等于第三方用户证据**

所以「用户是否真的需要它」这一层，仍然没有第三方证据。自访谈把两处口径写细了，投递纵向记录补上了第一人称的真实投递分布，但两者证明的都是「我这么用」，不是「这类人都这么用」。

补齐计划：已完成两条不依赖外部人员的替代证据——本人 30 天内沟通记录的纵向记录（docs/投递纵向记录_V1.0.md）与牛客网公开讨论材料 16 条（docs/公开讨论材料_V1.0.md）；**经本人决定不做**第三方访谈、外部盲测与主观质量评分——这三项如实声明，不留「还在等」的印象。

一、问题与目标用户

求职者在直聊类平台主动投递时有两件重复劳动：

- **逐条写打招呼**：复制同一段会被招聘者一眼看出模板，手写一条要 5–10 分钟，而招聘者通常只读前两句；
- **为每个岗位改简历**：成本高于打招呼本身，多数人干脆一份通用简历投所有岗位。

现有通用 AI 写作工具的问题在于：要用户自己调提示词，且倾向产出「贵公司岗位非常感兴趣」这类无信息量套话，甚至编造简历里没有的经历和数字。

目标用户：互联网技术岗与产品岗的主动投递者——JD 硬性要求明确，与简历的对照关系最清晰，最容易验证生成质量。 **目标**：从粘贴 JD 到复制一条敢直接发送的话术 ≤ 30 秒。

> **依据强度**：痛点首先来自本人（作者就是目标用户之一）；2026/09 做了一份结构化自访谈，把两处口径写细了，但**第三方用户证据仍为零**——没有第三方访谈、问卷或竞品实测，所以上面的描述里推演成分依然存在。这是复盘里的第一个短板。

二、四条关键判断

1. 不信任模型：把不可让渡的规则放在模型外面

个人信息与教育经历绝不允许被改写。最直接的做法是写进提示词，我没有这么做——模型是不可靠的执行者，越界是静默的。最终方案是：模型只输出「原文 → 建议写法」的片段，**替换由本地确定性代码执行**，锚点匹配不上就丢弃。 这样「逐字不变」从一句依赖模型自觉的期望，变成了一条普通单元测试。

2. 先建评测，再调提示词

第一轮真实模型回归，8 组样本里违规 11 条。逐条归因后分成三类：只写长度区间没强调下限、检查器把「2022.06」误判为编造、模型把 JD 的年限要求写成用户自己的经验。 分别对应三种处置，修正后条级合规率 95.8%。**这套归因链比「效果变好了」有用**——每次改提示词，回归脚本会告诉你哪条规则退化了。

3. 把「复制率」从条级改成会话级

原目标「零修改直接复制率 ≥ 50%」的口径是被复制话术中未经编辑的比例，这个数天然接近 1/3——用户在三选一里挑一条发，等于要求每 2 条就有 1 条不用改。 改为**会话级 ≥ 80%**（这一次有没有拿到敢发的话），更贴合产品目标；条级 40% 降为质量辅助线。

4. 导出的保真度不靠重新排版，靠原地回填

用户要的是「和原来排版一样」的简历。把简历解析成结构化数据再渲染，那是**重排**——字体、页边距、表格、页眉页脚都得重新对齐，误差不可控。 我选的是相反方向：原文件当模板，只替换被采纳建议所在的那几行文字，其余字节原样保留。

三、被放弃的方案

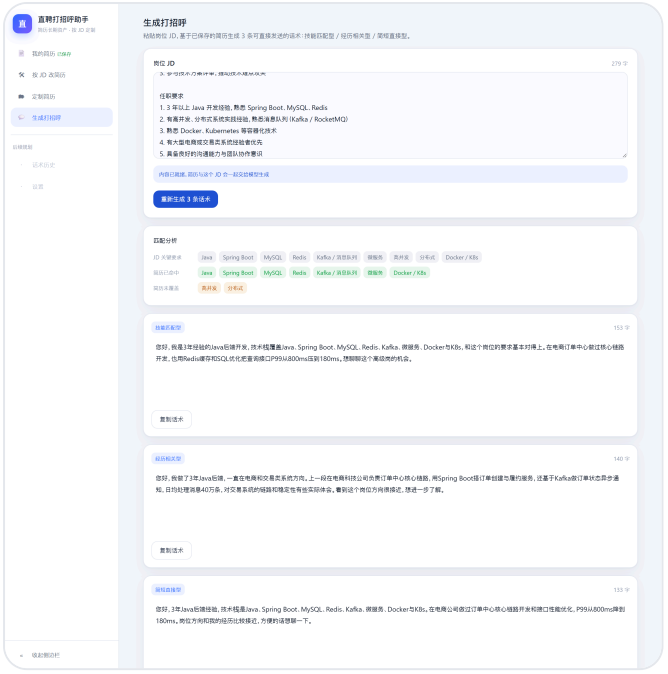
看起来更省事的路	为什么不做
让模型自己「输出前自检」字数与数字	等于让它猜自己的输出：结论不可验证，还要为这段自审多付输出 token。数字是否出现在简历里、有没有用黑名单词，本地是几行确定性比较。
PDF 先转 Word 再导出	转换本身是版面重建而非还原，同一份 PDF 用不同工具转出来的结果并不相同，导出结果不可预测。
在原 PDF 上覆盖改写	要嵌入中文字体，与「导出不触网」冲突；改写让文字变长还会溢出原行。
引入第二个模型做审核	既增加成本，又把「有没有编造」这个可判定的问题重新变成不可判定的问题。

四、评测体系

项目	结果
事实溯源	8 组样本 × 2 轮，**0 例编造**；数字、技能、学历均可在简历原文找到
话术条级合规率	95.8%（48 条中 2 条长度越界）。这是**合规率**不是质量率——它判不了「像不像真人写的」，那一层改由下面的人评承担
主观质量（人评）	8 组真实输出 × 3 维度，由本人按 1–5 判据打分：**A 像真人写的 3.00**、**B 敢直接发 3.75**、**C 无夸大 3.63**，总均分 **3.46**。6 组「基本可直接发」、2 组「改一两处才敢发」，**没有一组是「一个字不改直接发」**；最低分样本是目标人群之外的平面设计（A=2）。评分者即产品作者、单人评分无可算的一致率，**这是方法演示，不是外部证据**
Word 往返一致性	不采纳任何建议时，导出文件的正文与原件逐字一致；导出件能被自己的导入通道重新解析
模型选型	同批 8 组样本跑了三个候选：在用的模型延迟与 token 用量都最低，另两个分别是它的 5.6 倍与 30 倍耗时，且质量并未更好
AI 的增量	与纯规则基线（关键词匹配 + 模板）对照：规则版 22 处违规对模型版 6 处，重复句比例 0.514 对 0.03——模板方案的输出里一半的句子是原样复用的
外部盲测	**未做，且不计划做**：评分表与汇总工具已就绪（8 组样本，含评分者一致率），但经本人决定不组织外部盲测

关键点：**判定规则只有一份**——线上逐条自检与离线质量回归调用同一个函数；提示词是目标，自检是红线。

五、结果



本地计算匹配分析（JD 关键要求 / 已命中 / 未覆盖），再流式出现 3 条话术。

建议逐条可采纳，右侧实时合成；个人信息与教育经历永远保持原文。

指标	目标	实测
首字到达	≤ 5 s	0.46 s（16 次平均）；线上 2.20 s
完整结果	≤ 30 s	1.69 s（16 次平均）；线上 2.89 s
Word 保真导出	≤ 1 s	48–60 ms（全程本地，断网可用）
限流与配额	超限不消耗模型调用	压测三场景（60 / 200 / 160 个请求），模型调用数与配额完全一致
成本硬顶	公开传播可控	单日 300 次调用上限；一次完整生成约 1443 输入 + 596 输出 token

部署形态：一个 Cloudflare Worker 同时托管前端与 API，同域、无 CORS；密钥只在 Worker Secret。 简历与 JD 只存在浏览器本地，服务端只转发不存储。

六、定位与边界

会不会加重招聘方的无效沟通?

这是这个产品最该被质疑的地方，而且**这个质疑有一半成立。**如果把目标定成「把打招呼的数量做上去」，正确的做法是批量发送、模板库、一键投递——**这几件事我都没做。**

硬约束	为什么这么定
不自动发送	话术必须由本人复制、本人在平台上发出，产品不碰发送环节
匹配分析在生成之前	作用是帮用户判断「这个岗位该不该投」，而不是投得更多
禁止夸大	同时保护招聘方与求职者——自访谈里那次「写得比实际强、面试被问住」正是这条的由来

另有一条来自使用成本的观察：AI 辅助之后，写一条仍要 1–2 分钟用于审核，**这不是零成本群发。**受访者的原话是「如果不是我比较心仪的岗位，我一般不会使用」。

没有验证的部分：「匹配分析能不能真的减少低匹配投递」目前是设计意图，没有数据。

单周期产品，为什么接受

求职是低频刚需，用户拿到 offer 就会离开。我不认为这是缺陷。但这里有一个刻意的区分：**话术是消耗品，简历是资产**——简历被做成长期可维护的对象，话术历史只留最近 20 条。用户下次换工作时，资产还在。

出问题的时候，用户看到什么

模型不是每次都听话，服务也不是每次都活着。三条边界，**每一条要么有实现，要么写明没做和原因**:

场景	已实现	没做的部分与原因
模型输出了 bad case	服务端逐条本地自检（数字与技能能否溯源、套话黑名单、表情符号），命中就**在卡片上标出具体问题**；用户仍拿到 3 条完整话术，可以自己改。不做硬拦截，因为自检会误判——首轮回归就把「2022.06」当成了编造	**没做**「怎么改」的引导与「为什么标这条」的解释。原因是这一步的设计取决于用户看到标注时的真实反应，需要用户反馈才能定，而第三方用户证据为零，现在做等于猜
超时、解析失败、上游报错	首字超时与整体超时分开识别，用户看到的是「**生成超时，请重试**」；整份输出解析失败**重试一次**；单行解析失败**只丢那一行**、其余照常渲染；生成过程可随时中止；图片识别不可用时提示「直接粘贴 JD 文本不影响生成」	**没做**模型整体不可用时的自动降级（切备用模型或规则兜底）。规则兜底的质量已实测过，重复句比例 0.514，放出来会打脸「话术可以直接发送」这个承诺——宁可不给结果，也不给一个看着能用、实际不能用的结果。代价是用户只能等一会儿再来
成本	限流判断**先于**模型调用，超限不消耗额度；按 IP 每分钟 10 次 / 每日 100 次；全局单日上限 300 次由持久化计数器承载，跨实例与重新部署都不重置	**没做**「预算不足时自动切更便宜的模型」。换模型就是换输出质量，而质量是主指标——省下的钱会直接变成采纳率下降。控成本的做法是**限制次数**，不是**降低单次质量**

七、复盘：已知不足

- **用户侧证据缺失**：没有访谈、问卷或竞品实测，痛点来自推演。这是当前最大的短板。
- **评测样本是自造的**：8 组样本格式规整，没有真实简历的噪声（多栏、中英混排、排版混乱）。
- **核心价值指标不采集**：会话级复制率与建议采纳率的口径与本机导出已落地，但经本人决定不采集回流数据——价值维度因此降为观察项，结论保持「有条件通过」。
- **单位经济性只到 token 实测**：token 是实测值，但供应商单价未从账单核对，因此成本仍是公式而非结论。
- **简历与 JD 可能被模型服务商用于训练**：按 DeepSeek 隐私政策（2026/09 查阅原文），其在加密与去标识化前提下可能将输入及输出用于模型训练；政策给的退出开关只在它自己的对话产品里，API 调用没有对应入口，**本产品无法替用户关闭**。产品能做的是把风险讲清楚——生成前告知 +「数据与隐私」页 + 一键清除。
- **保真导出只覆盖 Word 来源**：PDF 与粘贴来源的用户只能拿到文本。
- **国内直连依赖 IPv6**：IPv4 出口到 Cloudflare 的部分网段被阻断，纯 IPv4 网络需要代理。

补齐顺序：第三方访谈与外部盲测经本人决定不做，作品集里如实声明；能继续推进的是把公开讨论材料整理出来（检索方案见 `docs/公开讨论调研_V0.1.md`），以及把已认领的四条取舍讲透。

八、我的角色

判断全部认领完毕，四件事已作答

2026/09/21 补上了此前空着的四件事：目标用户是**从本人痛点反推**出来的（优先选有量化标准的岗位， 先做容易做的部分，在最小成本下把产品做出来）；最先坚持的硬边界是 **「锁定区块交给本地代码」**（能交给代码的就交给代码，AI 生成有不确定性，AI 的产出还要交用户审核）； **「先建评测，再调提示词」是本人提的**，「判定规则只有一份」是工程侧的落地选择；实际参与的是提出起点、定硬边界、提出话术历史需求、提供样本、逐条验收、决定原地回填，以及决定上线与域名这七项。

「你的确认」两张表的 13 行也已逐条认领：**8 行本人决定**（含不做自动发送、保真导出只支持 Word 来源、 锁定区块交给本地代码、自检命中只标注不丢弃、原文件模板只留在内存、存档对不上就不导出）、 **5 行我认可这个建议**——分得清哪些是自己拍的板、哪些是接受的专业建议，比十三行全写「本人决定」更可信。

未参与的一项照实写成协作产出：**每日额度不是本人决定的**（上线与域名是）。清单只记录参与了什么， 没做的部分写明在旁，不留含义模糊的空格。

更细的证据链见仓库文档：作品集主文档 V1.3（三分钟读完）、六版 PRD 与六版技术方案增补、质量回归 V0.2、模型选型对比 V1.0、 项目评审标准 V0.2、上线前评审记录 V1.0、运维与回滚 V1.0。

在线可体验 | 557 项自动化测试 | 可回滚 | 数据不出本机