

大摩闭门会：Ignis Health 论人工智能在医疗领域的应用

20260526

IgnisHealth: AI 驱动医疗软件研发范式转移与商业模式重构

摘要

- 2026 年 1 月 LLM 性能实现阶跃式提升，模型幻觉与不可预测性显著降低，推动医疗后台应用从基础代码辅助转向根本性工作流重塑。
- 研发效率指数级增长：原定 2027 年的积压任务提前至 2026 年 Q2 完成；工程师转向“氛围编程”，可同时处理 10-15 项跨前后端任务。
- AI Agent 团队已实现 Jira 工单自动领取、计划制定、协作开发及自主部署，大幅削减传统开发中的沟通与测试环节。
- 核心护城河由“领域专长+专有流程”构成：通过自研数据标准化层解决医疗数据异构性，将业务逻辑与 AI 模型深度整合。
- Token 使用量激增：部分团队 Token 支出已达薪资成本的 40%-100%，公司视其为效率杠杆而非单纯成本，不因单价波动缩减投入。
- 商业模式面临颠覆：传统“软件即产品”模式过时，公司计划 2026 年秋季推行基于 Token 使用量或服务组合的新计费模式。
- 坚持不裁员策略：利用 AI 提升全员能力上限，通过缩短交付周期（重构代码从 2-3 年缩至数天）提升客户满意度与参与度。

Q&A

请介绍一下 Ignis Health 的业务概况？

Ignis Health 专注于远程医疗分析领域，自 2015 年起便开始运营。公司为美国大型医院系统内的远程医疗团队提供一套以分析工具为核心的软件应用，旨在提供业务、财务及临床方面的洞见。此外，公司还提供其他软件解决方案，以提升这些团队的效率和服务交付能力。公司的业务重点在于后台应用层面，支持高质量的医疗服务，但不直接涉及患者护理解决方案。

从 2025 年到 2026 年，大型语言模型供应商的能力经历了怎样的演变，贵公司又是如何利用这些 AI 能力为客户创造价值的？

从 2025 年到 2026 年初，LLM 技术对公司业务的实用性呈现指数级增长。公司从两个维度评估该技术：一是内部应用以提升自研应用的开发效率，二是如何利用该技术为客户提供更优质的产品。在 2025 年，尽管公司大力推动团队进行开发，但直至该年第四季度初，多数尝试均未成功，主要面临幻觉问题和结果不可预测性等挑战，团队和试用客户均提出了大量反对意见。然而，到 2026 年 1 月，情况发生显著变化，模型性能大幅提升，团队对该技术的热情高涨，产品的使用量、使用率和易用性均呈指数级增长。这种阶跃式变化体现在信任度和质量水平的提升上。在软件工程方面，工程师使用的工具也发生了转变。2025 年，工程师更倾向于使用在现有 IDE 平台上增加 AI 功能的工具，如 Cursor，以加速编写基础代码。但进入 2026 年 1 月，随着模型性能和用户友好度的显著提高，工程师开始转向能从根本上改变工作方式的应用，例如 Claude Code。

从 2025 年到 2026 年，贵公司团队在应用 AI 方面的学习曲线呈现出怎样的变化？

团队的学习曲线可以分为两个阶段。第一个阶段是 2025 年全年直至 2026 年 1 月第三周左右，这一时期的学习曲线非常陡峭。尽管 AI 工具被宣传为自主化且输出可预测，但实际应用中存在大量幻觉问题。鉴于公司所处的医疗领域对数据处理有极高要求，并受 HIPAA 法规的严格约束，处理包含公共健康信息的数据时必须极其谨慎，这使得在 2025 年操作这些工具的难度很大。第二个阶段从 2026 年第一季度开始，学习曲线在某种意义上更为陡峭，但其挑战已从工具的使用层面转向如何有效管理和引导 AI 所赋予的巨大创造力。团队效率大幅提升，原计划于 2027 年处理的积压任务，现已在 2026 年第二季度前完成。同时，AI 的强大能力也促使团队着手开发一些从未设想过的项目，功能的深度和广度均在变化。工程师的工作内容正发生根本性改变，从编写代码转向了所谓的“氛围编程”，能够同时处理十到十五项任务，并能跨越前后端界限，消除了大量工作障碍和不必要的会议。

在利用 AI 为客户创造价值方面，贵公司取得了哪些具体的成功案例？

公司组建了专门的 AI 团队，该团队能像真人一样在 Jira 等工单管理系统中运作，自动领取工单、制定计划、将任务分配给不同的 AI 智能体，并自主协作完成任务、整合发布版本、执行质量检测，甚至在部分场景下独立完成部署。这种高度自主化的流程摒弃了传统开发模式中大量的前后端沟通、交付确认和反复测试环节，显著提升了效率。对客户而言，最直观的感受是公司响应速度的提升，例如软件故障的修复效率。此外，原计划于 2026 年晚些时候或 2027 年年中交付的功能得以提前兑现，客户对此极为满意。公司不仅交付了承诺的功能，还利用 AI 带来的创造力为产品增添了独特的亮点，在某些情况下甚至实现了软件的自我迭代。目前，公司正与每位客户进行一对一沟通，展示新的工作方式和成果。客户的参与度非常积极，他们看到了产品功能的持续演进和未来的潜力，并开始重新

规划自身流程，计划借助 Ignis Health 将大量琐碎、耗时的任务（如重复的数据录入）完全自动化。远程医疗的运营流程极其复杂，涉及临床、IT、合同、许可认证等多个跨组织部门的协调，而 AI 技术使公司能够更有效地协同各方，为客户创造巨大价值。

贵公司如何将专有数据与 AI 模型进行整合，具体涉及哪些供应商，以及数据基础设施是如何与 AI 模型关联的？

公司处理的数据实际上归客户（即医疗系统及其患者）所有，公司本身并不拥有这些数据，而是拥有处理这些数据的流程以及叠加其上的业务逻辑。公司的核心愿景是为医疗管理机构和临床医生打造类似彭博终端的应用，以解决他们在远程医疗运营中缺乏可见性的问题。由于医疗数据领域的特殊性，每家医疗系统甚至同一科室内部的术语和操作流程都存在差异，因此公司自主开发了工具来应对这些挑战。公司构建了自有的数据标准化层，并在数据基础上叠加了大量的流程知识，确保应用具备足够的灵活性，以适应不同用户的术语习惯，同时保持数据本质的一致性。最终，公司开发了一套专有流程，特别是在数据归一化、标准化以及为不同服务线叠加特定运营知识方面，形成了独特的解决方案。

从 2025 年的阻力到 2026 年的突破，除了模型质量和准确性的阶跃式提升外，还有哪些关键因素促成了这一转变？

模型质量和准确性的阶跃式提升确实是关键的变化。在医疗健康领域，许多终端用户，尤其是医生，过去对技术（如各类转录或自动化数据录入服务）的体验并不理想，他们常常觉得这些技术让工作变得更复杂而非更简单。然而，当前 AI 技术的发展带来了新的乐观前景，有望使医疗行业再次成为一个令人兴奋领域。

回顾 2025 年，除了模型本身的技术局限外，团队对终端用户体验及实际工作流程的设想是否与预期存在偏差，导致一些原本设想的流程演进方向未能实现？

这确实产生了一定影响。最初的思路是直接为工程师团队开放工具权限并推广某个特定应用和流程，但很快发现此路不通。随后策略调整为提供多种工具供团队自行尝试，以寻找最适合的解决方案。尽管模型技术在进步，但部分工具的用户体验仍有待提升，导致团队使用不便。从产品角度看，这引发了对产品形态的根本性思考：它是否还能被称为传统的软件产品。产品的定义正在发生根本性变化，我们和客户都在适应这一过程，同时努力调整市场预期，避免描绘过于理想化或悲观的场景，以呈现更为现实的情景。

考虑到 2026 年软件开发周期已显著加快，贵公司作为一家自力更生的企业，是如何考量 AI 开发与投资的投资回报，尤其是在代币等方面的支出？

随着新的软件开发方式日趋成熟，尤其是在 2026 年，采纳并调整代码开发方式已是必然选择。作为一家无外部投资、无负债的公司，我们对支出极为审慎，不

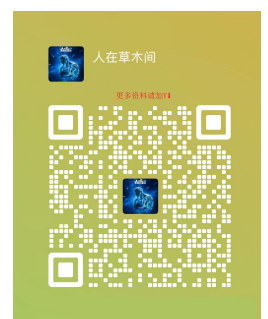
会盲目投入资金解决问题，每一笔投资都必须经过系统化的量化评估。这种方法论与行业普遍做法，尤其是在硅谷，有所不同。在预算方面，我们非常谨慎。例如，在员工培训上，我们会制定内部使用指南和最佳实践标准，而非放任自流，以避免产生反效果。软件行业存在“技术债务”概念，即旧代码需要长期维护。过去，优秀的团队重构或重写代码需要两到三年；现在，若方法得当，可能仅需一周甚至几天。但这也可能导致代码生成失控，形成新的问题。因此，我们看待 token 使用的目标并非无限制消费，这主要基于两点考量：一是预算控制；二是为了避免代码量膨胀到人力无法维护的程度，从而过度依赖 AI 工具，陷入恶性循环。

摩根士丹利技术团队预测，代币补贴时代可能即将结束，未来成本或将边际上升。这一预期将如何影响贵公司当前的投资策略，尤其是在利用工具探索可能性边界的实验性投入方面？

我们基本不考虑这个问题，因为这种视角过于短浅。我们真正关注的是 AI 带来的可能性、效率提升、开发与产品上市的加速，以及创造力的释放，目标是成为下一个引领全新设计的分析解决方案的提供者。我们认为，那种销售软件后再由客户聘请人员进行定制的传统模式已过时。我们一直鼓励团队积极尝试所有可用工具，并通过每周多次的内部会议，分享探索过程中的有效与无效经验。公司内部设有专门用于试点项目（即“moonshot”项目）的预算、时间和工作量，实验始终在进行。因此，关于因成本上升而放缓脚步的假设，对我们而言并不成立，这甚至不在我们关注的前一百个事项之列。

随着 AI 代理团队开始负责从开发、质量保证到安全等软件开发环节，这对贵公司在软件开发和技术工程师方面的招聘策略产生了何种影响，以及未来的预期是怎样的？

至今为止，我们没有因为效率提升而裁减任何员工。裁员是一种落后且短视的做法。团队成员都非常优秀，既然公司有充足的资金储备，我们选择为每位成员投入资源，旨在实现整个团队能力的指数级提升。这一策略已持续了四五个月，并且依然有效。核心思路是必须在人才和技术上持续投入，并耐心等待成果。这一理念也延伸至与客户的沟通中。当我们向客户介绍那些非可视化的能力提升时，例如更具创意的方案或解决复杂问题的能力，他们往往难以直观理解，必须亲眼看到成果。我们告知客户，项目推进速度已大幅加快，原计划 3 至 12 个月后启动的项目现已在进行中。然而，客户听完后常常只关注到开发成本的下降，并询问是否可以削减开支。这是一种误解。实际上，我们现阶段的成本反而更高，因为既要支付人力成本，也要承担 AI 相关费用。或许十年后情况会有所不同，但当前正处于需要加倍投入的阶段，不仅要加大投资，还要彻底重塑工作方式与产品服务。



行业普遍观察到，在医疗领域，数据的价值相对于其他受益于 AI 甚至被 AI 颠覆的行业而言被低估了。您如何看待这一现象？

我同意数据在医疗行业被严重低估的观点，但这背后有其原因。首先，医疗是一个高度监管的行业，涉及敏感的患者数据，操作无法像金融等行业那样灵活。其次，医疗行业早期未能投资于合适的数据基础设施，导致许多决策依赖于定性信息、轶事证据或咨询顾问的个案分析，缺乏基于数据的系统化方法。我们之所以进入这个领域，正是因为它高度监管且极其复杂。即使在同一医疗系统内，不同部门的运作方式也差异巨大，这既反映了医疗服务的复杂性，也暴露了数据基础设施的落后。直到 2026 年，我们仍在与一些医疗系统讨论从本地部署迁移至云端的话题，这说明该领域的技术采纳速度非常缓慢。因此，数据定价过低有多重原因，任何新的服务或应用都必须与现有产品有根本性差异，才能真正带来改变。正如一位身兼医生与商人的朋友所说：在医疗领域，唯一能快速传播的只有病毒。任何应用或服务的市场渗透和影响力显现都需要时间。

AI 技术正对商业模式产生颠覆性影响。您认为 AI 将如何影响各类软件的商业模式？从长远来看，什么样的因素能构成持久的护城河或竞争优势？

“软件即产品”的概念正在被颠覆，我们向客户提供的产品形态正发生显著变化，商业模式也即将迎来大幅调整。目前我们尚未调整商业模式，但已开始与部分客户合作探讨新的解决方案，并计划在 2026 年秋季与大多数客户共同商议未来的合作策略：是加大投入还是节约成本。这不仅是预算问题，更关乎企业竞争力。在竞争对手全力推进 AI 的背景下，不进行投资意味着落后。我们希望成为首批探索新业务模式形态的企业，其基础可能基于 token 使用量、服务与软件的组合，或是服务与 token 使用量的结合。目前判断具体形态为时尚早，我们正审慎地建立流程，评估投资回报的量化与质化、财务与非财务指标。我们与客户的已从简单的供需模式，演变成为一种更深度的协作关系。至于竞争优势，我认为适用于所有公司的核心竞争力在于领域专长（domain expertise）。例如，我们团队的工程师能够与医生进行深度交流并理解其真实需求，从而快速提供精准的解决方案。深刻理解特定行业（如医疗、金融、机器人技术）的专业知识，将成为企业的关键优势。正如彼得·蒂尔所言，成功在于选择最小的细分领域并深耕。我们自 2015 年起就专注于一个细分领域。我们的优势在于能够消化客户提出的复杂问题，并将其转化为有效的产品或服务解决方案。这种由医疗服务提供商、工程师和 AI 技术三者结合而成的能力，构成了我们强大而持久的护城河。

您如何看待贵公司业务中代币需求的未来发展趋势？

我预见公司业务和 token 使用量将出现飙升。通过采取正确的措施，公司将持续增长，增长曲线可能非常陡峭，以至于我们无需过度关注利润底线。许多普通用户仍将 ChatGPT 等工具当作搜索引擎使用，并未发掘其强大潜力，且倾向于使用免费版本，从而错失了最新、最强大的模型。在复杂应用场景下，模型的选择至关重要。因此，无论是在软件开发还是其他领域，token 的使用量都将持续增长。

我观察到团队的 token 利用率从零开始，一路攀升至薪资成本的 10%、30%、40%，部分团队甚至超过 100%。这种增长速度有时令人不安，但只要使用方式得当，这种增长就是合理的。我们不会遇到使用量下降的情况，当续约出现问题导致工具无法访问时，团队的反应足以证明其对工作的重要性。因此，我不担心业务成本问题，使用量的增长趋势是确定且不会改变的。

