



案例研究: AI 内容优化与创作

GenAI 驱动的内容优化和本地化试验

4 款 AI 工具参与测试

编写和测试 2 种语言的提示

使用英语提示优化的德语内容的平均质量评分: 86%

LIONBRIDGE 开展本次试验的背景

Lionbridge 使用内部团队、资源和内容进行了此次试验。具体来说,我们使用了内部德语内容:现有的博客文章。尽管聚焦 Lionbridge 内容(具体为德语内容),但本案例研究旨在确立规范和获得研究结果,借此为我们覆盖各行各业和各种语言的广大客户提供更好的服务。我们将运用本次试验的研究结果和观察所得,满足客户对各种内容和语言的需求。这些信息还将应用于我们自己的内部营销内容创作和优化流程。

挑战

优化内容(尤其是非本地语言)内容,可能既费时又耗财。而且翻译现有内容有时无法达到出色的 SEO 效果,因为翻译有时只是简单地直接转换语言,而没有体现出语言的微妙区别以及因不同地区而异的关键词、用户行为和/或数据。GenAI 工具或许可以帮助我们优化多语言内容,同时省时省力省成本。

背景

四款生成式人工智能 (GenAI) 工具: Amazon Bedrock、ChatGPT (Lionbridge Azure 实例使用 ChatGPT 4)、Google Bard 和 Meta 开发的 Llama, 都可用于创作和优化内容。

它们可以分析搜索模式、竞争者数据和用户行为, 以提供基于数据的关键词和内容建议。

这些工具可帮助确保优化后的内容富有吸引力、具有强大的搜索引擎优化 (SEO) 价值且尽可能契合用户期望, 在某些情况下甚至还能生成优于人类创作水平的所需文本。

例如, 在需要按照严格的字符限制创作和翻译内容时, GenAI 工具的水平 (尤其是在 SEO 方面) 有时可以匹敌人类。

解决方案

我们对 GenAI 工具进行了三阶段试验,以探索哪个工具的搜索和内容创作(尤其是非本地语言的搜索和内容创作)效果最佳。我们的团队重点研究了生成提示(尤其是针对多语言内容)的优秀佳实践。我们还探索了以新语言优化已有内容,从而提升网站 SEO 的优秀实践。

方法

Lionbridge 使用了四款 AI 工具进行此项研究: Amazon Bedrock、ChatGPT (Lionbridge Azure 实例使用 ChatGPT 4)、Google Bard 和 Meta 开发的 Llama。我们分别对以下三个阶段进行了两次试验: 搜索、创作和分析。

搜索: 在该阶段,我们命令每个 AI 工具完成关键词搜索,并选出两个最适合创作德语内容的问题。

我们向这些工具馈送了以下各项内容:

- 角色提示,要求工具以德语内容编辑的角色,“提供 3 个新的德语关键词和 3 个德语问题”。
- 搜寻需要优化的内容
- 现有关键词

此外,我们还选择了两个补充问题,用于后续生成小篇幅的内容段落。

在关键词搜索方面,人工介入(通常称为“人机回圈”)似乎非常重要,对 Bard 而言尤其如此。这款工具的搜索表现十分出色,但提供的选项太多,需要主题专家选择内容创作所需的关键词或问题。



创作: 在该阶段, 我们生成了两组德语提示。这些内容创作提示中包含上一步确定的关键词。其中一组提示基于内容编辑的角色而生成, 另一组用于任务指示。

我们要求 Lionbridge AI 部门人员审核并帮助修改所有提示, 以充分提高其有效性。最终敲定提示后 (在下面的“提示”部分列出), 我们使用 Lionbridge 资源将提示本地化成德语。我们将这两组已完成的德语提示应用到所有四款 AI 工具中, 并收集了工具建议的内容并对其进行了打分。

在生成提示时, 我们输入了以下内容:

- 在搜索阶段确定的问题和关键词 (德语)
- 我们需要模仿的优质内容示例 (德语)
- 源文章的文本 (德语)
- 角色提示 (英语和德语)
- 任务提示 (英语和德语)

我们有两组提示, 因此在每个 AI 工具中都需要将项目运行两次。

分析: 在该阶段, 我们请十位德语商务和语言专家对内容进行了审核。为确保“盲审”, 我们没有标记这些内容是由哪一个工具生成的。

我们将输出添加到电子表格以进行评价, 并使用了两种评分方式:

- 含有三个值的下拉菜单: “不能使用”、“编辑后可以使用”和“无需编辑, 可以直接使用”
- 内容评分字段, 分值范围为 1-100

通过审核收集而来的反馈, 以确定:

- 哪种工具效果最佳?
- 生成提示的优秀实践有哪些? 以其他语言生成提示的表现如何?
- 这些工具在关键词搜索方面的表现如何?
- 这些工具在生成有关内容的问题方面表现如何?
- 这些工具在生成文章视角方面表现如何?
- 这些工具在回答其生成的问题方面表现如何?
- 如何确保实施最佳的 AI 内容创作流程?
- 如何生成有效的提示?

规划



选出 3 款要使用的
GenAI 工具



选出 3 篇文章进行
优化试验



寻找 10 位专家组成评
审小组



通过搜索确定关
键词和问题



生成提示并本地
化成德语

使用的提示示例

角色提示

- 您是 Lionbridge 的内容编辑,负责创作与语言服务相关的内容。
- 您会德语,住在德国。
- 您创作德语内容,目标读者是德国、奥地利和瑞士等国家/地区的德语人士。

受众提示

- 市场营销或传媒人员
- 高级岗位人员、高级经理、总监、副总裁、总裁、首席营销官、首席执行官
- 德语人士,住在德国、奥地利和瑞士等国家/地区

创作提示

- 请额外创作两段文本附加到这篇文章...
- 每一段都应回答以下问题...
- URL 的主要关键词是...
- 次要关键词是...

多学科团队

- 10 位会德语的提示工程师
- 1 位 Lionbridge SEO 专家
- 1 位 Lionbridge 数字营销专家



试验结果

AI 应该在各种优化项目中担任重要角色。

我们的试验结果证明, AI 工具确实有助于有效且高效地优化和更新现有内容,甚至包括除英语之外的其他语言的内容。

可喜的是,随着试验的进行,输出的质量也逐步提高。人工介入在内容优化或创作过程中非常重要。必须要由人工从 AI 输出中选择关键词进行优化。此外还需要通过人工审核来确认最终的内容输出(随着输出质量显著提高,这一项在未来可能会有所变化)。

下面列举了我们确定的一些优秀实践:

- 目前必须通过人工介入来确认最终的 AI 输出。
- 在试验过程中,随着我们不断学习到新的优秀实践,输入的质量也逐步提高。
- 当提交的请求具有以下特征时, AI 更有可能创作出优质的内容:
 - 篇幅较短
 - 内容可以轻松分解(例如,前十榜单等等)
- 目前,使用英语与 AI 技术沟通最为有效(这很好理解,因为这些工具是用英语训练而成的)。提供德语提示并不能帮助 AI 工具生成更优质的德语内容。值得注意的是,这些工具根据英语提示生成的德语内容更加优质。注意: AI 工具在以目标语言(非英语)提供源内容、关键词、问题和示例方面仍然会有所帮助。
- 在内容创作流程中,应该对内容创作者和目标受众进行充分描述。此外,最好建立提示数据库,这样用户就能重复利用提示并为客户创作一致的内容。

搜索输出质量最佳的应该是 Bard。

其搜索结果与专用的搜索工具相当 (例如 MarketMuse)。值得注意的是, Bard 提供的搜索输出需要由人工选出有价值的片段进行内容创作。Bard 使用德语指示时, 在这两个阶段的表现比较稳定。但使用英语指示时得到的评分结果为 82/100 到 66/100。这个数据集的体量较小, 因此 Bard (英语) 的平均得分是 74/100。

ChatGPT 的内容收到的负面评价稍多, 但其内容有时也让评审员大为赞叹。Meta (Llama) 的表现也不尽人意, 可能还没有准备好可靠地优化或创作非英语内容。本次小项目为期一个月, 在项目推进过程中, 输出的质量似乎在逐步提高。Bard 的搜索输出质量有所提高, ChatGPT 的德语内容创作质量有所提高。ChatGPT 根据德语指示得到的输出最初的得分只有 47/100,

但短短一个月后, 其输出得分便提高到了 76/100。根据英语指示生成的输出得分最初是 71/100, 一个月后提高到了 86/100 (英语的平均得分是 78.5/100)。

Bard 和 ChatGPT 在生成短篇幅内容方面 (例如, 只有一两个句子) 的表现都十分出色, 因此在这方面得到的负面反馈极少。

这里显示的评分是在两个试验阶段中针对长篇幅内容的得分。

阶段 1 评分

来源	不能使用	编辑后可以使用	无需编辑, 可以直接使用	平均得分
ChatGPT (德语输入)	9	9	0	47.61
Bard (德语输入)	2	14	2	68.33
ChatGPT (英语输入)	3	9	6	71.67
Bard (英语输入)	0	9	9	82.22
Meta (Llama) (英语输入)	7	8	3	46.50
总分	21	49	20	

阶段 2 评分

来源	不能使用	编辑后可以使用	无需编辑, 可以直接使用	平均得分
ChatGPT (德语输入)	1	7	0	76.25
Bard (德语输入)	2	4	2	66.25
ChatGPT (英语输入)	0	4	4	86.875
Bard (英语输入)	3	4	1	59.375
总分	5	19	7	

注: Meta (Llama) 未参与阶段 2 试验, 因为该工具尚不足以优化或生成非英语内容。



更多详情, 请访问

LIONBRIDGE.COM