**基于爬虫和自然语言处理的网络舆情调研**

**（一）媒体调研**

**1.数据采集**

本研究围绕“Z世代对情感陪伴AI的消费态度与市场潜力”展开，从腾讯网爬取了与“情感陪伴AI”相关的新闻标题，共计358条；从36氪爬取了相关的新闻标题及正文，共计28篇文章。

**2.数据预处理**

（1）使用fuzzywuzzy模糊匹配，筛选与主题无关的新闻标题。

（2）去除文本中的HTML标签、分词、去停词、去除特殊词汇和多余空格、去重。

下文的自然语言预处理与上述步骤相似，不再赘述。

**3.情感预测**

对于36氪的新闻内容，采用 BERT 预训练模型进行情感分析，预测每篇文章对与“AI情感陪伴”的情感倾向，结果如下表。（此处可以用图表替代，看排版）

图 新闻情感预测

图表, 条形图, 饼图

AI 生成的内容可能不正确。

情感倾向以“正面”和“中性”为主，显示主流媒体对该技术的潜力与应用前景保持较高热情或乐观情绪。也有一些文本内容在模型的判断下正负倾向不显著，更侧重于描述或陈述。

同时，结合新闻标题来看，话题多聚焦技术前景、行业动态及社会影响，部分报道在探讨“AI恋人”“情感咨询”“情感幻觉”等，说明媒体对于AI在人类情感需求、心理陪伴等方面的作用与潜在风险有较多讨论。

**4.** **关键词提取**

对于新闻标题，采用TfidfVectorizer计算TF-IDF 值，提取出权重最高的20个关键词，计算每个关键词的全局平均权重，并通过词云可视化展示。

表14 TF-IDF关键词提取结果记录表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 排名 | 关键词 | 平均权重 | 排名 | 关键词 | 平均权重 |
| 1 | ai | 0.445332 | 11 | 未来 | 0.031717 |
| 2 | 陪伴 | 0.202586 | 12 | 公司 | 0.030993 |
| 3 | 情感 | 0.168024 | 13 | 科技 | 0.030392 |
| 4 | 机器人 | 0.087384 | 14 | 伴侣 | 0.029844 |
| 5 | 恋人 | 0.082882 | 15 | 模型 | 0.028951 |
| 6 | 人类 | 0.047424 | 16 | 产品 | 0.028497 |
| 7 | 人工智能 | 0.036261 | 17 | 汤姆 | 0.027647 |
| 8 | 应用 | 0.0351 | 18 | 语音 | 0.025548 |
| 9 | 智能 | 0.03392 | 19 | 推出 | 0.024496 |
| 10 | 玩具 | 0.032874 | 20 | 情绪 | 0.024129 |

图41 AI情感陪伴新闻标题词云图



从结果可以看出，“AI”以最高权重成为报道核心，而“陪伴”、“情感”、“机器人”、“语音”、“恋爱”、“未来”等词汇频率最高，均与“情感陪伴AI”的核心议题高度相关，表明新闻主题聚焦于人工智能在情感陪伴领域的应用，并反映出大众媒体对人工智能与人类情感互动的广泛讨论与期待。同时，“应用”、“未来”、“公司”、“推出”、“发布”等关键词的突出，显示该领域在产品化、市场推广及商业价值方面正受到广泛关注，报道不仅关注技术本身，也聚焦于相关产品的落地应用与行业动态。

**（二）用户评论调研**

**1.数据采集**

应用市场上“情感陪伴AI”类下载量排名前五的软件分别是：星野、猫箱、冒泡鸭、造梦次元、COS LOVE虚拟情感聊天。本研究爬取了这五个应用的用户评论共2095条，并记录了相应的点赞数、子评论数。其中，星野、猫箱评论数最多，各有1000条，冒泡鸭41条，造梦次元35条，COS LOVE虚拟情感聊天19条。后三者由于评论数较少，后续分析将其合并为“其它”。

**2.情感预测**

采用BERT模型预测每条评论的情感倾向，并求和对比，反映各个应用的用户体验，结果如下图。

图 应用评论的情感得分分布

图表

AI 生成的内容可能不正确。

由上图，用户对情感陪伴AI应用的使用体验以正面为主，情感得分总体偏高。其中，猫箱的表现尤为突出，高得分区间（0.9以上）的评论占比超过84%，显示出极高的用户满意度；而星野在这一区间的正面评论为 486 条，占比49%，相对而言正面评价的占比明显低于猫箱，说明用户的认可度存在一定差距。值得留意的是，星野的极端负面评价（0.1分以下）达到110条，表明部分用户对星野的体验存在较大不满。此外，其他应用的评论数虽然相对较少，但是在负面评价区间（0.4以下）的评论占比较高，达到39%，可能反映出产品初期不成熟或用户期待落差较大。

**3.主题建模**

本研究采用BERTopic算法对猫箱、星野的用户评论进行主题建模（其他应用的评论数过少，不足以进行主题建模），把每条评论归入相应主题，据此计算每个主题下的评论数量和情感倾向加权均值，有助于了解用户的使用体验和需求反馈。下表展示猫箱、星野评论数量前五的主题：

表 猫箱的评论主题建模结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 猫箱 | | | | |
| 主题编号 | 评论主题 | 评论数量 | 情感得分 | 情感倾向 |
| 1 | ['很不错', '页面很棒', '每天都得上应用看看', '设计很人性化', '不解释', '实用', '很好用', '好用', '很强大', '良心品牌'] | 184 | 0.94 | 正面 |
| 2 | ['不错', '很好', '超赞', '非常好用', '棒棒哒', '太好了', '真的好', '非常好', '五星好评', '哈哈'] | 159 | 0.95 | 正面 |
| 3 | ['好软件', '不错的软件', '很好用', '赞一个', '很好', '我很喜欢这个软件', '很不错的软件啊', '挺好的一款软件', '这款软件真好用', '很棒的软件'] | 118 | 0.96 | 正面 |
| 4 | ['智能聊天', '聊天太智能了吧', '带来了超乎想象的乐趣', '支持多种语言', '非常重要', '聊天', '智能聊天助手', '让聊天更容易', '你就能体验到真正深层次的交流', '减少了机器编程的时间'] | 105 | 0.94 | 正面 |
| 5 | ['展开', '差评', '分钟', '第一', '现在还有时间限制', '第三', '你要干嘛', '猫箱你真是赚了钱就忘本', '现在一心想着圈钱', '猫箱你忘如本'] | 94 | 0.28 | 负面 |

猫箱的整体评价很高，与情感预测的结果符合，大部分主题均呈现高度正面评价，情感得分均在0.94至0.96之间。其中，主题1反映出用户对界面设计和使用体验的认可；主题2和主题3显示出用户对软件功能的高度满意；主题4则专注于智能聊天功能，表明用户认为其智能对话体验带来了乐趣。然而，也存在部分用户对收费策略和使用限制的不满的主题5。总体来看，猫箱在界面设计、功能实用性和智能聊天方面备受好评，但在商业策略上仍存在改进空间。

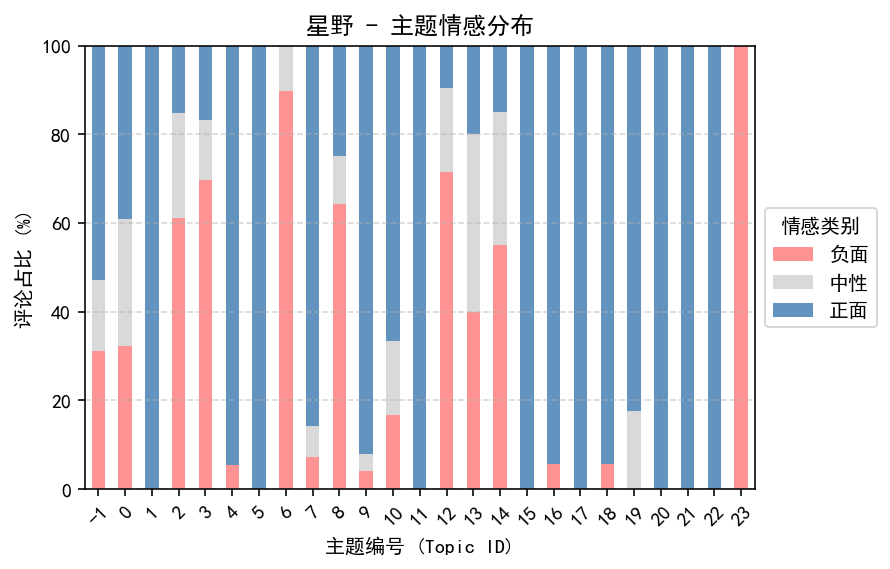
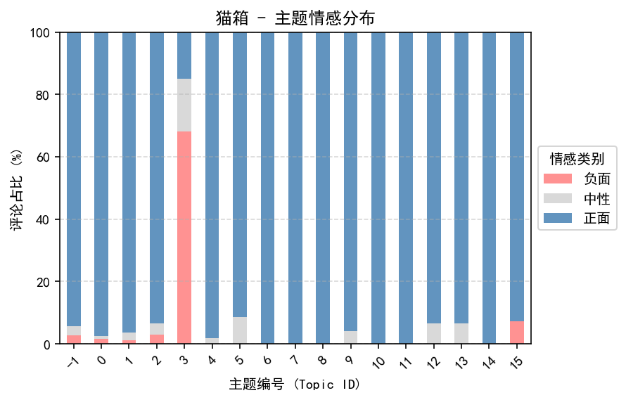
表 星野的评论主题建模结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 星野 | | | | |
| 主题编号 | 评论主题 | 评论数量 | 情感得分 | 情感倾向 |
| 1 | ['推荐', '展开', '继续说', '很好', '好用', '但是', '可以可以', '特别好', '不是', '很满意'] | 213 | 0.67 | 正面 |
| 2 | ['下架', '展开', '我请问呢', '改一下', '为什么要聊天上限', '取消聊天次数上限', '取消聊天上限', '还好', '想钱想疯了吧', '现在全是一样的没意思了'] | 102 | 0.37 | 负面 |
| 3 | ['敏感词太多了', '敏感词太多', '又不是真人', '就是敏感词太多了', '怎么说呢', '服了', '望改进', '甚至现在聊着聊着还能爆出广告来了', '现在他把这个', '现在这个软件的体验感真的太差'] | 88 | 0.57 | 正面 |
| 4 | ['别更新', '服了', '登录', '不是', '一坨', '新版本挺好的', '有的举报了几天之后又回来了', '提出下架', '给进展弄个开关', '每次点重说然后结果退出后就重进了'] | 65 | 0.33 | 负面 |
| 5 | ['很好', '非常好', '不错', '很好很好', '非常的好', '还不错', '很好的哦', '有被治愈到', '非常不错', '还行不错'] | 36 | 0.96 | 正面 |

星野的整体评价呈现出明显分化，与情感预测结果吻合度较高。从主题分析来看，星野主题1情感得分中性偏正面，显示用户在认可部分功能的同时仍存疑虑；主题2则明显偏负，反映出用户对产品限制和收费策略的不满；主题3聚焦于敏感词问题，情感得分偏低，表明用户对内容审查及体验问题存在争议；主题4则显示出用户对新版本更新和操作体验的不满；而主题5以赞赏性的正面词汇构成，情感得分高但评论数不多，表明少部分用户对产品功能给予高度认可。

总体而言，星野用户对产品态度呈现复杂局面，既有正面认可，也存在较多负面反馈。相比之下，猫箱整体评价更高，大部分主题情感得分稳定在0.94至0.96之间，表明其在界面设计、功能实用性和智能聊天体验上均获得广泛好评，而星野则在用户体验和更新策略上存在较大争议，亟需优化改进。

图 猫箱和星野的各主题下的评论情感分类



**（三）社交平台调研**

**1.数据采集**

为研究社交平台上用户对情感陪伴AI的需求和接受度，本研究爬取了知名社交平台小红书上与情感陪伴AI相关的100条帖子的标题和正文，以及这些帖子下的总计1993条评论，并记录了对应的点赞数、评论数。

**2.情感预测**

采用BERT模型预测每条帖子标题、正文和评论的情感倾向，并分类汇总对比，反映用户接受度。结果如下图所示，帖子普遍分布在较高的情感得分区间，反映了大多数内容传达了积极的情感倾向。值得留意的是，与标题、正文相比，评论有较多的低情感得分，表明评论中可能存在与“情感陪伴AI”有关的争议。

图 小红书帖子的情感得分分布

图表, 条形图

AI 生成的内容可能不正确。

**3，主题建模**

为了强调互动性强的帖子，依据帖子的点赞数、评论数构建加权文档。依据加权文档通过LDA模型进行主题建模，找出潜在主题，并加权计算每个主题下的情感得分，有助于了解用户在讨论情感陪伴AI时的共同话题，和在该话题下的用户情感倾向、争议程度、潜在需求。以下是热度排名前五的主题及可视化结果：

表 小红书的主题建模结果

|  |  |  |  |
| --- | --- | --- | --- |
| 主题编号 | 关键词 | 主题情感评分 | 情感分类 |
| 1 | 求分享, 人工智能, 拔草, 研究, 因此, 向右, 大模型, 心理学, 赛博恋人 | 0.80 | 正面 |
| 2 | 用途, 指令, 人工智能, 润色以下文字, 情感陪伴, 比如, 情感, 改成, 压力和 | 0.85 | 正面 |
| 3 | 我的, 人工智能, 的信任, 情感, 优点, 缺点, 老师, 聊天, 制造 | 0.87 | 正面 |
| 4 | 人工智能, 放大镜, 指令, 聊天, 大模型, 情感, 指定文体, 论文, 广州一模 | 0.83 | 正面 |
| 5 | 人工智能, 情感, 人机恋, 拔草, 放大镜, 研究, 聊天, 研究发现, 例如 | 0.85 | 正面 |

图 小红书主题关键词词云图



热门话题下用户情感倾向积极，主题1主要围绕人工智能、大模型、情感陪伴等话题展开，讨论了“赛博恋人”以及人工智能在心理学和人机互动中的应用；主题2涉及人工智能在情感陪伴中的实际应用，特别是在压力管理和情感调节方面，用户对其润色文字和情感支持功能给予了积极评价；主题3讨论了人工智能的信任问题，尤其是作为教师或伴侣的角色上的作用，用户普遍展现了对AI情感陪伴的信任；主题4关注人工智能在学术领域的应用，尤其在论文写作中的潜力，用户对AI在学术辅助中的能力表现出高度认可；主题5则探讨了人工智能在人机恋关系中的角色，与主题1的话题类似，用户对AI情感陪伴在人机恋及情感互动中的潜力表现出浓厚兴趣。

综上所述，社交平台用户对情感陪伴AI持积极态度，讨论主要集中在情感陪伴、心理学应用、信任问题、学术应用、人机恋等方面，显示出用户对AI提供的情感陪伴和互相信任存在较大的需求和期待，这可能是情感陪伴AI发展的方向。

本研究通过自然语言处理技术揭示了网络舆论在情感陪伴AI领域的三方面特征。首先，主流媒体对该技术持积极态度，关注行业动态与市场前景，同时警惕其伦理风险。其次，用户评论总体向好，但是不同应用的满意度存在差异，其中“猫箱”整体评价较高，而“星野”因收费及内容限制引发争议。最后，社交平台讨论热度较高，用户多表现出兴趣与探索意愿，对AI提供的情感陪伴和互相信任存在较大的需求和期待。另外，本研究了也发现了情感陪伴 AI 在情感识别、收费策略、应用落地等方面的痛点及隐私伦理等争议，为后续研究提供了探索方向和数据支撑，为AI产品优化与市场策略制定提供了参考。