

AI 能学会 😊 吗？——大语言模型 Emoji 与文本情绪一致性探究

摘要：“😊、😊、💪、🎉”均为网络沟通中常见的“emoji”，又称“表情符号”。当前研究发现，emoji 与文本间的情绪一致性和情感搭配策略，是人类使用 emoji 的重要特征。当前，生成式人工智能（Gen AI）高速发展下，大语言模型也被证实能够模仿人类理解、使用 emoji。然而，当前关注大语言模型 emoji 使用的研究仍较少从 emoji 与文本情绪的一致性出发，调查模型对 emoji 情感使用的理解与学习。因此，本研究关注大语言模型在 emoji 情感使用上与文本情绪的一致性特征，探讨大语言模型 emoji 使用的情感偏好与情绪策略。基于对文本、emoji 的情感分析，研究发现：大语言模型所用 emoji 以积极情感为主，常为积极或中性。积极意义常通过表达喜悦、友好、关爱的 emoji 传达，中性意义常通过映射客观现实事物的 emoji 传达；同时，大语言模型在绝大多数情况下所用 emoji 与文本情感保持一致，在少数情况下所用 emoji 与文本情感冲突。两者一致时，模型常用 emoji 来阐释或增强当前主题、情感表达。两者冲突时，模型常用积极 emoji 调节负面文本语气。这些发现能导向大语言模型与人类在 emoji 理解、使用上的相似性。本研究进一步提出了相应的理论、实践启示，并讨论了现存局限。

关键词：Emoji；表情符号；情感分析；大语言模型；人工智能；AI

1. 引言

“Emoji”，又称“表情符号”（如😊、😊、💪、🎉），是用户在网络沟通中传达非语言信息的重要手段。当前研究（Novak, 2015; Sampietro, 2019; Caspi & Raz, 2025）对 emoji 使用的情感价值展开了探讨，并发现 emoji 与文本间的情绪一致性和情感搭配策略，是人类使用 emoji 的重要特征。同时，当前生成式人工智能（Gen AI），特别是大语言模型的 emoji 使用能力是一项研究焦点（Dunn & Hopkinson, 2024; Jahan et al., 2024; Zhou et al., 2024）。然而，目前大语言模型 emoji 使用研究仍较少从 emoji 与文本情绪的一致性角度出发，探讨模型对 emoji 情感的理解与使用。因此，本研究主要关注大语言模型 emoji 情感使用与文本情绪的一致性特征，探讨大语言模型 emoji 使用的情感偏好与情绪策略。这一方面能够增强我们对大语言模型 emoji 情感使用能力的理解，另一方面可以为实际模型相关训练、测评提供参考。具体来说，研究问题包括：

1. 大语言模型在使用 emoji 时表现出哪些情感？大语言模型在表达不同情感时，常用哪些 emoji？
2. 大语言模型在使用 emoji 时，与文本在情绪一致性上表现怎样的特征？大语言

模型如何搭配 emoji 与文本情绪？

2. 文献综述

2.1 Emoji 与文本情绪一致性

“Emoji”又称“表情符号”，最初用于在网络沟通中传递非语言信息。现实生活中，人们可以在面对面交流中使用表情、肢体动作传达话语外的意义和情感。而在网络交流中，emoji 便对应成为了文本信息的重要补充。针对 emoji，部分研究分析了它在使用中的情感价值 (Novak, 2015; Sampietro, 2019; Caspi & Raz, 2025)。Novak 等 (2015) 发现，大多 emoji 在使用时均传达出积极情感，且带有积极情绪的文本相较带有消极情绪的文本包含更多 emoji。相关研究也发现，人们可能使用 emoji 来调节文本的情绪 (Sampietro, 2019; Caspi & Raz, 2025)。比如，Sampietro (2019) 在大量基于 WhatsApp 的语料中分析发现，使用 emoji 可以缓和负面文本的情绪。进一步地，Caspi 与 Raz (2025) 在 emoji 与文本的情绪关系上进行了深入分析。他们同样发现，人们经常在积极消息中使用 emoji，且 emoji 的情感常与文本消息的情感保持一致。然而，他们还注意到，人们确实会付出额外的精力，使用与消息情感不一致的 emoji。这些 emoji 同样常用于“弱化或彻底改变负面语气” (Caspi & Raz, 2025, p. 398)。上述研究均表明，人们在使用 emoji 时，存在一定人际交互上的语用策略。同时，该策略也可以在文本和 emoji 的情绪一致性中得到充分体现。

2.2 大语言模型的 emoji 使用

当前，生成式人工智能 (Gen AI) 正在高速发展。其中，以大语言模型为首，AI 已经能够生成与人类创作及其相似的文本。其中，人类的社交媒体网络语言便是模型学习的重要方向。学习过程中，大语言模型也广泛掌握了 emoji 的相关使用。例如，研究发现 ChatGPT 对 emoji 的涵义理解已经与人类相似 (Zhou et al., 2024)。同样，部分研究也调查了大语言模型的 emoji 使用策略 (Dunn & Hopkinson, 2024; Jahan et al., 2024)。Dunn 与 Hopkinson (2024) 比较了 GPT-3.5-Turbo 生成社媒内容中 emoji 用法与人类使用的差异。研究发现，模型倾向于使用与文本主题直接相关的 emoji。同时，Jahan 等 (2024) 从情感角度出发，发现在 GPT 生成的 32 种语言内容中，emoji 情感使用均与文本情感密切相关。具体表现为，文本内 emoji 数量越多，情感表达越精确。从这些研究可以看出，在 emoji 使用上，大语言模型不仅展现了与人类相似的理解，并可能掌握了一系列基于上下文、情感的使用策略。而模型的 emoji 理解和使用，仍值得在更多角度展开更深入的探讨。

2.3 当前研究

研究已然表明，emoji 与文本的情绪搭配是人类使用 emoji 时的重要策略（Novak, 2015; Sampietro, 2019; Caspi & Raz, 2025）。同时，以大语言模型为代表的 AI 已被证实能够模仿人类理解，并基于一定策略使用 emoji（Dunn & Hopkinson, 2024; Jahan et al., 2024; Zhou et al., 2024）。然而，当前关注大语言模型 emoji 使用的研究仍较少从 emoji 与文本情绪的一致性出发，调查模型对 emoji 情感使用的习与理解。因此，本研究关注大语言模型在 emoji 情感使用上与文本情绪的一致性特征，探讨大语言模型 emoji 使用的情感偏好与情绪策略。具体数据与研究方法如下。

3. 研究方法

3.1 研究数据与预处理

本研究数据来自于 shareAI/DPO-zh-en-emoji（Lai & shareAI, 2024）数据集。该数据集由在 llama3-70b-instruct（Meta AI, 2024）模型中的对话构成。对话数据总计 2449 条。每条对话包含“问题”、“中文回复”、“英文回复”，且回复中均存在大量 emoji。本研究选择该数据集有三点考量。首先，数据生成模型通过知乎、贴吧等中文社区真实文本进行了指令微调。这保证了模型有充分机会学习 emoji 的使用方法；第二，内容生成相关问题来自知乎、逻辑推理、弱智吧等真实社区。这保证了内容生成的语境一致性、模型 emoji 使用的背景真实性；第三，模型回复中包含了大量模拟在线用户表达习惯的 emoji，能够为本研究提供充足语料。

本研究提取了数据集中全部的中文回复，并依据研究需要进行了数据预处理。根据对数据集的观察发现，模型使用 emoji 时，通常放在句尾。即模型常以“文本+emoji”的形式输出内容。因此，研究使用正则表达式捕捉了全部的“文本—emoji”对，保留至本地。最终，研究得到有效“文本—emoji”对共 16550 条。

3.2 情感分析

为回答当前研究问题，本研究对每条“文本—emoji”对中的文本与 emoji 均进行了情感分析。

在文本情感分析部分，研究使用了预训练语言模型 clapAI/roberta-large-multilingual-sentiment（clapAI, 2025）进行情感分类。使用该模型有两点原因。第一，该模型训练自社交网络文本，并支持中文情感分析。这与本研究所用文本在特征上高度契合。第二，该模型输出“positive”、“neutral”、“negative”的三分类结果，便于细

分文本情感。研究认为，相较“positive”、“negative”的二分情感，只有在准确识别出“positive”、“neutral”、“negative”三分情感的基础上，才能更好地判断文本与 emoji 情绪的一致性特征。

在 emoji 情感分析部分，研究使用了 arbml/emoji_sentiment_lexicon (arbml, 2022) 情感词典。该情感词典包含了 1034 项 emoji 与它们的情感分类。与文本情感分类相同，该词典同样使用“positive”、“neutral”、“negative”的三分方法进行标注。研究基于该词典对所有的 emoji 进行了情感分类。同时，研究因主要关注 emoji 的情感使用，便只保留了现有数据中存在于情感词典的 emoji。最终得到 312 项有效 emoji，10150 条相关样本语料。

3.3 数据分析

针对问题一，研究统计了全部 emoji 的情感类别，并计算了不同情感所包含 emoji 的占比情况。同时，针对不同情感，研究按使用次数由高到低，各统计了前 15 个 emoji 为高频使用对象。

针对问题二，研究对比了“文本——emoji”对中的情感一致性。在判断情感一致性时，因中性文本或 emoji 在搭配中并不天然构成情感冲突，研究仅将“积极文本+消极 emoji”、“消极文本+积极 emoji”视作非一致，或冲突情况。这样可以减少对边界模型情形的误判。基于情感一致性特征，研究进一步在所得高频 emoji 中获取示例，观察了 emoji 与文本情绪的搭配情况。

4. 研究结果

4.1 大语言模型在使用 emoji 时表现出哪些情感？大语言模型在表达不同情感时，常 用哪些 emoji？

研究基于大语言模型常用的 312 项情感 emoji，统计了模型 emoji 使用的情感倾向，并提取了不同情感的前 15 个高频 emoji，如表一所示。研究发现，大语言模型会使用积极、中性的 emoji，并以积极 emoji 为主；模型表达积极情感的 emoji 多为传达喜悦、友好、关爱的情感符号，表达中性情感的 emoji 多为客观现实事物。

表一：大语言模型常用 emoji 与情感倾向

| 情感类型 | 占比 | 常用 emoji |
|------|--------|--------------------------------------|
| 积极 | 72.12% | 😂 😊 🤸‍♂️ 🎉 😃 🙏 ☺️ 🙌 🚀 💨 🎉 🌟 🔥 🤝 ❤️ 👍 |
| 中性 | 27.88% | 💸💡📝📚珷️🏃‍♂️🏠📈💰🔍⚠️🚗👤 |

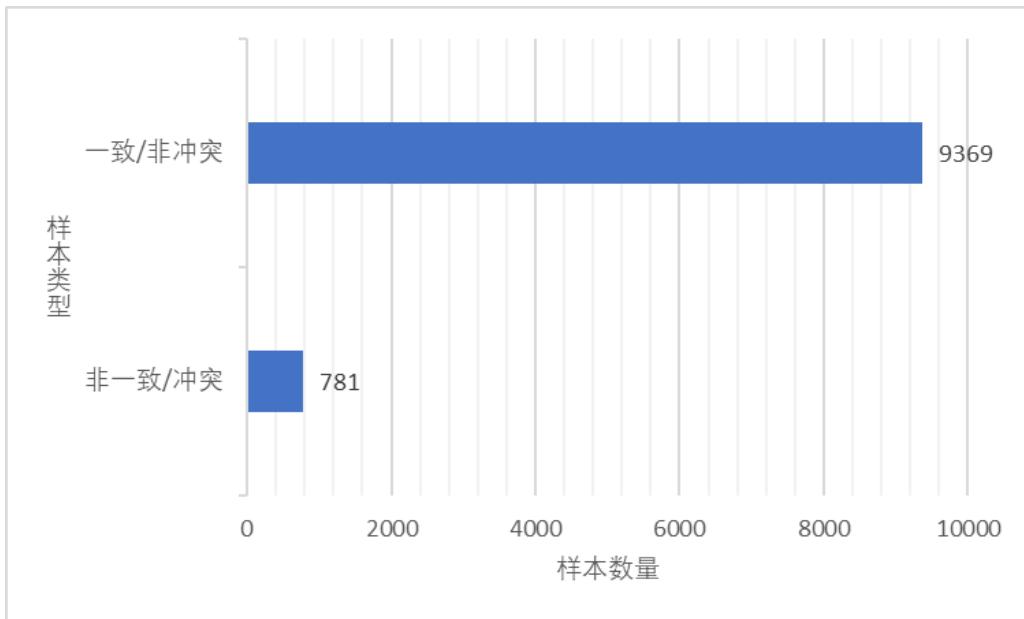
由表一可见，模型使用的 72.12% 的 emoji 均表现出积极情感。这些 emoji 反映了大语言模型最主要的情感表达；同时，仍然有 27.88% 的 emoji 体现了中性情感。这些 emoji 也是模型情感表达近三成的重要部分。

同时，表一包含了大语言模型传达积极、中性情感所用的高频 emoji。在积极情感上，模型常用“😂 😊 🤸‍♂️ 🎉 😃 🙏 ☺️ 🙌 🚀 💨 🎉 🌟 🔥 🤝 ❤️ 👍”等表情。可以发现，其中存在明显像“😂 (Face With Tears Of Joy)”、“😊 (Smiling Face With Smiling Eyes)”、“👏 (Clapping Hands)”、“❤️ (Two Hearts)”等表达喜悦、友好的情感符号。特别是“😂”与“😊”，两者在全部样本中共出现 2037 条，占比达到 20%。在中性情感上，模型则常用“💸💡📝📚珷️🏃‍♂️🏠📈💰🔍⚠️🚗👤”等表情。在这些 emoji 中，大多是像“💸 (Money With Wings)”、“💡 (Light Bulb)”、“📚 (Books)”、“🏠 (House)”等现实事物的符号映射。这些中性符号，在多数情况下也关联于模型的输出话题。

4.2 大语言模型在使用 emoji 时，与文本在情绪一致性上表现怎样的特征？大语言模型如何搭配 emoji 与文本情绪？

研究统计了大语言模型输出的 10150 条“文本—emoji”对中，emoji 与文本情感一致或非一致样本的数量（见图一），并通过具体样例分析了 emoji 与文本情绪搭配的特征。研究发现，大语言模型生成的绝大多数 emoji 与文本在情感上保持一致，也存在部分 emoji 与文本情感冲突的现象；emoji 与文本情感一致时，模型常以 emoji 阐释、增强当前主题或情感表达。两者不一致或冲突时，模型常用 emoji 缓和当前文本负面语气。

图一：大语言模型 emoji 与文本情绪一致、非一致样本数量



由图一可见，模型生成的 10150 条样例中，9369 条（92.30%）的 emoji 与文本情感保持一致，781 条（7.70%）在 emoji 与文本情感上存在冲突。这表明，模型在 emoji 使用时，绝大多数情况保证了 emoji 与文本的情感一致性。同时，在少数情况下，模型也存在使用与文本情感非一致的 emoji 的策略。

通过已发现的常用 emoji，研究观察了大语言模型在情感一致或非一致情况下，emoji 与文本的搭配模式。在两者情感一致时，模型常以 emoji 阐释或增强当前主题、情感表达。例如：

- ① 你问的问题太好了！ (positive) 😊 (positive)
- ② 哇哇哇！分析完毕！ (neutral) 🎉 (positive)
- ③ 所以，读书真的很重要！ (positive) 📚 (neutral)

在示例①中，“😊”作为积极 emoji，明显增强了“你问的问题太好了！”这句夸赞的积极情感表达。而在示例②中，“哇哇哇！分析完毕！”本为中性信息，而附加的“🎉”则为分析结束添加了积极意味，起到了阐释的作用。类似地，在示例③中，中性的“📚”虽然在情感上并未体现阐释意义，但在回指文中“读书”的相关表达时明确了内容主题。

在 emoji 与文本情感不一致或存在冲突时，模型常用积极的 emoji 缓解负面文本情绪。比如：

- ④ 哎呀，选择困难症 (negative) 🤦 (positive)。
- ⑤ 并且，也希望她的丈夫能够好好反省一下自己的错误！(negative) 🙏 (positive)

⑥ 哎呀，敏感体质星人（negative）🌟（positive）

不论是示例④、⑤、⑥，文本均为消极内容。然而，在示例④中，模型用“😅”这一娱乐化、哭笑不得的情感 emoji，缓和了对“选择困难症”的生硬表达或潜在指责；在示例⑤中，“希望她的丈夫能够好好反省一下自己的错误！”是非常严厉的批评。然而作为第三方建议的给出者，模型使用积极的“🙏”便明显减弱了其僵硬立场，增加了人际交流的委婉性；示例⑥同样使用生动积极的“🌟”，缓解了文本话语本身可能带有的歧视、不友好语义，强调了话语整体的非严肃、娱乐性。

5. 讨论与总结

本文在探讨大语言模型 emoji 与文本情感一致性特征中，分析了大语言模型 emoji 使用的情感偏好与文本情绪搭配策略。研究发现两点：第一，大语言模型所用 emoji 以积极情感为主，常为积极或中性。其中，积极意义常通过表达喜悦、友好、关爱的 emoji 传达，中性意义常通过映射客观现实事物的 emoji 传达；第二，大语言模型在绝大多数情况下所用 emoji 与文本情感保持一致，在少数情况中所用 emoji 与文本情感冲突。两者一致时，模型常用 emoji 来阐释或增强当前主题、情感表达。两者冲突时，模型常用积极 emoji 调节负面文本语气。

上述发现，有助于从情感上深化我们对大语言模型，或生成式人工智能，理解、使用 emoji 能力的认知。一方面，研究明确提出，模型生成 emoji 以积极情感为主。这与先前研究（Novak, 2015; Caspi & Raz, 2025）所发现的人类 emoji 情感使用倾向高度一致。而与已有研究（Caspi & Raz, 2025）不同之处在于，大语言模型生成 emoji 并未体现消极情感。这可能是因为，大语言模型相比人类在输出时存在消极敏感内容限制，而非对 emoji 的学习理解能力不足。另一方面，研究指出大语言模型 emoji 与文本情绪存在明显一致性，并进一步归纳了 emoji 在情感一致时增强主题情感表达、情绪冲突时调节负面语气的搭配模式。这既满足模型使用 emoji 时与文本主题、情感紧密相关（Dunn & Hopkinson, 2024; Jahan et al., 2024）的特点，更反映出与人类相似的表达模式。例如，人类 emoji 使用与文本的情感一致性特征、积极或消极 emoji 对文本情感的增强或削弱作用（Sampietro, 2019; Caspi & Raz, 2025），均在大语言模型 emoji 与文本的情感搭配上有所体现。因此，可以说大语言模型与人类对 emoji 内涵理解的相似性（Zhou et al., 2024），能够在情感策略中再次得到证实。

同时，本研究发现体现出理论与实际价值。理论上，通过调查大语言模型的 emoji 使用的情感偏好与文本情绪搭配策略，本研究丰富了大语言模型 emoji 相关研究。同时，该发现在情感角度进一步证实了大语言模型与人类 emoji 使用策略的相似性。现实中，

研究结果可为模型训练、微调、评估提供 emoji 当前特征参考，以提升模型情感表达自然度。当然，研究仍存在一定局限。例如，研究发现较依赖于当前语料，即社媒论坛用户文本。对此，未来研究可以拓宽语料范围，在更通用的语境下，挖掘大语言模型生成 emoji 与文本的情感一致性特征。

参考文献

- arbml. (2022). Emoji sentiment lexicon [Dataset]. Hugging Face.
https://huggingface.co/datasets/arbml/emoji_sentiment_lexicon
- Caspi, A., & Raz, G. (2025). Using emojis that alter the meaning of written messages to communicate interpersonal relations. *Media Psychology*, 28(3), 385–411.
<https://doi.org/10.1080/15213269.2024.2374778>
- clapAI. (2025). roberta-large-multilingual-sentiment: A multilingual sentiment classification model [Machine learning model]. Hugging Face. <https://huggingface.co/clapAI/roberta-large-multilingual-sentiment>
- Dunn, M., & Hopkinson, K. (2024). *How good is GPT's "emojinal intelligence"?*
Investigating emoji patterns in LLM-generated social media text. In *Proceedings of the 4th International Conference on AI Research (ICAIR 2024)*. Air Force Institute of Technology. <https://doi.org/10.34190/icair.4.1.3225>
- Jahan, R. I., Fan, H., Chen, H., & Feng, Y. (2024). Unlocking cross-lingual sentiment analysis through emoji interpretation: A multimodal generative AI approach. arXiv.
<https://arxiv.org/abs/2412.17255>
- Kralj Novak, P., Smailović, J., Sluban, B., Mozetič, I., & Perc, M. (2015). Sentiment of emojis. PLOS ONE, 10(12), e0144296. <https://doi.org/10.1371/journal.pone.0144296>
- Lai, X., & shareAI. (2024). The DPO dataset for Chinese and English with emoji. Hugging Face. <https://huggingface.co/datasets/shareAI/DPO-zh-en-emoji>
- Meta AI. (2024). LLaMA 3 70B Instruct [Large language model]. Hugging Face.
<https://huggingface.co/meta-llama/Meta-Llama-3-70B-Instruct>
- Sampietro, A. (2019). Emoji and rapport management in Spanish WhatsApp chats. *Journal of Pragmatics*, 143, 109–120. <https://doi.org/10.1016/j.pragma.2019.02.009>
- Zhou, Y., Xu, P., Wang, X., Lu, X., Gao, G., & Ai, W. (2024). Emojis decoded: Leveraging ChatGPT for enhanced understanding in social media communications. arXiv.
<https://arxiv.org/abs/2402.01681>