AUGMENTING IMAGE ANNOTATION: A HUMAN–LMM COLLABORATIVE FRAMEWORK FOR EFFICIENT OBJECT SELECTION AND LABEL GENERATION

He Zhang
College of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802, USA
hpz5211@psu.edu

Xinyi Fu*
The Future Laboratory
Tsinghua University
Beijing, China
fuxy@tsinghua.edu.cn

John M. Carroll
College of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802, USA
jmc56@psu.edu

ABSTRACT

Traditional image annotation tasks rely heavily on human effort for object selection and label assignment, making the process time-consuming and prone to decreased efficiency as annotators experience fatigue after extensive work. This paper introduces a novel framework that leverages the visual understanding capabilities of large multimodal models (LMMs), particularly GPT, to assist annotation workflows. In our proposed approach, human annotators focus on selecting objects via bounding boxes, while the LMM autonomously generates relevant labels. This human-AI collaborative framework enhances annotation efficiency by reducing the cognitive and time burden on human annotators. By analyzing the system's performance across various types of annotation tasks, we demonstrate its ability to generalize to tasks such as object recognition, scene description, and fine-grained categorization. Our proposed framework highlights the potential of this approach to redefine annotation workflows, offering a scalable and efficient solution for large-scale data labeling in computer vision. Finally, we discuss how integrating LMMs into the annotation pipeline can advance bidirectional human–AI alignment, as well as the challenges of alleviating the "endless annotation" burden in the face of information overload by shifting some of the work to AI.

1 INTRODUCTION

With the rapid development of deep learning and big data technologies, image annotation, an essential component in computer vision tasks has found widespread applications in fields (Najafabadi et al., 2015) such as autonomous driving (Huang et al., 2018), intelligent surveillance (Dharmawan et al., 2022), medical imaging (Hu et al., 2003), and emotional-behavior analysis (Zhang et al., 2024a). However, traditional image annotation tasks primarily rely on manual processes for selecting objects and assigning labels (Zhang et al., 2012). This approach is not only time- and labor-intensive (Zhang et al., 2012) but also prone to causing annotator fatigue due to prolonged repetitive work, which in turn affects the quality and consistency of the annotations (Herde et al., 2021).

To alleviate the burden of manual annotation, crowdsourcing methods have been widely adopted. For example, by leveraging Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHAs) (Von Ahn et al., 2003; Barnard et al., 2003; Chew & Tygar, 2004) and crowdsourcing platforms, large-scale annotation projects can be decomposed into numerous small tasks that are distributed to participants worldwide, with each participant responsible for annotating only a subset of images (Luz et al., 2015). This approach not only effectively reduces costs and improves efficiency but also enhances annotation accuracy through multiple rounds of verification and the collective intelligence of the crowd (Nowak & Rüger, 2010). Crowdsourcing is generally well-suited for simple, highly repetitive, and standardized tasks, such as basic image annotation or

*Corresponding Author

classification (e.g., the CAPTCHAs commonly seen on websites' login page). However, due to limitations in human knowledge, cultural differences, and experience, crowdsourcing is not suitable for generating more detailed labels (as illustrated in Figure 2, which simulates tasks that may be beyond the capabilities of non-experts) (Nassar & Karray, 2019).

In addition, automated tools and semi-automated annotation methods have gradually emerged in recent years. By leveraging pre-trained object detection models, semantic segmentation algorithms, and image captioning technologies, some systems attempt to automatically generate candidate bounding boxes and initial labels, with subsequent human correction through a human–machine collaborative process (Zhang et al., 2012; Cheng et al., 2018). Although these methods have improved annotation efficiency to some extent, they still suffer from limitations in generalization and user interaction. Existing automated tools often rely on specific datasets and scenarios, and when faced with diverse or complex situations, models are prone to misclassification and may struggle to accurately capture all details (Fernandes et al., 2024).

Recently, large language models (LLMs) and large multimodal models (LMMs) such as GPT have achieved groundbreaking progress in the field of natural language processing, and their exceptional semantic understanding and generation capabilities have provided new perspectives for annotation tasks (Gilardi et al., 2023; Belal et al., 2023; Zhang et al., 2024b; Nasution & Onan, 2024; Nguyen & Rudra, 2024; Zhu et al., 2024; He et al., 2024; Tan et al., 2024; Zendel et al., 2024; Wang et al., 2024; Li et al., 2024; Zhang et al., 2023; Lu et al., 2023; Zhang & Fu, 2025). This paper proposes a human–AI collaborative annotation framework (as shown in Figure 1), where humans are responsible for selecting target regions in images while an LMM automatically generates labels that align with the image context. The framework has the following two main advantages: (1) reduced human workload by delegating the laborious task of label assignment to the LMM, human annotators can focus solely on target selection, thereby significantly enhancing overall efficiency; and (2) bidirectional human–AI alignment in terms of knowledge and annotation accuracy. From the human annotators' perspective, they provide guidance (through the selection of regions) to help the LMM more effectively address specific task objectives and use their expertise to verify, correct, and offer feedback on the labels generated by the LMM. From the LMM's perspective, it can offer more detailed labels to compensate for the human annotators' potential lack of domain-specific knowledge. Furthermore, by leveraging the LMM's visual analysis capabilities, human annotators are not confined to the limitations of the annotation task, enabling them to broaden the scope of the task.

2 Proposed Framework

Our proposed framework (shown in Figure 1) streamlines the image annotation process through a systematic workflow. The process begins with a collection of raw images containing various objects of interest. Human annotators then review these images and draw bounding boxes around target objects, helping AI focus on the target and establishes connections between objects and the possible labels. These annotated images are then processed by a LMM, which analyzes the content within each bounding box using prompts such as "*Please tell me what is selected by the bounding box in each image.*" The LMM leverages its natural language understanding capabilities to generate precise labels for the outlined objects. This approach yields specific, high-quality annotations, for example, identifying specimens like "Saddle-Billed Stork", "Elephant Rhinoceros", "Giraffe", and "Ankole-Watusi", that serve as valuable input for downstream tasks such as object recognition or classification. If an image does not have a bounding box, the LMM will analyze the entire image. In some cases where the image contains only a single subject, a bounding box might not be necessary.

2.1 Comparison of the Traditional Workflow and the LMM-Enhanced Annotation Workflow

Figure 2 compares the workflow of traditional annotation tasks with the framework proposed in this paper. The traditional workflow places the entire burden of annotation on human annotators, who must both select objects and assign labels. These annotators first choose a specific predefined task, such as identifying animals, and maintain this focus throughout the process. They carefully draw bounding boxes around relevant objects, a step that demands precision since box accuracy directly influences annotation quality. The annotators then assign labels to each bounded region, a task that often requires specialized knowledge. When identifying animal species, for instance, annotators must navigate challenges such as blurry images or complex scenes. While this method can yield high-quality results, it suffers from three key limitations: heavy labor requirements, potential inconsistencies, and poor scalability.

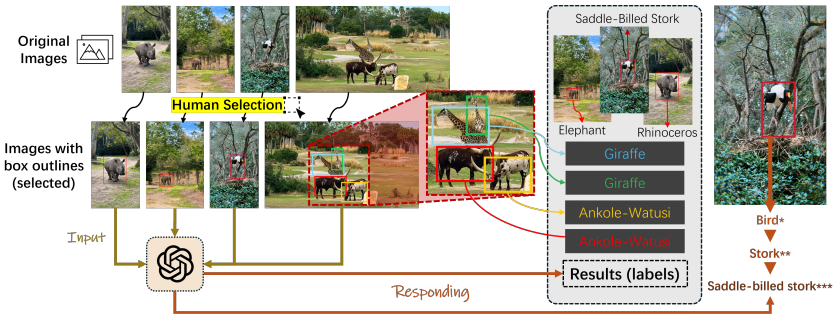

Figure 1: Image Annotation Workflow. The figure illustrates the steps involved in the image annotation process. It begins with a collection of original images, followed by human selection of relevant images. Selected images are then annotated with bounding boxes to highlight objects of interest. The final output consists of labeled bounding boxes, which are used for downstream tasks in computer vision. The rightmost part indicates the task levels that annotators with different knowledge levels can complete. A single asterisk (*) marks labels that can be annotated by the average person. Double asterisks (**) mark labels that can be annotated by those with some foundational or passing-knowledge. Triple asterisks (***) signify labels that only expert groups are deemed capable of annotating (Note: GPT-4o can annotate at this level).

The LMM-enhanced workflow addresses these limitations by dividing responsibilities between humans and machines. Human annotators now focus solely on drawing bounding boxes around objects of interest, without restricting themselves to specific categories. This approach reduces cognitive load while establishing the necessary context for subsequent machine processing. The LMM then analyzes these bounded regions and generates appropriate labels based on contextual prompts. In the example case of animal identification, the LMM can supply precise species names without requiring specialized knowledge from human annotators. This hybrid approach offers several advantages over the traditional workflow. By delegating the classification task to LMMs, it reduces the need for specialized expertise, lowers annotation costs, and significantly improves scalability. The division of labor between human visual expertise and machine classification capabilities creates a more efficient and sustainable annotation process, particularly for large datasets.

2.2 Testing and Results

To validate our LMM-enhanced annotation approach, we extended our evaluation to the Asirra dataset (Elson et al., 2007)[1], employing GPT-4-mini for rapid annotation testing[2]. The results revealed remarkable accuracy in primary classification tasks, achieving a 99.63% success rate in distinguishing between cats and dogs. Beyond basic classification, the proposed framework demonstrated sophisticated labeling capabilities. The LMM successfully generated detailed breed-specific annotations, such as ”Dachshund (Dog)”, ”German Shepherd (Dog)”, ”Siamese cat (Cat)”, and ”Himalayan cat (Cat)”. This granular classification ability highlights the system’s potential for specialized annotation tasks that traditionally require expert knowledge. These results underscore two key advantages of our approach: exceptional accuracy in basic classification tasks and the ability to provide detailed, breed-specific labels without additional human expertise. This combination of high accuracy and detailed classifi-

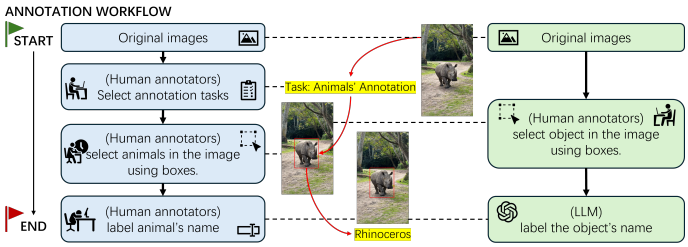

Figure 2: A Synergistic Loop Illustrating Bidirectional Human–AI Alignment.

[1]Testing utilized the annotated version available on Kaggle, https://www.kaggle.com/datasets/alvarole/asirra-cats-vs-dogs-object-detection-dataset

[2]Given the predominantly single-subject nature of the images, we omitted bounding box selection. Additionally, we utilized basic functionality prompts

cation capabilities suggests that LMM-enhanced annotation systems can effectively bridge the gap between efficiency and annotation depth.

3 Framework Supports Bidirectional Human-AI Alignment

This framework emphasizes bidirectional alignment (Shen et al., 2024) between human annotators and AI systems, specifically examining the interplay of labor distribution, knowledge transfer, and collaboration. On the one hand, from the human perspective, annotators provide essential guidance by selecting objects of interest through bounding boxes, helping LMMs focus on relevant areas and generate accurate labels. This division of labor reduces the cognitive burden on human annotators while LMMs complement potential gaps in human knowledge by generating detailed, contextually relevant labels, particularly in domains requiring specialized expertise like fine-grained animal species identification. On the other hand, from the AI perspective, LMMs benefit from the structured input provided by human annotators, enhancing their ability to understand and interpret visual content. Through prompt engineering and human selection, LMMs align their outputs with task objectives. This bidirectional interaction not only improves annotation accuracy but also creates a collaborative environment where human and AI systems learn from each other, forming a synergistic loop (as shown in Figure 3). Over time, this alignment leads to more robust and adaptable annotation systems capable of handling complex tasks across diverse domains.

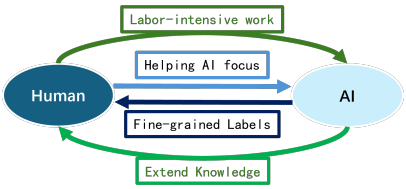

Figure 3: A Synergistic Loop Illustrating Bidirectional Human–AI Alignment.

4 Future Work: Data Explosion and Endless Annotation

In recent years, the exponential growth of digital data has spawned what's commonly called a "Data Explosion/Information Explosion." (Turi, 2024; Sweeney, 2001) As information proliferates across sectors, the demand for annotated datasets has surged to train and maintain high-performance AI models (Liang et al., 2022; Zini & Awad, 2022). Traditional annotation methods, heavily reliant on manual labor, struggle to keep pace with this relentless influx, creating a seemingly endless annotation backlog.

Our framework accelerates the annotation process by leveraging LMM capabilities. By shifting repetitive, labor-intensive tasks to AI, human annotators can focus on critical decisions like object selection and quality validation. Furthermore, it is essential to consider the economic and ethical implications of this approach. From an economic perspective, using LMM for annotation can undoubtedly reduce substantial labor costs, but it also raises further demands for computing resources (Bhattacharya et al., 2024). This trade-off can be analyzed in future work by comparing the savings from reduced human labor against the costs associated with using the LMM's API, potentially providing a clearer understanding of sustainability and return on investment. Additionally, deploying LMM locally and utilizing smaller-scale models might further reduce costs. From an ethical perspective, although automated annotation can improve efficiency, it also raises concerns about job cuts (Zarifhonarvar, 2024). Future work could focus on mitigating the negative impact on human workers, possibly by redefining their roles in the annotation workflow and directing them toward more strategic, high-level tasks. This balanced approach not only advances technological progress but also addresses broader societal impacts.

Despite these improvements, maintaining annotation quality across expanding datasets remains challenging. Future research could explore ways to enhance our framework's scalability. For instance, integrating active learning techniques (Prince, 2004) could help the system prioritize the most informative samples, optimizing both human and AI efforts. Additionally, employing image segmentation techniques (Han et al., 2024) to replace reliance on manual object boxing could enable the framework to operate more autonomously, adapting to new tasks and domains with minimal human intervention, ultimately transferring the endless annotation tasks brought by the data explosion entirely to AI.

Acknowledgments

We would like to extend our gratitude to Wen Chen and Xiaomeng Li for their assistance with the preliminary step of this work. We appreciate the insightful comments and suggestions provided by the anonymous reviewers. This work was supported by the Beijing Natural Science Foundation-Youth Project (Grant No.4254082).

REFERENCES

Kobus Barnard, Pinar Duygulu, David Forsyth, Nando De Freitas, David M Blei, and Michael I Jordan. Matching words and pictures. *The Journal of Machine Learning Research*, 3:1107–1135, 2003.

Mohammad Belal, James She, and Simon Wong. Leveraging chatgpt as text annotation tool for sentiment analysis, 2023. URL https://arxiv.org/abs/2306.17177.

Pronaya Bhattacharya, Vivek Kumar Prasad, Ashwin Verma, Deepak Gupta, Assadaporn Sapsomboon, Wattana Viriyasitavat, and Gaurav Dhiman. Demystifying chatgpt: An in-depth survey of openai's robust large language models. *Archives of Computational Methods in Engineering*, pp. 1–44, 2024. doi: https://doi.org/10.1007/s11831-024-10115-5.

Qimin Cheng, Qian Zhang, Peng Fu, Conghuan Tu, and Sen Li. A survey and analysis on automatic image annotation. *Pattern Recognition*, 79:242–259, 2018. doi: https://doi.org/10.1016/j.patcog.2018.02.017.

Monica Chew and J Doug Tygar. Image recognition captchas. In *International Conference on Information Security*, pp. 268–279. Springer, 2004. doi: https://doi.org/10.1007/978-3-540-30144-8_23.

Andi Dharmawan, Agus Harjoko, Faisal Dharma Adhinata, et al. Region-based annotation data of fire images for intelligent surveillance system. *Data in brief*, 41:107925, 2022. doi: https://doi.org/10.1016/j.dib.2022.107925.

Jeremy Elson, John (JD) Douceur, Jon Howell, and Jared Saul. Asirra: a captcha that exploits interest-aligned manual image categorization. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, pp. 366–374, New York, NY, USA, October 2007. Association for Computing Machinery. ISBN 9781595937032. doi: 10.1145/1315245.1315291. URL https://doi.org/10.1145/1315245.1315291.

Rodrigo Fernandes, Alexandre Pessoa, Marta Salgado, Anselmo De Paiva, Ishak Pacal, and António Cunha. Enhancing image annotation with object tracking and image retrieval: A systematic review. *IEEE Access*, 12:79428–79444, 2024. doi: 10.1109/ACCESS.2024.3406018.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120, 2023. doi: 10.1073/pnas.2305016120. URL https://www.pnas.org/doi/abs/10.1073/pnas.2305016120.

Kai Han, Victor S Sheng, Yuqing Song, Yi Liu, Chengjian Qiu, Siqi Ma, and Zhe Liu. Deep semi-supervised learning for medical image segmentation: A review. *Expert Systems with Applications*, pp. 123052, 2024. doi: https://doi.org/10.1016/j.eswa.2023.123052.

Zeyu He, Chieh-Yang Huang, Chien-Kuang Cornelia Ding, Shaurya Rohatgi, and Ting-Hao Kenneth Huang. If in a crowdsourced data annotation pipeline, a gpt-4. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642834. URL https://doi.org/10.1145/3613904.3642834.

Marek Herde, Denis Huseljic, Bernhard Sick, and Adrian Calma. A survey on cost types, interaction schemes, and annotator performance models in selection algorithms for active learning in classification. *IEEE Access*, 9:166970–166989, 2021. doi: 10.1109/ACCESS.2021.3135514.

Bo Hu, S. Dasmahapatra, P. Lewis, and N. Shadbolt. Ontology-based medical image annotation with description logics. In *Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence*, pp. 77–82, 2003. doi: 10.1109/TAI.2003.1250173.

Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

Zejian Li, Chenye Meng, Yize Li, Ling Yang, Shengyuan Zhang, Jiarui Ma, Jiayi Li, Guang Yang, Changyuan Yang, Zhiyuan Yang, Jinxiong Chang, and Lingyun Sun. Laion-sg: An enhanced large-scale dataset for training complex image-text models with structural annotations, 2024. URL https://arxiv.org/abs/2412.08580.

Weixin Liang, Girmaw Abebe Tadesse, Daniel Ho, Li Fei-Fei, Matei Zaharia, Ce Zhang, and James Zou. Advances, challenges and opportunities in creating data for trustworthy ai. *Nature Machine Intelligence*, 4(8):669–677, 2022. doi: https://doi.org/10.1038/s42256-022-00516-1.

Yuxuan Lu, Bingsheng Yao, Shao Zhang, Yun Wang, Peng Zhang, Tun Lu, Toby Jia-Jun Li, and Dakuo Wang. Human still wins over llm: An empirical study of active learning on domain-specific annotation tasks, 2023. URL https://arxiv.org/abs/2311.09825.

Nuno Luz, Nuno Silva, and Paulo Novais. A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, 44:187–213, 2015. doi: https://doi.org/10.1007/s10462-014-9423-5.

Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of big data*, 2:1–21, 2015. doi: https://doi.org/10.1186/s40537-014-0007-7.

Lobna Nassar and Fakhri Karray. Overview of the crowdsourcing process. *Knowledge and Information Systems*, 60:1–24, 2019. doi: https://doi.org/10.1007/s10115-018-1235-5.

Arbi Haza Nasution and Aytuğ Onan. Chatgpt label: Comparing the quality of human-generated and llm-generated annotations in low-resource language nlp tasks. *IEEE Access*, 12:71876–71900, 2024. doi: 10.1109/ACCESS.2024.3402809.

Thi Huyen Nguyen and Koustav Rudra. Human vs chatgpt: Effect of data annotation in interpretable crisis-related microblog classification. In *Proceedings of the ACM Web Conference 2024*, WWW '24, pp. 4534–4543, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400701719. doi: 10.1145/3589334.3648141. URL https://doi.org/10.1145/3589334.3648141.

Stefanie Nowak and Stefan Rüger. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the International Conference on Multimedia Information Retrieval*, MIR '10, pp. 557–566, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605588155. doi: 10.1145/1743384.1743478. URL https://doi.org/10.1145/1743384.1743478.

Michael Prince. Does active learning work? a review of the research. *Journal of engineering education*, 93(3):223–231, 2004.

Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, Sushrita Rakshit, Chenglei Si, Yutong Xie, Jeffrey P. Bigham, Frank Bentley, Joyce Chai, Zachary Lipton, Qiaozhu Mei, Rada Mihalcea, Michael Terry, Diyi Yang, Meredith Ringel Morris, Paul Resnick, and David Jurgens. Towards bidirectional human-ai alignment: A systematic review for clarifications, framework, and future directions, 2024. URL https://arxiv.org/abs/2406.09264.

Latanya Sweeney. Information explosion. *Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies*, pp. 43–74, 2001.

Zhen Tan, Dawei Li, Song Wang, Alimohammad Beigi, Bohan Jiang, Amrita Bhattacharjee, Mansooreh Karami, Jundong Li, Lu Cheng, and Huan Liu. Large language models for data annotation and synthesis: A survey, 2024. URL https://arxiv.org/abs/2402.13446.

Abeba N Turi. Data explosion, algorithm economy, and the ai fervidness. In *Innovation, Sustainability, and Technological Megatrends in the Face of Uncertainties: Core Developments and Solutions*, pp. 3–22. Springer, 2024.

Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *Advances in Cryptology—EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4–8, 2003 Proceedings 22*, pp. 294–311. Springer, 2003.

Xinru Wang, Hannah Kim, Sajjadur Rahman, Kushan Mitra, and Zhengjie Miao. Human-llm collaborative annotation through effective verification of llm labels. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3641960. URL https://doi.org/10.1145/3613904.3641960.

Ali Zarifhonarvar. Economics of chatgpt: A labor market view on the occupational impact of artificial intelligence. *Journal of Electronic Business & Digital Economics*, 3(2):100–116, 2024. doi: https://doi.org/10.1108/JEBDE-10-2023-0021.

Oleg Zendel, J. Shane Culpepper, Falk Scholer, and Paul Thomas. Enhancing human annotation: Leveraging large language models and efficient batch processing. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*, CHIIR '24, pp. 340–345, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704345. doi: 10.1145/3627508.3638322. URL https://doi.org/10.1145/3627508.3638322.

Dengsheng Zhang, Md Monirul Islam, and Guojun Lu. A review on automatic image annotation techniques. *Pattern Recognition*, 45(1):346–362, 2012. doi: https://doi.org/10.1016/j.patcog.2011.05.013.

He Zhang and Xinyi Fu. Benchmarking zero-shot facial emotion annotation with large language models: A multi-class and multi-frame approach in dailylife, 2025. URL https://arxiv.org/abs/2502.12454.

He Zhang, Chuhao Wu, Jingyi Xie, ChanMin Kim, and John M. Carroll. Qualigpt: Gpt as an easy-to-use tool for qualitative coding, 2023. URL https://arxiv.org/abs/2310.07061.

He Zhang, Xinyang Li, Yuanxi Sun, Xinyi Fu, Christine Qiu, and John M. Carroll. Vrmn-bd: A multi-modal natural behavior dataset of immersive human fear responses in vr stand-up interactive games. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 320–330, 2024a. doi: 10.1109/VR58804.2024.00054.

He Zhang, Chuhao Wu, Jingyi Xie, Yao Lyu, Jie Cai, and John M. Carroll. Redefining qualitative analysis in the ai era: Utilizing chatgpt for efficient thematic analysis, 2024b. URL https://arxiv.org/abs/2309.10771.

Yiming Zhu, Zhizhuo Yin, Gareth Tyson, Ehsan-Ul Haq, Lik-Hang Lee, and Pan Hui. Aptpipe: A prompt-tuning tool for social data annotation using chatgpt. In *Proceedings of the ACM Web Conference 2024*, WWW '24, pp. 245–255, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400701719. doi: 10.1145/3589334.3645642. URL https://doi.org/10.1145/3589334.3645642.

Julia El Zini and Mariette Awad. On the explainability of natural language processing deep models. *ACM Comput. Surv.*, 55(5), December 2022. ISSN 0360-0300. doi: 10.1145/3529755. URL https://doi.org/10.1145/3529755.