Demystify Adult Learning: A Social Network and Large Language Model Assisted Approach

Fang Liu
Singapore University of Social Sciences
Singapore, 599494
liufang@suss.edu.sg

Bosheng Ding
Nanyang Technological University
Singapore, 639798
bosheng001@e.ntu.edu.sg

Chong Guan*
Singapore University of Social Sciences
Singapore, 599494
guanchong@suss.edu.sg

Zhang Wei
Singapore Institute of Technology
Singapore, 138683
wei.zhang@singaporetech.edu.sg

Dusit Niyato
Nanyang Technological University
Singapore, 639798
dniyato@ntu.edu.sg

Justina Tan
Singapore University of Social Sciences
Singapore, 599494
justinatanpi@suss.edu.sg

Abstract—**Adult learning is increasingly recognized as a crucial way for personal development and societal progress. It however is challenging, and adult learners face unique challenges such as balancing education with other life responsibilities. Collecting feedback from adult learners is effective in understanding their concerns and improving learning experiences, and social networks provide a rich source of real-time sentiment data from adult learners. Machine learning technologies especially large language models (LLMs) perform well in automating sentiment analysis. However, none of such models is specialized for adult learning with accurate sentiment understanding. In this paper, we present `A-Learn`, which enhances <u>a</u>dult <u>learn</u>ing sentiment analysis by customizing existing general-purpose LLMs with domain-specific datasets for adult learning. We collect adult learners' comments from social networks and label the sentiment of each comment with an existing LLM to form labelled datasets tailored for adult learning. The datasets are used to customize `A-Learn` from several base LLMs. We conducted experimental studies and the results reveal `A-Learn`'s competitive sentiment analysis performance, achieving up to 91.3% accuracy with 20% improvement over the base LLM. `A-Learn` is also employed for word cloud analysis to identify key concerns of adult learners. The research outcome of this study highlights the importance of applying machine learning with educational expertise for teaching improvement and educational innovations that benefit adult learning and adult learners.**

Index Terms—**social network, large language model, generative artificial intelligence, sentiment analysis, applied artificial intelligence, adult learning**

I. INTRODUCTION

Adult learning is essential for personal growth and societal advancement [1]. It involves the continuous acquisition of knowledge and skill throughout one's adulthood [2]. This process contributes to a more competent workforce and better employment prospects, and positively affects the economy with better job opportunities and increased productivity. It also promotes social inclusion by supporting marginalized groups to participate more fully in society [3]. Unlike traditional education, which is typically confined to formal learning in childhood and adolescence, adult learning brings unique challenges. Adult learners must balance their educational pursuits with other adult responsibilities, which interfere with their regular attendance of classes, timely completion of assignments, and overall focus on studies. Adult learners may also struggle with adapting to new learning environments and technologies, and face financial constraints [4].

Therefore, it is vital to gather feedback from adult learners and understand their unique needs to enhance the adult learning experience [5]. Traditionally, educators use surveys, in-class discussions, and course evaluation forms to collect feedback, but these methods may yield limited and potentially biased responses due to low participation rates. To address this, researchers are increasingly turning to social networks where a large volume of real-time comments from the learners are available. Analyzing such data allows for examining the social impacts of adult learning and offering insights into learners' real-time opinions, experiences, and sentiments [6].

However, the vast and ever-increasing volume of data from social networks makes traditional analysis, often manual, impractical. To overcome this challenge, researchers have studied machine learning (ML) models to automate data processing and analysis. Within ML, natural language processing (NLP) is particularly useful for analyzing social network data, where text is the dominant data modality. An important application of NLP is sentiment analysis, which accesses the emotional tone behind comments to infer sentiments such as `Positive`, `Neutral`, or `Negative`. NLP excels in such tasks, and recent advancements have led to the development of many large language models (LLMs) that are trained on huge text datasets for enhanced performance in sentiment analysis. Some LLMs are specifically tailored for certain tasks and platforms, e.g., Twitter-RoBERTa [7], trained on 58 million tweets, is highly effective for various tasks including sentiment analysis for Twitter comments.

Unfortunately, even advanced LLMs may not perform optimally across all subjects due to the broad nature of training data which may not cover specialized areas comprehensively. While an LLM might demonstrate high accuracy in specific applications, there is a risk of over-fitting, where the model cannot generalize well to new data. This leads to a significant increase in error rates when the model is applied to unfamiliar data. Additionally, the constant evolution of language, including the emergence of new buzzwords and acronyms commonly used by students on social media, presents a continuous challenge for pre-trained LLMs which may struggle to adapt to these changes.

In this paper, we propose a customized LLM tailored to adult learning which is crucial for effective analysis in this field. Utilizing a specialized dataset for adult learning is essential for providing domain-specific knowledge to the models and refining the performance on the targeted topics. Accurate labelling of this specialized dataset is important for extracting meaningful insights. However, sourcing such accurately labeled datasets presents a significant challenge for adult learning research. To address this, we introduce a system called `A-Learn` for adult learning sentiment analysis. `A-Learn` utilizes social network comments enhanced with LLM and accordingly customizes LLM models for adult learning. We outline our main contributions as follows.

- We collected an adult learning dataset from multiple social network platforms with over 55,000 comments from current and prospective adult learners;
- We generated two additional datasets: a large-scale dataset labelled by LLM for model fine-tuning and customization, and a small-scale dataset labelled manually by human experts for performance evaluation;
- We developed `A-Learn` to incorporate adult learning knowledge into existing LLMs with optimized training and produce LLMs customized for adult learning;
- We conducted experimental studies and demonstrated `A-Learn`'s competitive performance in sentiment analysis for adult learning, with up to 91.3% accuracy and a 20% improvement over the baseline.

The rest of the paper is organized as follows. The system architecture of `A-Learn` is presented in Section II. Section III describes our methodology in detail and Section IV presents an experimental study of the proposed solution as well as result discussions. Finally, we conclude the paper in Section V.

II. SYSTEM ARCHITECTURE

In this section, we present the system architecture of `A-Learn` and an illustration of it is available in Fig. 1.

A. LLM for Sentiment Analysis

The sentiment analysis of adult learning can be conducted in different ways and LLM is one of the effective solutions to automatic the analysis. In a typical scenario, an LLM takes student comments (e.g., from social media and questionnaires) as input. The data is pre-processed such as modality alignment before the ML analysis, where sentiment analysis is treated
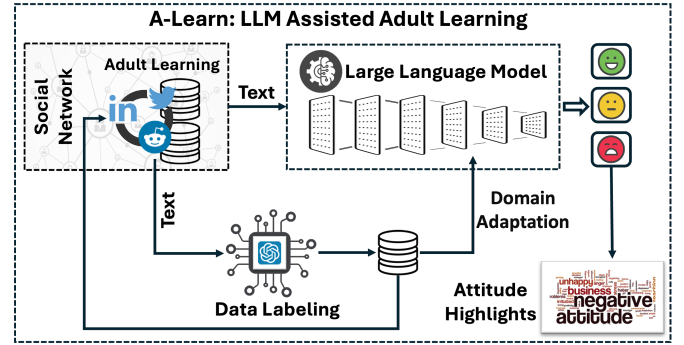

Fig. 1. The system architecture of `A-Learn` for adult learning sentiment analysis. `A-Learn` is ML-based and processes the social network comments as input to generate sentiment labels as output. `A-Learn`'s sentiment analysis plays a crucial role in identifying key concerns and facilitating educational innovations in adult learning.

as a classification problem and the model's role is to generate sentiment labels including `Positive` for satisfaction, `Negative` for dissatisfaction, and `Neutral` otherwise. The sentiment labels allow for the grouping of comments, where in-depth investigation can be performed for each group to gain a deep understanding of adult learners for teaching quality enhancement and educational innovation.

B. LLM Customization for Adult Learning

Despite being a promising technology, existing LLM is not specialized for adult learning. There are significant differences between general-purpose text analysis and adult learning sentiments, where some phrases, acronyms, etc., are not common in generic settings. This calls for a customization, or domain adaptation, of existing LLM to adult learning scenarios. In `A-Learn`, we first collect an adult learning dataset from various social media platforms to capture the adult learning concerns in different aspects. Yet, no one has indicated the sentiment of any comments. Thus, we apply an existing LLM, ChatGPT in specific, to label the comments in the dataset. The labels may not be fully correct from the views of human experts, but they are cheap to obtain and very likely to be correct. Now, with the dataset specialized for adult learning as well as associated sentiment labels, we customize the generic LLMs to adapt to adult learning with accurate sentiment analysis. This is a typical transfer learning concept, which has been verified in other applications [8]. We expect the transfer or customization, to be effective with our specialized dataset and training strategies.

III. METHODOLOGY

We present the detailed methodology of `A-Learn` in this section, for both data collection and knowledge adaptation.

A. LLM-assisted Sentiment Label Acquisition

Adult learning as a specific topic of sentiment analysis involves unique patterns that are not common in generic settings. Frequent phrases such as time management, work-life balance, teaching quality, and resource allocation, may

"i had to deal with this same issue in a computer applications class. very basic things like how to open, save, copy, paste, and rename files. along with learning microsoft word, excel and powerpoint. this was an advanced version of this class in that it lasted half as long 6 weeks instead of 12 . so, we were supposed to be moving much quicker than the full length class. i was astounded to learn that hardly anyone in the class could keep up. it was ridiculous and i was super frustrated. i finished my work first in every class and many of the other students in the class would ask me for help. i helped as much as i could until they started asking me for help before i had finished my work. i really love helping people understand things, but i can't help the entire class and get my own work done at the same time."

(a) Mixed Sentiment

"… the long and short of it is my one professor is a real piece of work, and i highly doubt my ability to do well in his course 😊. not only are all the lectures pre-recorded, but they are also very long and rambling as well often going over the allotted time for the course for the week and are essentially required viewing if you expect to do well on the tests. worse is the fact that the professor seems very reluctant to actually interact with the students in any real, meaningful way **TL;DR.** god help you if you actually want any real feedback on how you did on the tests 👍…"

(b) Negative Sentiment

Fig. 2. Sample comments with different emotions/sentiments about time management and teaching quality for adult learning.

Fig. 3. An illustration of `A-Learn` with social network comments as the input. Domain knowledge is extracted from the comments and represented as sentiment initiates, used to fine-tune different layers in LLM's transformer encoder. The customized model can identify the sentiment of a comment.

not be well modeled in general-purpose LLMs. Besides, the sentiment can be complicated with mixed emotions. We show one sample comment in Fig. 2(a), where a student expresses enthusiasm for helping peers, but feels struggles when additional demands arise and affect his/her own learning progress. Fig. 2(b) presents a sample comment with negative sentiment, featuring network slang and emojis, which add analysis complexity.

To understand the unique patterns of adult learning, we consider social networks that offer comprehensive information regarding student comments about adult learning. However as mentioned above, we lack sentiment labels of the comments and face challenges to interpret the data and extract insights. As such, we introduce a labelling process, which shall not be fully manual considering the labelling complexity and cost. We follow a hybrid approach with the assistance of both LLM and human experts. We task ChatGPT to label comments with sentiment and let human experts verify the labels, assuming verification is much easier and faster than labelling. According to recent studies [9], [10] for other applications, such an approach can be cost-efficient with big time-savings.

B. `A-Learn` with LLM Customization

It takes a huge dataset to train an LLM to achieve optimal performance, but a huge dataset is often very costly to obtain. This is no difference for adult learning and we shall choose the data-lite strategies. We first adopt pre-trained LLMs with optimal performance for understanding generic text content. Each LLM consists of many parameters and the optimal values for generic usage are known and downloadable for these LLMs. We take such optimal values as the starting point for customizing the parameters to adult learning. We adjust and optimize the values as a model re-training or fine-tuning process, and our ultimate goal is to adopt the parameter values for general-purpose text analysis to the domain of adult learning. This process is illustrated in Fig. 3. LLM tokenizes the input text and converts them into numerical representations, which are then fed into the transformer encoder for language
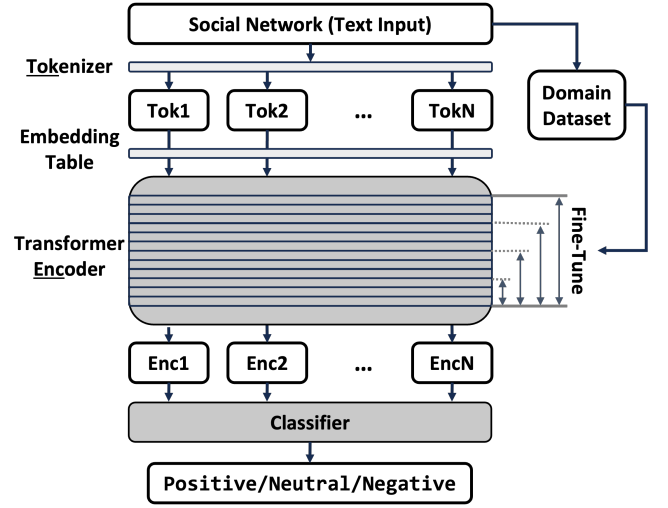
comprehension. We follow transfer learning to re-train or fine-tune the models with the adult learning dataset. This process adjusts the weights of various layers in the transformer architecture. Specifically, we can freeze some layers near the input and fine-tune the rest layers near the output. The more layers we freeze, the more information from the pre-trained model is retained. Conversely, adjusting more layers allows the penetration of more domain-specific information, which however demands more data to perform well. The output from the transformer encoding is passed to a classifier, which produces the final three-fold sentiment labels, `Positive`, `Negative`, and `Neutral`.

Let $\mathcal{M}(x; \theta)$ be the sentiment analysis modal which analyzes the comment x and determines its sentiment label as $y = \mathcal{M}(x; \theta)$ with model parameters θ. We assume two connected stages of the whole model as $(\mathcal{M}_f, \mathcal{M}_t)$ where $\mathcal{M}_f$ and $\mathcal{M}_t$ are the sets of frozen and trainable layers, respectively, along with the corresponding model parameters (θ_f, θ_t). In `A-Learn`, we initialize the parameters as θ_f^0 and θ_t^0 which are copied from the optimal parameter values from the base LLMs. The aim of customization is to find the optimal parameter values θ_t^* for adult learning sentiment analysis as,

$$\theta_t^* = \arg\min_{\theta_t} \mathcal{L}(\mathcal{D}; \theta_f^0, \theta_t), \qquad (1)$$

where $\mathcal{L}(\cdot)$ is the cross-entropy loss function, $\mathcal{D}$ is our labelled dataset, and θ_f remains unchanged as θ_f^0. The final model of `A-Learn` is $\mathcal{M}(\cdot; \theta^*)$ where $\theta^* = (\theta_f^0, \theta_t^*)$. Specifically, re-training means all transformer layers are trainable where $\mathcal{M}_f$ is null and the optimal model is $\mathcal{M}(\cdot; \theta^*)$ where $\theta^* = \theta_t^*$. With comments and `A-Learn`'s generated labels, educators can perform in-depth analysis for improved teaching.

IV. Experimental Study

We present our experimental study in this section to demonstrate `A-Learn`'s sentiment analysis performance for adult learning and interpret the experiment results to find insights. `A-Learn` is data-driven and we start with data description.

A. Dataset and Data Pre-processing

In this part, we first introduce the data collection for our experiments, followed by data pre-processing descriptions.

1) Data Collection: Social networks are popular among various groups of users. It provides rich and extensive data that facilitates in-depth analysis of human opinions and sentiments. In `A-Learn`, we scrap plenty of comments related to adult learning from multiple social media platforms, including Reddit, Twitter, and Tumblr. They are selected because of their diverse user demographics (e.g. various age groups and backgrounds) and well-supported APIs. Our comment search is based on keywords such as `adult learning`, `part-time studies`, `continuous learning`, and `work study`. In total, we have 31,724 Reddit comments collected from 2017 to 2023, 22,270 Twitter tweets in 2023, and 1,043 Tumblr comments from 2011 to 2023.

2) Data Pre-processing: The raw data (comments) collected from the above stage is not ready for ML models and necessary pre-processing is needed to facilitate the ML-based analysis. Firstly, we remove duplicated and incomplete content. Then, we scan the content and eliminate special text components, including HTML links and tags, emojis, and newline characters. All contents are normalized and converted to lowercase to ensure consistency. Finally, the contents are processed through tokenization, lemmatization, and stop-word removal to prepare them to be compatible with ML.

3) Datasets for Experiments: We present two datasets specialized for adult learning with different labelling schemes.

a) LLM Labels for Training: We use LLM to generate a relatively large-scale labelled dataset. Particularly, we use `GPT-3.5` and get its opinion about the sentiment of each of the 1,000 randomly selected social network comments. Some labels are randomly selected for manual inspection by human experts, to evaluate if the labelling quality meets our expectations. Note that we do not expect fully correct labels in `A-Learn` training.

b) Expert Labels for Testing: We need fully correct labels for performance evaluation. Thus we introduce human expert support. We randomly select comments from our collected dataset. We let one human expert read each comment and assign a sentiment label to it, and the label is verified by two other human experts. Only if the three experts reach a consensus, we include the comment along with its label into the testing dataset. We stop the labelling once we have 300 labelled comments with one-third for each category of sentiment. This will be our testing dataset.

B. A-Learn Performance Analysis

`A-Learn` considers several widely used or latest LLMs as base models and then customizes them with our labelled adult learning dataset. We first present the experimental setup.

TABLE I
THE PERFORMANCES IN ACCURACY (%) AND TRAINING TIME IN MINUTE OF BASE MODELS AND A-LEARN CUSTOMIZED FROM THE MODELS. A-LEARN ALWAYS ACHIEVES BETTER ACCURACY COMPARED TO THE CORRESPONDING BASE MODELS WITH SIGNIFICANT IMPROVEMENT AND MANAGEABLE TIME USAGE FOR TRAINING.

Model Name	Base	A-Learn	Improvment	Time
Bert	41.0	61.7	20.7	11
ALBERT	33.0	35.7	2.7	10
RoBERTa	32.0	56.3	24.3	12
Distilbert	36.0	59.0	23.0	6
Distilbert-SST2	44.0	59.3	15.3	7
Twitter-RoBERTa	76.3	**91.3**	15.0	12

1) Experimental Setup: We use Google Colab to implement `A-Learn`. For the base model, we select the widely used BERT [11] and five of its variants, including ALBERT [12], RoBERTa [13], DistilBERT, DistilBERT-SST2 [14], and Twitter-RoBERTa [7]. Here, DistilBERT-SST2 and Twitter-RoBERTa are two specialized models trained with the SST2 dataset (a benchmark for sentiment analysis) and a Twitter dataset, respectively. Some models have different versions and we choose the default one, normally with a suffix of `-base`, which is excluded in this paper for simplicity.

In the following part, we use `Base` as a reference of the original LLMs and `A-Learn` is the LLM customized for adult learning. `A-Learn` is fine-tuned with the LLM-labelled dataset for at least ten epochs. Accuracy, in the percentage of the number of correct predictions to the total number of predictions, is used as the main performance metric based on the testing results on expert-labelled dataset. All experimental procedures are executed on NVIDIA V100 GPUs.

2) Sentiment Analysis Performance: We evaluate `A-Learn`'s performance on improving pre-trained `Base` models for adult learning sentiment analysis. The statistical results of our experiments are presented in Table I. The results demonstrate a consistent and substantial accuracy improvement of `A-Learn` with six different models after model customization with domain-specific data. The accuracy improvement ranges from 2.7% to 24.3% and the average is 16.8%. This validates our hypothesis that incorporating domain-specific knowledge into existing models can boost the model's performance in adult learning sentiment analysis. Notably, `A-Learn` based on Twitter-RoBERTa reaches an accuracy of 91.3%, with a big increase from 76.3% when Twitter-RoBERTa is not customized. This underscores the importance of customizing the base models with the data of a similar format (e.g., from Twitter) for sentiment analysis.

`A-Learn` is also efficient. The training time is about ten minutes for all tested LLMs. Compared with training a general-purpose model which takes days or even weeks, `A-Learn` is much faster. One reason is that `A-Learn` starts its training with the optimal model parameters for generic applications that share similarities with adult learning. This leads to reduced training difficulty and accelerated training.

TABLE II
PERFORMANCE SENSITIVITY OF A-LEARN TO DIFFERENT NUMBERS OF
TRAINABLE LAYERS IN TERMS OF ACCURACY (%). THE IMPROVEMENT
COLUMN REFERS TO THE A-LEARN'S ACCURACY IMPROVEMENT IN
PERCENTAGE OVER THE BASE MODEL.

Model	No. of Trainable Layers	Accuracy	Improvement
Base	0 (0%)	76.3	—
A-Learn	1 (8%)	72.0	-5.6
	7 (58%)	83.7	9.7
	12 (100%)	**91.3**	19.7

TABLE III
THE PERFORMANCES IN ACCURACY (%) OF BASE MODELS AND A-LEARN
FOR SENTIMENT ANALYSIS WITH DIFFERENT LABELLING METHODS. THE
IMPROVEMENT IN PERCENTAGE IS BASED ON THE COMPARISON WITH
BASE MODEL.

Customization Data	Accuracy	Improvement (%)
Base	76.3	-
LLM $y \to \hat{x}$	64.3	-15.6
LLM $x \to \hat{y}$ (ours)	**91.3**	19.7

C. Sensitivity Analysis

In this section, we explore different A-Learn settings and identify the optimal configuration of A-Learn.

1) Number of Trainable Layers: We have proved that customization is effective for adult learning sentiment analysis. In this part, we further investigate how to customize the knowledge to achieve optimal performance. We take the best-performing model, Twitter-RoBERTa, as the Base model and we conduct sensitivity analysis. We examine A-Learn's performance sensitivity to the number of trainable layers, and we show the results in Table II. This analysis aims to identify the ideal configuration of the trainable layers to reach the best performance of adult learning sentiment analysis.

Specifically, we investigate the transformer stage of Twitter-RoBERTa, which comprises 12 layers. Generally, the layers closer to the model output are more application-specific, and should be trainable to adapt to our application. In our experiment, let n be the number of trainable layers nearest to the model output, while the remaining layers are frozen. For example, $n = 0$ refers to the Base model where no layers is trainable. In A-Learn, we explore different settings of n ranging from 1 to 12. Due to the limit of space, we report the representative ones with the minimum $n = 1$ layer, about half of the layers with $n = 7$, and all the layers with $n = 12$.

The results show an increase in accuracy as the number of trainable layers n increases, reaching a peak accuracy of 91.3% when all layers are trainable. This trend suggests that limiting the number of trainable layers can negatively impact the model's ability to adapt to a specific task, potentially due to insufficient adjustment to the unique patterns of our task. Notably, the accuracy drops to 72% when $n = 1$, even lower than the performance of Base. This implies insufficient domain adaptation for the new application, while the analysis of the original model is disrupted with model changes. In conclusion, allowing more layers to be trainable helps A-Learn better adapt to the sentiment analysis application of adult learning.

2) Different LLM-assisted Data: In sentiment analysis, there are two primary approaches using LLMs for data labelling: 1) assign sentiment labels to comments, and 2) generate comments given specific labels. Let a comment be x. The role of LLM in the first approach is to generate $\hat{y} = \text{LLM}(x)$ where x is the input (real data) of the LLM and $\hat{y}$ is LLM's inferred sentiment label (not ground-truth). We represent this approach as $x \to \hat{y}$. For the second approach, represented by $y \to \hat{x}$, we provide the sentiment label y (without adding any comments), and let the model generate a list of comments corresponding to the label, i.e., $\hat{x} = \text{LLM}(y)$ for each synthetic comment $\hat{x}$. We select Twitter-RoBERTa as the Base and report A-Learn's performance in Table III.

Our default setting of LLM-assisted labelling is shown in the last line of $y \to \hat{x}$, where A-Learn is 91.3% accurate. Compared with the base model without customization, it is nearly 20% more accurate. With another scheme $y \to \hat{x}$, the accuracy declines compared to the base model, by 15.6% with 64.3% sentiment analysis accuracy. A potential reason is that social network comments usually are complex and the texts are often unstructured and informal. There are also new and non-common phrases with the use of slang, abbreviations, and various languages or dialects. Generating synthetic comments $\hat{x}$ purely from labels y with LLM is challenging. We notice that LLM tends to standardize the output comments with similar language styles, length, etc., where the complexity of the real comments is not well captured. As a result, customization with such data does not lead to improved performance for A-Learn, and even worse, the accuracy drops. We assume that the dataset fails to bring extra value in adult learning to A-Learn and meanwhile disrupts the base model with optimal performance for general-purpose applications.

In summary, our experiments demonstrate that the LLM-labelled dataset is useful for customizing the general-purpose LLM to adult learning with optimized configurations. This aligns with the findings from previous works [9] that tagging unlabeled data using LLMs is an effective method compared to traditional data collection schemes. This is especially true for our application where the comments own unique features specialized to adult learning.

D. A-Learn-enabled Insights Visualization

In this part, we apply A-Learn in its optimal configuration for analyzing the comments collected from all three platforms, Reddit, Twitter, and Tumblr. We apply A-Learn on the collected comments that are not used in neither the training nor testing dataset of small-scale, meaning most of the comments collected are used in this part. After A-Learn processing, all these comments are assigned with a sentiment label. We pick out the comments with the same label, specifically Positive and Negative. For each group of comments with the same label, we conduct a word cloud analysis to visualize the most frequent words for each sentiment of adult learning. We show

Fig. 4. The word clouds with highlighted adult learning concerns for both negative and positive sentiments. Larger words appear more frequently in the comments and imply significant concerns of adult learners.

the word cloud figures in Fig. 4 where the words with limited linguistic information like stop words are excluded.

1) Negative: Negative sentiment is important as it often sparks insights into teaching improvement and educational innovation. As seen from the figure, negative comments are largely personal with domestic challenges. Unlike traditional learners, adult learners must balance their studies with family and work responsibilities. This is evidenced by the highlighted words such as `time job`, `family vacation`, `family time`, etc. Additionally, adult learners often lack frequent interactions with peers and instructors and spend much of their study time `home alone`. Older students may face difficulties related to their age and physical fitness, indicated by terms such as `old enough`, while younger students may experience pressure related to their `first job` and financial obligations such as `make money`.

2) Positive: Positive sentiment is related to professional development and advice for effective learning. For instance, adult learners emphasize the need for `work smarter` and `work hard` at their jobs, to free up time for adult learning. This highlights the continued importance of balancing work and study. Adult learning shall be persistent, with phrases `every day`. Interestingly, utilizing online resources and open-sourced tools such as `khan academy` seems to be effective in enhancing adult learning. Overall, `A-Learn` based word clouds for different sentiments can shed light on teaching improvement, policy adjustment, etc., with special attention on adult learning rather than traditional learning.

V. CONCLUSION

In this paper, we introduced `A-Learn` to demystify the sentiment of adult learners by analyzing their social network comments. `A-Learn` leveraged advanced LLMs and was tailored for adult learning. It can be based on different LLMs and several of them were investigated in this paper. `A-Learn` customizes the base LLMs, often general-purposed, with adult learning dataset extracted from social networks and labelled by an LLM. Experiment results demonstrated that customizing existing LLMs with domain-specific data is effective, and `A-Learn` can achieve high accuracy in sentiment analysis in adult learning scenarios. Specifically, `A-Learn` achieves its peak performance when Twitter-RoBERTa is used as its `Base` model. Its accuracy can be up to 91.3%, surpassing the non-customized base model by 19.7%. Besdies, customization itself demands an optimized process, and we identified the optimal configuration of `A-Learn` training through sensitivity studies. Overall, this is an interdisciplinary research combining the strengths of ML and social science. The research outcome holds substantial implications for continuing education and insights for better-supporting adult learners.

REFERENCES

[1] M. Tennant, *Psychology and adult learning.* Routledge, 2007.
[2] J. Mezirow, *Transformative dimensions of adult learning.* ERIC, 1991.
[3] R. Desjardins and A. Ioannidou, "The political economy of adult learning systems—some institutional features that promote adult learning participation," *Zeitschrift für Weiterbildungsforschung*, vol. 43, no. 2, pp. 143–168, 2020.
[4] C. Guan, J. Mou, and Z. Jiang, "Artificial intelligence innovation in education: A twenty-year data-driven historical analysis," *International Journal of Innovation Studies*, vol. 4, no. 4, pp. 134–147, 2020.
[5] T. Shaik, X. Tao, Y. Li *et al.*, "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," *IEEE Access*, vol. 10, pp. 56 720–56 739, 2022.
[6] K. D. Philp and M. G. Gill, "Brokering youth learning opportunities in after-school settings: The role of adult social networks and network orientations." *Translational Issues in Psychological Science*, vol. 8, no. 2, p. 257, 2022.
[7] U. Sirisha and S. C. Bolem, "Aspect based sentiment & emotion analysis with roberta, lstm," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 11, 2022.
[8] F. Liu, W. Zhang, I. Atmosukarto *et al.*, "Transline: transfer learning for accurate and explainable power line anomaly detection with insufficient data," *CCF Transactions on Pervasive Computing and Interaction*, vol. 5, no. 3, pp. 241–254, 2023.
[9] B. Ding, C. Qin, L. Liu *et al.*, "Is gpt-3 a good data annotator?" in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 11 173–11 195.
[10] M. Li, T. Shi, C. Ziems *et al.*, "Coannotating: Uncertainty-guided work allocation between human and large language models for data annotation," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 1487–1505.
[11] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
[12] Z. Lan, M. Chen, S. Goodman *et al.*, "Albert: A lite bert for self-supervised learning of language representations," in *International Conference on Learning Representations*, 2019.
[13] Y. Liu, M. Ott, N. Goyal *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[14] V. Sanh, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter." in *Proceedings of Thirty-third Conference on Neural Information Processing Systems (NIPS2019)*, 2019.