SigBart: Enhanced Pre-training via Salient Content Representation Learning for Social Media Summarization

Sajad Sotudeh
IRLab, Georgetown University
Washington, D.C., USA
sajad@@ir.cs.georgetown.edu

Nazli Goharian
IRLab, Georgetown University
Washington, D.C., USA
nazli@ir.cs.georgetown.edu

ABSTRACT

Our approach to automatically summarizing online mental health posts could help counselors by reducing their reading time, enabling quicker and more effective support for individuals seeking mental health assistance. Neural text summarization methods demonstrate promising performance owing to their strong pre-training procedure. *Random token/span masking* technique is often relied upon by existing pre-trained language models; an approach that overlooks the importance of content when learning word representations. In an attempt to rectify this, we propose using source and summary alignments as a saliency signal to enhance the pre-training strategy of language model for better representation learning of important content, paving the way for a positive impact on the model fine-tuning phase. Our experiments on a mental health-related dataset for user post summarization (MentSum) reveal improved performance, as evidenced by human evaluation metrics, surpassing the current state-of-the-art system.

CCS CONCEPTS

• **Computing methodologies → Natural language generation**.

KEYWORDS

Text Summarization, Social Media Text Summarization, Pretrained Text Summarization

ACM Reference Format:
Sajad Sotudeh and Nazli Goharian. 2024. SigBart: Enhanced Pre-training via Salient Content Representation Learning for Social Media Summarization. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3589335.3652505

1 INTRODUCTION AND RELATED WORK

Social media platforms such as Reddit have become popular spaces for people to discuss a range of topics, including mental health. These forums are crucial for individuals seeking mental health support or advice. However, for mental health professionals navigating and understanding the vast amount of user-generated content is challenging [4, 9, 23, 26]. Creating summaries of these posts is not only a technical task but also a means to grasp the nuance aspects of mental health more effectively [13, 20, 24, 33]. In this context, summarization efforts aim to create concise, abstractive summaries, known as tl;dr (Too Long; Didn't Read), from user posts.

The majority of research in the mental health domain has primarily concentrated on predictive tasks [6–8, 18, 22, 32]. However, the development of the MentSum dataset, as introduced in [25], has made the task viable. Building on this, [24] achieved the state-of-the-art (SOTA) performance by implementing a curriculum-guided abstractive summarizer on this dataset. In a broader context, pre-trained language models [10, 16, 17, 30] have revolutionized NLP by two-step process of pre-training and fine-tuning. Existing models [1, 2, 14] in the social media domain have primarily focused on fine-tuning. However, none have focused specifically on pre-training the language model for summarization-specific purposes in the social media domain. Our work fills this gap by enhancing the pre-training stage, identifying salient content, and masking it for representation learning. Despite some work showing that further pre-training on in-domain data can enhance model performance[12, 15, 29], none have considered content importance [1]. Our work is the first to incorporate content saliency in the pre-training stage for the social media domain. We identify important content by aligning it with the tl;dr summary, mask it, and have the pre-trained language model (i.e., Bart) predict these masked spans. This replaces Bart's random text infilling pre-training objective with masked important content. We then use this pre-trained model to fine-tune the state-of-the-art summarizer, BART + R3F [2], for mental-health social media summarization. The work that comes closest to ours is Pegasus [30], which masks the top-m important sentences. However, we diverge by using token alignments with the summary as a measure of content saliency. Our evaluations reveal our method outperforms the SOTA system in fluency, informativeness, and faithfulness, while also improving automatic evaluation metrics. In summary, our contributions include proposing and evaluating a novel model that pre-trains the language model on masked salient content over social media data to improve mental health post summarization.

2 APPROACH

Aiming to improve important span representations (as defined earlier), we replace Bart's typical random masking approach with *salient span masking*. We explain the steps in the following in more detail.

Identification of salient spans. To identify the salient spans, we employ two orthogonal methods to make alignments between

[1] In our framework, content importance in pre-training is defined as those spans of text that are likely to be present in summary.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3652505

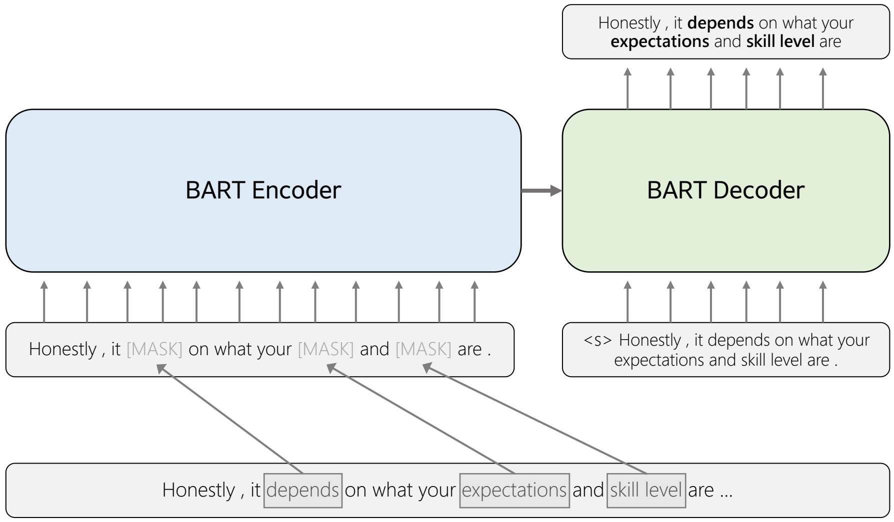

Figure 1: An example sentence from the pre-training dataset. As shown, the important spans with varying lengths are first selected, masked, and fed into the encoder. The autoregressive decoder then reconstructs the input text by predicting the masked input spans.

users' posts and their associated TL;DR summary. Specifically, we frame the alignment problem as a sequence-tagging task, where the post terms are tagged if they are copied into the TL;DR. Formally written, let $\mathbf{b} = \{b_0, b_1, ..., b_n\}$ be the binary tags over the post words (excluding stop words) [2] $\mathbf{x} = \{x_0, x_1, ..., x_n\}$ and TL;DR terms $\mathbf{y} = \{y_0, y_1, ..., y_m\}$, with n and m being the length of the user post and TL;DR, respectively. To align the user post to the TL;DR, we denote the word x_i as copied if (1) it is included in the longest possible subsequence of words $s = \{x_{i-j:i;i;i+k}, (j \leq i, k \leq (n-i))\,|s \in \mathbf{x}, s \in \mathbf{y}\}$, and (2) there exists no earlier sequence u with $s = u$. While the proposed method generates supervised labels for token masking, it does not guarantee that enough labels will be generated due to the abstractive nature of the pretraining dataset. To address this issue, we identify frequently occurring spans across the data collection and annotate their occurrence within each sample. Our investigations show that around 32% of tokens (on average) are identified as being part of the salient spans. [3]

Pre-training BART on salient spans. Our goal is to learn effective representations of the salient content (i.e., odds-on-copied spans as discussed in Section 2) for better summarization in the fine-tuning stage. In this sense, after identifying and masking salient spans, we pre-train the BART model over the masked spans. A sample of pre-training iteration is shown in Figure. 1. The BART model [4] is trained to generate the output sequence by predicting the masked tokens within the input. We choose to pre-train the model using *text infilling* pre-training objective, considering its effectiveness as reported by Lewis et al. [16]. Our pre-training strategy, as well as the associated pre-trained checkpoint, is referred to as SIGBART, which is short for **Sig**nificant **Bart**.

[2]To increase the match between user post and summary, we use the lemmatization.
[3]BART randomly masks 30% of tokens (default setting).
[4]We used `bart-large` as the initial checkpoint.

3 EXPERIMENTAL SETUP

The experimental setup is set to address the following research questions;

- **RQ1**: Does pre-training on salient content/spans lead to improved summaries in terms of automatic and qualitative metrics as compared to the SOTA BART + R3F model? [5] and
- **RQ2**: Does the proposed pre-training method lead to improvements as compared to the existing baselines that differ in pre-training and fine-tuning stages, in terms of ROUGE evaluation metrics?

Datasets. We use WEBIS-TLDR-17[27] for pre-training [6], which comprises 3.8M user posts and their paired TL;DR summaries. For fine-tuning, we use MENTSUM [25] with over 24k mental-health post, summary pairs. For more information on the datasets, please refer to the respective papers.

Comparison. We compare our model with several extractive and abstractive baselines as listed below. The BART + R3F baseline is for evaluating RQ1, and the rest for RQ2.

- **CURRSUM** [24]: a SOTA curriculum-guided abstractive model that exploits SUPERLOSS (a confidence-aware curriculum loss) as a progressive training signal to enhance the learning efficiency and performance of the summarization model.
- **BART + R3F** [2]: a SOTA abstractive system that discourages representation change during fine-tuning when possible without hurting the system performance. In the experiment, we include this model to assess the impact of SIGBART's pre-training strategy in comparison to a scenario when it is not employed (i.e., RQ1).
- **MATCHSUM** [31]: an extractive summarizer that matches text spans of source and summary in semantic space for selecting the most important sentences in fine-tuning time.

[5]Note that the fine-tuning stage of both models is the same, and the only difference is the pre-training stage to validate our hypothesis.
[6]The pre-training dataset comes from social media domain.

Model	Rg-1 (%)	Rg-2 (%)	Rg-L (%)	BS (%)
SigBart (ours)	**30.79***	**9.41***	**21.72***	**86.69***
Bart + R3F [2021b]	29.66	8.78	21.25	85.79

(a)

Model	Rg-1 (%)	Rg-2 (%)	Rg-L (%)	BS (%)
CurrSum† [2022b]	30.16	8.82	21.24	86.32
MatchSum† [2020]	26.29	6.32	17.12	–
Pegasus [2020]	29.48	8.26	20.39	–
SocialPegasus	29.71	8.49	20.66	–
Bart† [2020]	29.13	7.98	20.27	–

(b)

Table 1: Results for Rouge (F1) and Bertscore (F1) metrics comparing our SigBart model with (a) Bart + R3F (RQ1); (b) existing baseline systems (RQ2), on the MentSum dataset. indicates statistical significance (paired t-test, $p < 0.05$) against Bart + R3F. † shows the results reported from prior works on MentSum [24, 25].

- **Bart** [16]: an abstractive model that pre-trains the encoder-decoder framework for summarization task.
- **Pegasus** [30]: an abstractive model that adds a new pre-training objective where gap sentences are masked to be predicted by the remaining sentences in the input document.
- **SocialPegasus**: a Pegasus model that we have pre-trained on the same pre-training dataset as SigBart with Pegasus's pre-training objectives. The purpose of including this baseline is to compare the effectiveness of our pre-training approach compared to the Pegasus pre-training.

Implementation. We use Huggingface Transformers [28] to implement SigBart model and the Bart + R3F model for fine-tuning. We pre-train the Bart-large model on dual NVIDIA RTX A6000 GPUs for 7 days, 5 epochs, with a learning rate of $3e-5$ and warm-up steps of 2000. For fine-tuning, we fix the learning rate to $3e-5$ with a weight decay of 0.01. We fine-tune the Bart + R3F and SigBart for 8 epochs with intervals of 0.5 epoch for validation, and use the checkpoint with the highest rouge-1 score in inference time.

4 RESULTS AND DISCUSSION

4.1 Automatic evaluation.

Table 1 presents experimental results with respect to automatic evaluation metrics over MentSum dataset. In what follows, we present the results and discussions around each research question.
RQ1: As demonstrated in Table 1 (a), SigBart improves rouge-1 by 3.8%, rouge-2 by 6.6%, rouge-l by 2.2%, and Bertscore by 1.1% in comparison to the Bart + R3F model. As the fine-tuning stage of these two models are the same, with difference being on pre-trained stage, this improvement backs our hypothesis on effectiveness of learning salient representations during the pre-training stage.
RQ2: Table 1 (b) provides the automatic evaluation of other existing baselines that have different pre-training and fine-tuning. The results indicate that SigBart outperforms these existing systems, including the best performing baseline (i.e., CurrSum) by relative improvements of 2.1% (rouge-1), 6.3% (rouge-2), 2.3% (rouge-l), and 0.5% (Bertscore). Moreover, a comparative analysis of Pegasus and SocialPegasus shows only a marginal improvement, indicating that pre-training Pegasus on short documents, akin to those found in social media, may not lead to considerable benefits. SocialPegasus, despite its approach, still lags behind SigBart model. This discrepancy could potentially be attributed to the divergence between the pre-training strategies of SigBart and SocialPegasus, the latter of which adopts a similar but deviating pre-training.

System	Flue.	Info.	Faith.
Bart + R3F	4.43	3.92	4.26
agr. rate (%)	47	52	53
SigBart	**4.51**	**4.18**	**4.41**
agr. rate (%)	51	54	54
Relative imp. (%)	+1.8	+6.4	+3.4

Table 2: Human evaluation scores, agreement rates, and relative improvement of SigBart over Bart + R3F, on MentSum.

4.2 Human evaluation.

We undertook a human evaluation to contrast SigBart with the Bart + R3F baseline (i.e., the baseline with the same fine-tuning, yet different pre-training), using three qualitative metrics: Fluency (is the summary easy to read and understand?), Informativeness (does the summary provide key information about the user post?), and Faithfulness (is the summary's information supported by the user post?). The evaluation task was double-blinded, involving 50 randomly picked cases from the MentSum's test set. To minimize bias, we randomized the order of the summaries for evaluation. For the scoring process, each case was independently reviewed by two assessors, scoring each sample from 1 (worst) to 5 (best) based on these metrics. A third assessor resolved any scoring differences exceeding 2 points.

As presented in Table 2, SigBart model consistently surpasses the baseline in all qualitative metrics. The improvements are specifically prominent in the informativeness and faithfulness metrics, presenting relative improvements exceeding 6% and 3%, respectively. This highlights the effectiveness of the SigBart-specific pre-training method in improving the qualitative metrics. We also attribute gains in fluency to the augmented transferability achieved

User-written	Baseline-generated	SigBart-generated
spooned a girl i like and didn't make a move cause ...	***cuddled with girl** i like, got blocked by her best friend.*	***cuddled with a girl** i like, she was really close to me and **i didn't make a move**.*
	rouge (%): (19.35 / 13.79 / 19.35) Human: (4, 2, 3)	rouge (%): (71.79 / 37.84 / 51.28) Human: (3, 4, 5)
*i went to walmart with $1000 in cash in my wallet, got distracted by some dumb sh*t while meandering around and somehow lost my wallet and now i'm a broke *****.*	***went to walmart** with my wallet in a pamphlet holder caddy, lost it in the clearance aisle and now **i'm out $1000** and my credit card.*	*i **lost my wallet** at **walmart** and now **i'm out $1000**.*
	rouge (%): (57.63 / 24.56 / 37.29) Human: (4, 3, 3)	rouge (%): (45.45 / 23.81 / 36.36) Human: (4, 2, 5)
made a lunch lady believe i've immigrated to study and stay at my cousins because of how awkward the situation was when in fact i've lived in england my whole life.	*made a lady think i've **come over to england to study** and am **staying at my cousins** when in fact i was born in england.*	*i made a stranger think i've been living in england **for my whole life** when i was born in saudi.*
	rouge (%): (64.41 / 38.60 / 61.02) Human: (4, 4, 4)	rouge (%): (51.86 / 23.08 / 37.04) Human: (3, 3, 2)

Table 3: Three samples of the the user-written, baseline (Bart + R3F) generated, and SigBart-generated tl;dr summaries. The text that is unfaithful to the post is in Gray and the salient information that is picked up by the summarization systems is shown in Bold. We further show the rouge evaluation results for each system summary (rouge-1, rouge-2, rouge-l), as well as the human-assigned scores (Fluency, Informativeness, Faithfulness) below each generated summary. The user post is partially shown to preserve user privacy.

by continued pre-training on in-domain data collections [12]. Cohen's inter-rater agreement indicates moderate agreement.

5 ERROR ANALYSIS

Evaluating underperformed cases, we found that SigBart generates text spans by muddling different source materials, which is a pressing issue of SOTA abstractive summarizers [5, 19], requiring them to deal with complex reasoning over user post. For instance, SigBart unfaithfully produces phrase "`...i was born in saudi...`" by attending to "`born`" and "`saudi`" from two different regions of the user post. Furthermore, SigBart misses producing segments that are not directly mentioned within the user post, but inferred from it. This is a common phenomenon in social media summarization as also denoted by Sotudeh et al. [25], and it remains yet an open question for pre-trained summarization models. For example, we have a phrase "`...saved 317 gb worth of data...`" in the gold summary, where "`317`" is inferred from the user post; neither SigBart nor Bart + R3F infers "`317`" to produce in the summary. Lastly, the current aligner employs exact matching between the user post and the summary, leading to certain limitations. For instance, a phrase like "`...saw food being wasted...`" in the gold summary does not align with "`...seeing perfectly good food thrown away...`" in the user post. Similarly, "`...meandering around...`" in the gold summary does not align with "`...start walking away...`" in the user post. Despite these phrases conveying similar meanings, they are not detected as alignments by the current system. Future research could address this by employing semantic textual alignment methods, as suggested by Ernst et al. [11]. Examples of generated summaries on two samples are shown in Table 3.

6 ETHICS AND PRIVACY

In compliance with similar research practices on health domain [3, 21], Our research strictly adhered to ethical standards by using only anonymized data, ensuring no identification and exploration of users. We maintained stringent data security measures, with access restricted to the research team, to uphold confidentiality and integrity in our mental health study. While our automated summarization system advances state-of-the-art capabilities, it is crucial to acknowledge the potential for misrepresentation or oversimplification of complex mental health narratives within the posts. Given the sensitivity of such topics, special care must be exercised when implementing these technologies on mental health-related datasets at a production scale, ensuring that their application does not compromise the nuanced understanding of mental health issues.

REFERENCES

[1] Armen Aghajanyan, Anchit Gupta, Akshat Shrivastava, Xilun Chen, Luke Zettlemoyer, and Sonal Gupta. 2021. Muppet: Massive Multi-task Representations with Pre-Finetuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. ACL, Online and Punta Cana, Dominican Republic, 5799–5811. https://doi.org/10.18653/v1/2021.emnlp-main.468

[2] Armen Aghajanyan, Akshat Shrivastava, Anchit Gupta, Naman Goyal, Luke Zettlemoyer, and Sonal Gupta. 2021. Better Fine-Tuning by Reducing Representational Collapse. In *International Conference on Learning Representations*. https://openreview.net/forum?id=OQ08SN70M1V

[3] Adrian Benton, Glen A. Coppersmith, and Mark Dredze. 2017. Ethical Research Protocols for Social Media Health Research. In *EthNLP@EACL*.

[4] Judith Willemijn Borghouts, Elizabeth Victoria Eikey, Gloria Mark, Cinthia De Leon, Stephen Matthew Schueller, Margaret Schneider, Nicole A. Stadnick, Kai Zheng, Dana B. Mukamel, and Dara H. Sorkin. 2021. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. *Journal of Medical Internet Research* 23 (2021). https://api.semanticscholar.org/CorpusID:232339811

[5] Sihao Chen, Fan Zhang, Kazoo Sone, and Dan Roth. 2021. Improving Faithfulness in Abstractive Summarization with Contrast Candidate Generation and Selection. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, 5935–5941. https://doi.org/10.18653/v1/2021.naacl-main.475

[6] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. In *ICWSM*.

[7] Arman Cohan, Sydney Young, and Nazli Goharian. 2016. Triaging Mental Health Forum Posts. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, Kristy Hollingshead and Lyle Ungar (Eds.). Association for Computational Linguistics, San Diego, CA, USA, 143–147. https://doi.org/10.18653/v1/W16-0316

[8] Arman Cohan, Sydney Young, Andrew Yates, and Nazli Goharian. 2017. Triaging content severity in online mental health forums. *Journal of the Association for Information Science and Technology* 68 (2017).

[9] Serina R. Deen, Amy Withers, and David J. Hellerstein. 2013. Mental Health Practitioners' Use and Attitudes Regarding the Internet and Social Media. *Journal of Psychiatric Practice* 19 (2013), 454–463. https://api.semanticscholar.org/CorpusID:205424410

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[11] Ori Ernst, Ori Shapira, Ramakanth Pasunuru, Michael Lepioshkin, Jacob Goldberger, Mohit Bansal, and Ido Dagan. 2021. Summary-Source Proposition-level Alignment: Task, Datasets and Supervised Baseline. In *Proceedings of the 25th Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Online, 310–322. https://doi.org/10.18653/v1/2021.conll-1.25

[12] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 8342–8360. https://doi.org/10.18653/v1/2020.acl-main.740

[13] Fazida Karim, Azeezat Abimbola Oyewande, Lamis F Abdalla, Reem Chaudhry Ehsanullah, and Safeera Khan. 2020. Social Media Use and Its Connection to Mental Health: A Systematic Review. *Cureus* 12 (2020). https://api.semanticscholar.org/CorpusID:220598690

[14] Byeongchang Kim, Hyunwoo Kim, and Gunhee Kim. 2019. Abstractive Summarization of Reddit Posts with Multi-level Memory Networks. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 2519–2531. https://doi.org/10.18653/v1/N19-1260

[15] Jan-David Krieger, Timo Spinde, Terry Ruas, Juhi Kulshrestha, and Bela Gipp. 2022. A Domain-Adaptive Pre-Training Approach for Language Bias Detection in News. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries* (Cologne, Germany) *(JCDL '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 7 pages. https://doi.org/10.1145/3529372.3530932

[16] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703

[17] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Ro{BERT}a: A Robustly Optimized {BERT} Pretraining Approach. https://openreview.net/forum?id=SyxS0T4tvS

[18] Sean MacAvaney, Bart Desmet, Arman Cohan, Luca Soldaini, Andrew Yates, Ayah Zirikly, and Nazli Goharian. 2018. RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses. In *CLPsych@NAACL-HTL*.

[19] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On Faithfulness and Factuality in Abstractive Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 1906–1919. https://doi.org/10.18653/v1/2020.acl-main.173

[20] John A. Naslund, Ameya P. Bondre, John B Torous, and Kelly A. Aschbrenner. 2020. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. *Journal of Technology in Behavioral Science* 5 (2020), 245 – 257. https://api.semanticscholar.org/CorpusID:219087850

[21] Jennifer Nicholas, Sandersan Onie, and Mark Erik Larsen. 2020. Ethics and Privacy in Social Media Research for Mental Health. *Current Psychiatry Reports* 22 (2020), 1–7.

[22] Philip Resnik, Anderson Garron, and Rebecca Resnik. 2013. Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students. In *EMNLP*.

[23] Koustuv Saha, Asra Yousuf, Ryan L. Boyd, James W. Pennebaker, Munmun De, and Choudhury. 2022. Social Media Discussions Predict Mental Health Consultations on College Campuses. *Scientific Reports* 12 (2022). https://api.semanticscholar.org/CorpusID:244708187

[24] Sajad Sotudeh, Nazli Goharian, Hanieh Deilamsalehy, and Franck Dernoncourt. 2022. Curriculum-guided Abstractive Summarization for Mental Health Online Posts. In *Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI)*, Alberto Lavelli, Eben Holderness, Antonio Jimeno Yepes, Anne-Lyse Minard, James Pustejovsky, and Fabio Rinaldi (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 148–153. https://doi.org/10.18653/v1/2022.louhi-1.17

[25] Sajad Sotudeh, Nazli Goharian, and Zachary Young. 2022. MentSum: A Resource for Exploring Summarization of Mental Health Online Posts. In *Proceedings of the Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 2682–2692. https://aclanthology.org/2022.lrec-1.287

[26] Severi Uusitalo, Peter Eskolin, and Petros Belimpasakis. 2009. A solution for navigating user-generated content. *2009 8th IEEE International Symposium on Mixed and Augmented Reality* (2009), 219–220. https://api.semanticscholar.org/CorpusID:37779571

[27] Michael Völske, Martin Potthast, Shahbaz Syed, and Benno Stein. 2017. TL;DR: Mining Reddit to Learn Automatic Summarization. In *Proceedings of the Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, Copenhagen, Denmark, 59–63. https://doi.org/10.18653/v1/W17-4508

[28] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[29] Tiezheng Yu, Zihan Liu, and Pascale Fung. 2021. AdaptSum: Towards Low-Resource Domain Adaptation for Abstractive Summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, 5892–5904. https://doi.org/10.18653/v1/2021.naacl-main.471

[30] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. In *ICML*.

[31] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2020. Extractive Summarization as Text Matching. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 6197–6208. https://doi.org/10.18653/v1/2020.acl-main.552

[32] Ayah Zirikly and Mark Dredze. 2022. Explaining Models of Mental Health via Clinically Grounded Auxiliary Tasks. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, Ayah Zirikly, Dana Atzil-Slonim, Maria Liakata, Steven Bedrick, Bart Desmet, Molly Ireland, Andrew Lee, Sean MacAvaney, Matthew Purver, Rebecca Resnik, and Andrew Yates (Eds.). Association for Computational Linguistics, Seattle, USA, 30–39. https://doi.org/10.18653/v1/2022.clpsych-1.3

[33] Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. 2021. DepressionNet: Learning Multi-modalities with User Post Summarization for Depression Detection on Social Media. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. ACM. https://doi.org/10.1145/3404835.3462938