

TherapyView: Visualizing Therapy Sessions with Temporal Topic Modeling and AI-Generated Arts

Baihan Lin^{1,*}, Stefan Zecevic², Djallel Bouneffouf² and Guillermo Cecchi²

¹*Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA*

²*IBM TJ Watson Research Center, Yorktown Heights, NY, 10598, USA*

Abstract

We present the TherapyView, a demonstration system to help therapists visualize the dynamic contents of past treatment sessions, enabled by the state-of-the-art neural topic modeling techniques to analyze the topical tendencies of various psychiatric conditions and deep learning-based image generation engine to provide a visual summary. The system incorporates temporal modeling to provide a time-series representation of topic similarities at a turn-level resolution and AI-generated artworks given the dialogue segments to provide a concise representations of the contents covered in the session, offering interpretable insights for therapists to optimize their strategies and enhance the effectiveness of psychotherapy. This system provides a proof of concept of AI-augmented therapy tools with e in-depth understanding of the patient's mental state and enabling more effective treatment. (Video demonstrating the system: <https://youtu.be/TdE5FR4i550>)

Keywords

Health AI, Computational psychiatry, Generative AI, Psychotherapy

1. Introduction

Mental health is a global issue affecting people of all ages, cultures, and countries. With the increasing demand for mental health services, innovative solutions are needed to address the workload on mental health providers. One promising area of innovation is natural language processing (NLP), which has been adopted in psychotherapy to help therapists provide better care [1, 2, 3]. Previous works have demonstrated the effectiveness of classical topic modeling in mental illness detection [4, 5], but recent advancements in deep learning have led to the emergence of neural network-based topic modeling methods as a better solution [6].

In this paper, we propose TherapyView, a data visualization system that uses neural topic modeling to learn the topical propensities of different psychiatric conditions from psychotherapy session transcripts. We conducted an empirical evaluation of existing neural topic modeling techniques with a focus on their application to the domain of psychotherapy, benchmarked on the Alexander Street Counseling and Psychotherapy Transcripts dataset. By leveraging temporal modeling of state-of-the-art topic models, we provide a visual representation of the topical tendencies of psychiatric conditions, enabling therapists to easily identify patterns and make informed decisions about their psychotherapy strategies.

Moreover, we introduce AI-generated arts on different temporal segments of the therapy sessions, providing a concise visual summary of the session. The user-friendly interface and interactive visualizations make it easy for therapists to understand and interpret the results, leading to improved treatment outcomes for patients. Our data visualization system is a powerful tool for advancing the field of psychotherapy and providing therapists with the information they need to make informed decisions.

2. The Temporal Topic Modeling Framework

Joint Proceedings of the ACM IUI Workshops 2025, March 24-27, 2025, Cagliari, Italy

*Corresponding author.

✉ baihan.lin@mssm.edu (B. Lin); szecevic@ibm.com (S. Zecevic); djallel.bouneffouf@ibm.com (D. Bouneffouf); gcecchi@us.ibm.com (G. Cecchi)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

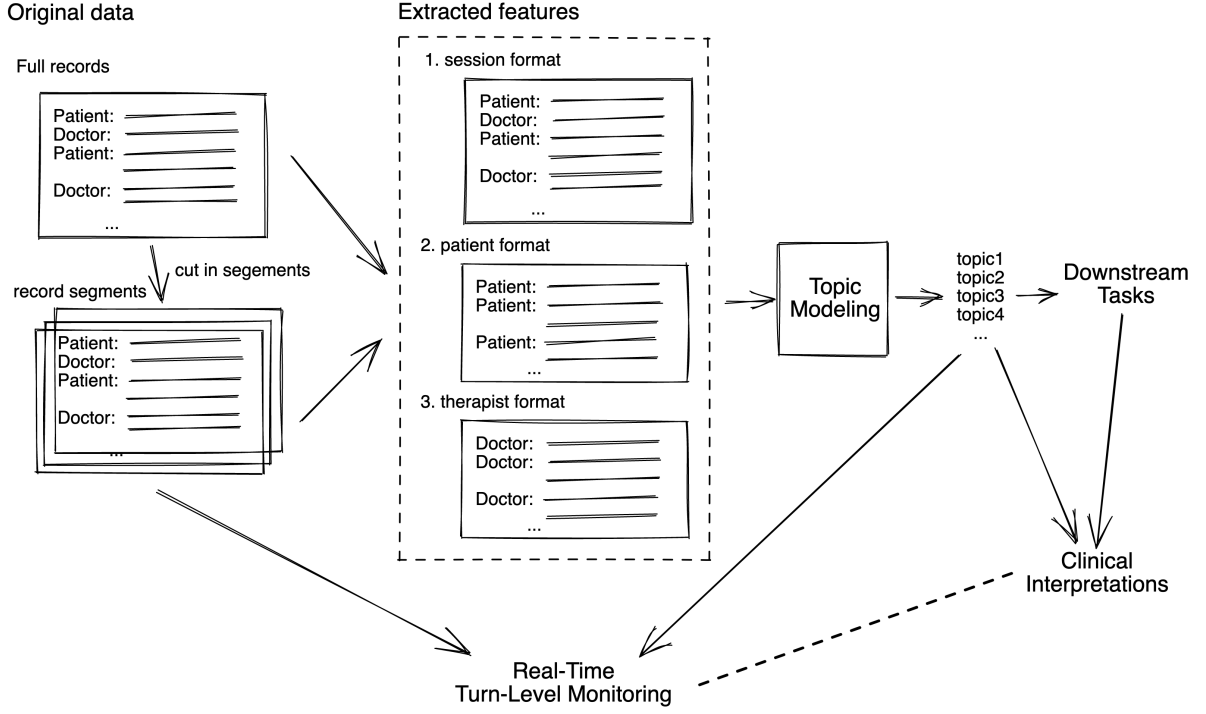


Figure 1: Psychotherapy topic modeling framework

Algorithm 1 Temporal Topic Modeling (TTM)

- 1: Learned topics T as references
 - 2: **for** $i = 1, 2, \dots, N$ **do**
 - 3: Automatically transcribe dialogue turn pairs (S_i^p, S_i^t)
 - 4: **for** $T_j \in \text{topics } T$ **do**
 - 5: Topic score $W_j^{p_i} = \text{similarity}(\text{Emb}(T_j), \text{Emb}(S_i^p))$
 - 6: Topic score $W_j^{t_i} = \text{similarity}(\text{Emb}(T_j), \text{Emb}(S_i^t))$
 - 7: **end for**
 - 8: **end for**
-

Our proposed analytical framework, outlined in Figure 1, leverages natural language processing (NLP) techniques and neural topic modeling to extract valuable insights from psychotherapy session transcripts. During the session, the dialogue between the patient and therapist is transcribed into pairs of turns, which are then used as the input data for our framework. We can take the full records of a patient or a cohort of patients belonging to the same condition, which can either be used as is before feature extraction or truncated into segments based on timestamps or topic turns.

We extract features from the transcript data using NLP techniques and fit them into neural topic models to generate a list of weighted topic words. These topic words provide important insights into the patient’s condition and are often highly interpretable, making them valuable in the context of psychotherapy.

Our framework offers a range of downstream tasks and user scenarios. We can use the extracted weighted topics to assess the progress of the therapy, identify potential issues in the patient’s mental state, or suggest adjustments to the therapist’s treatment strategies. These features can be incorporated into an intelligent AI assistant to help remind the therapist of important information during the session. Additionally, certain taboo topics such as those related to suicidal conversations can be flagged for the therapist’s attention.

To further analyze the transcript data, we use temporal topic modeling (TMM) to compute turn-

Table 1

Topic evaluations of the neural topic models

	Anxiety		Depression		Schizophrenia	
	TC	TD	TC	TD	TC	TD
NVDM-GSM [8]	0.653	-380.933	0.487	-316.439	0.527	-431.393
WTM-MMD [9]	0.927	-453.929	0.907	-359.964	0.447	-403.694
ETM [10]	0.893	-449.000	0.933	-367.069	0.973	-310.211
BATM [11]	0.720	-441.049	0.773	-443.394	0.500	-337.825

resolution topic scores. Algorithm 1 outlines the pipeline of our TMM analysis. For example, if we have learned ten topics, the topic score will be a ten-dimensional vector, with each dimension corresponding to a likelihood of the turn being in that topic. To account for the directional property of each turn with respect to a given topic, we compute the cosine similarity between the embedded topic vector and the embedded turn vector, instead of directly inferring the probability as in traditional topic assignment problems. The Embedded Topic Model (ETM), which we use for temporal modeling in the results section, also models each word with a categorical distribution whose natural parameter is the inner product between a word embedding and an embedding of its assigned topic. We use Word2Vec [7] as our word embedding for both the topics and the turns.

3. The Empirical Evaluations

To evaluate the performance of our proposed neural topic modeling approach, we compare four state-of-the-art models and analyze their learned topics on our dataset [8, 9, 10, 11]. We separate the transcript sessions into three categories based on the psychiatric conditions of the patients (anxiety, depression, and schizophrenia), and train the topic models for over 100 epochs at a batch size of 16. As with standard preprocessing for topic modeling, we set the lower bound of the word count to keep in topic training to be 3, and the ratio of the upper bound of the word count to keep in topic training to be 0.3. We evaluate the models using a series of validated measurements of topic coherence and diversity, as outlined in [12]. Specifically, we use an asymmetrical confirmation measure between top word pairs (smoothed conditional probability) for topic coherence and the ratio between the size of the vocabulary in the topic words and the total number of words in the topics for topic diversity.

Table 1 summarizes the evaluation results for the four models across the different psychiatric conditions, based on validated measurements of topic coherence and diversity proposed in [12]. We observe that the Embedded Topic Model (ETM) yields relatively high topic coherence and diversity across all three psychiatric conditions in the Alex Street datasets, making it a suitable choice for our deployed system. For the full evaluation results, see [6].

To ensure that the learned topics can be mapped from one clinical condition to another, we compute a universal topic model on the text corpus of the entire Alex Street psychotherapy database. Using this universal topic model, we compute a 10-dimensional topic score for each turn, corresponding to the 10 topics. The higher the score, the more positively correlated the turn is with the topic. This time-series matrix enables us to probe the dynamics of the dialogues within the topic space (e.g. visualized as a 3d trajectory in our demonstration system, TherapyView).

To provide interpretable insights, it is important to parse out the concepts behind the learned topics. To better understand the topics, we parse out the highest-scoring turns in the transcripts that correspond to each topic. For example, topic 0 is about figuring out self-discovery and reminiscence, while topic 1 is about play. Topic 2 is about anger, fear, and sadness, while topic 3 is about counts. Topic 4 is about tiredness and decision-making, while topic 5 is about sickness, self-injuries, and coping mechanisms. Topic 6 is about explicit ways to deal with stress, such as keeping busy and reaching out for help, while topic 7 is about numbers. Topic 8 is about continuation and perseverance, while topic 9 is mostly chit-chat, interjections, and transcribed prosody.



Figure 2: Screenshot of the launch page of TherapyView dashboard

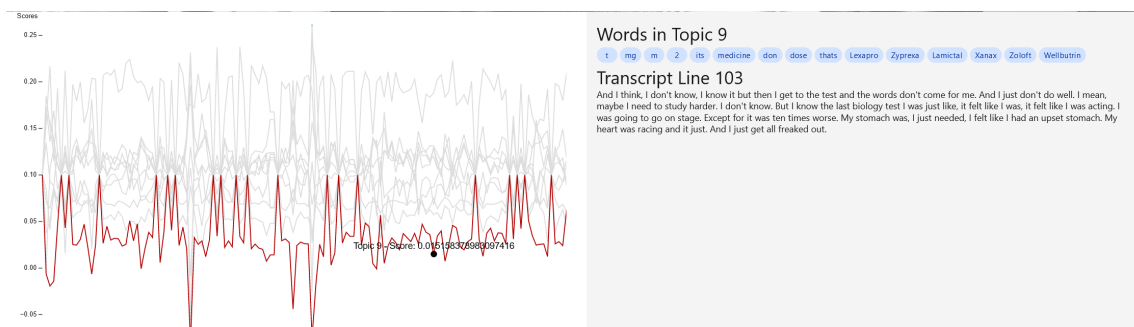


Figure 3: Screenshot of the highlighted details in the web dashboard

4. TherapyView Demonstration System

4.1. The web dashboard

These metrics and insights are visualized in a web dashboard (Figure 2). It consists of two parts: a Jupyter notebook that generates and serves the data, and a visual interactive dashboard that displays the data. There are four different visualizations in the dashboard:

- Images as a visual summary of this therapy session, powered by OpenAI's DALL-E 2 API.
- Line graph of topical tendency over time (as the dialogue turns), which is highlightable for more details (Figure 3).

- 3D plot showing the relationship of the selected three out of the ten topics over time.
- Full readout of the transcript (which can be replaced with user-specified inputs).

Each AI-generated image represents a single 1,000 character excerpt from the loaded transcript. These images act as a visual timeline, potentially surfacing notable changes in the patient during a session. The vague nature of these images is supplemented by the numerical data provided by the neural topic model. The therapist can explore each of the topics in detail through the charts described above. If the therapist finds a topic score change of interest, they can retrieve the corresponding line in the transcript and analyze the raw text.

This dashboard allows therapists to identify elements of concern by presenting them visually. By quickly identifying these elements, a therapist can provide the appropriate treatment in a timely fashion. These visualizations may also help surface behaviors that may remain un-noticed by the therapist without the help of the dashboard.

4.2. The system architecture

The system architecture of the dashboard consists of two components: an API and a web application. The API is a single Jupyter notebook written in Python. This notebook contains all the logic for generating the visualizations in the dashboard. The “Jupyter Kernel Gateway” package turns each cell into an API endpoint. The web component is a React single page application that queries the API for the data, displays it, and adds interactivity.

This separation of concerns allows scientists to quickly iterate and experiment. Jupyter notebooks are generally understood by AI researchers, and it does not require any special knowledge for them to add new features to the API.

This is a design optimized for rapid prototyping and experimentation. It is not production ready. Commercialization of this dashboard will require that the Jupyter notebook be replaced with a more robust, efficient, and permanent solution.

4.3. The limitations of the AI image generation

Out of all the visualizations on the dashboard, the generated images are of special interest. Every refresh of the dashboard generates a new set of images, making the results unpredictable. This was included because novel AI approaches like DALL-E, even if they are imprecise, have the potential to provide new perspectives for a therapist to consider.

Integrating DALL-E with real-world therapy does have some challenges: (1) the API only allows a maximum of 1,000 characters per image request. This means that DALL-E cannot use any context outside of small chunks, which may limit the kind of insights that it has the potential to visualize; (2) in the demo, a number of prompts were rejected by DALL-E for “safety” reasons. OpenAI prevents certain topics to be visualized for ethical reasons. Psychotherapy sessions can involve many sensitive topics and harmful behaviors. Further development of this approach will require many safeguards to ensure privacy and ensure ethical use.

5. Ethical Statements

Till now, there is still a severe global shortage of workforce in mental health [13]. What we are proposing here, is not to replace existing work force of psychiatrists or therapists, but to assist them. In education setting, having an interpretable model that can inform the next-generation psychiatrists about the strategies adopted by experienced therapists. In this way, it can potentially alleviate this societal issue in both assisting and educating junior psychiatrists.

As more and more successful applications of AI are deployed in clinical domains, there are many ethical considerations we practitioners of machine learning should be aware of and take into consideration. When dealing with patient data, the privacy and security is a top priority. Following the suggestion

of best practices from [14], all examples in this paper as well as the dataset we analyzed are properly anonymized with pre- and post-processing techniques. In addition, the dataset itself was sourced with proper license from ProQuest’s Alexander Street platform. We remove all personally identifiable information (meta data, user name, identifiers, doctors’ name) from the dataset.

Since the clinical domain of this work is mental health and psychological well-being, there are additional ethical considerations. Emerging techniques in wearable devices, digital health records, brain imaging measurements, smartphone applications and social media are gradually transforming the landscape of the monitoring and treatment of mental health illness. However, most of these attempts are proof of concept as identified by this review [15], and requires extensive caution to prevent from the pitfall of over-interpreting preliminary results. The limitations of these prior studies, including our work here, reside in the difficulty of a systematic clinical validation and a uncertain future expectation of the technological readiness for patient care and therapeutic decision making approved by authorities. For instance, it was recently shown that despite the high predictability of statistical learning-based methods in analyzing large datasets in support of clinical decisions in psychiatry, existing machine learning solutions is highly susceptible to overfitting in realistic tasks which has usually a small sample sizes in the data, missing data points for some subjects, and highly correlated variables [16]. These properties in real-world applications limits the out-of-sample generalizability of the results.

Another ethical boundary to maintain is to make sure that the AI systems we use to diagnose, interpret and predict the mental health illness don’t lead to increased risks to the patients. This requires both the practioners and machine learning researcher to fully aware of potentially bias and ethical challenges, such as gender bias, language-related ambiguity and ethnicity-related mental illness connections [17], in order to deploy the AI system in a responsible and safe way. One way of promoting this awareness is to engage the public in discussions about the usage of AI in mental health and well beings [18].

In our work, we analyze a dataset with over 950 sessions of psychotherapy transcripts. Although it is the biggest dataset we find in this research domain, due to the anonymous nature of the dataset and a lack of details behind the collection process and demographics of the transcription, we cannot guarantee that the generalizability and representativeness to all populations. However, we believe that the insights we gain from these interpretable investigations is unlikely to increase the unforeseeable risks of the patients involved and can be potentially useful.

6. Conclusions

In conclusion, this data visualization demonstration system presents a visual journey through the doctor-patient dialogues in therapy sessions via temporal topic modeling and image generation. The results of our study show that the Embedded Topic Model yields high topic coherence and diversity, making it a strong candidate for use in this domain. Our incorporation of temporal modeling and interactive modules on the web dashboard provide additional interpretability, allowing therapists to better understand the progression of psychiatric conditions over time. The use of AI-generated artworks further enhances the interpretability of the results, providing therapists with a visual representation of the core themes of a given therapy session. The results of this study provide valuable insights into the session trajectories of patients and therapists and have the potential to improve the effectiveness of psychotherapy. This is just the first step in a potential turn-level resolution temporal analysis of topic modeling, and we look forward to further exploring this area in future studies.

In future work, we plan to use the learned topic scores to predict psychological or therapeutic states with other digital traces [19, 20]. Additionally, we aim to train chatbots as reinforcement learning agents using these states, incorporating biological and cognitive priors [21, 22, 23, 24], and studying their factorial relations with other inference anchors, such as working alliance and personality [25, 26, 27]. Our ultimate goal is to construct a complete AI knowledge management system for mental health, utilizing different NLP annotations in real-time (as in review [28]), and drive AI-augmented therapy sessions [29, 30].

In conclusion, our proposed TherapyView system represents a novel approach to psychotherapy,

leveraging the latest advancements in deep learning and data visualization to help therapists provide better care for their patients. The use of NLP and AI-generated arts in our system enables therapists to quickly identify patterns in patient data and tailor their treatment strategies accordingly. With further development and testing, TherapyView has the potential to revolutionize the field of psychiatry and improve patient outcomes.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools other than the proposed approach using Generative AI tools.

References

- [1] H.-Y. Shum, X.-d. He, D. Li, From eliza to xiaoice: challenges and opportunities with social chatbots, *Frontiers of Information Technology & Electronic Engineering* 19 (2018) 10–26.
- [2] M. T. ZEMČÍK, A brief history of chatbots, *DEStech Transactions on Computer Science and Engineering* 10 (2019).
- [3] N. Rezaei, P. Wolff, B. H. Price, Natural language processing in psychiatry: the promises and perils of a transformative approach, *The British Journal of Psychiatry* (2022) 1–3.
- [4] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, J. Boyd-Graber, Beyond lda: exploring supervised topic modeling for depression-related language in twitter, in: *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015, pp. 99–107.
- [5] Q. T. Zeng, D. Redd, T. Rindflesch, J. Nebeker, Synonym, topic model and predicate-based query expansion for retrieving clinical documents, in: *AMIA Annual Symposium Proceedings*, volume 2012, American Medical Informatics Association, 2012, p. 1050.
- [6] B. Lin, D. Bouneffouf, G. Cecchi, R. Tejwani, Neural topic modeling of psychotherapy sessions, in: *International Workshop on Health Intelligence*, Springer, 2023.
- [7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems* 26 (2013).
- [8] Y. Miao, E. Grefenstette, P. Blunsom, Discovering discrete latent topics with neural variational inference, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 2410–2419.
- [9] F. Nan, R. Ding, R. Nallapati, B. Xiang, Topic modeling with wasserstein autoencoders, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 6345–6381.
- [10] A. B. Dieng, F. J. Ruiz, D. M. Blei, Topic modeling in embedding spaces, *Transactions of the Association for Computational Linguistics* 8 (2020) 439–453.
- [11] R. Wang, X. Hu, D. Zhou, Y. He, Y. Xiong, C. Ye, H. Xu, Neural topic modeling with bidirectional adversarial training, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 340–350.
- [12] M. Röder, A. Both, A. Hinneburg, Exploring the space of topic coherence measures, in: *Proceedings of the eighth ACM international conference on Web search and data mining*, 2015, pp. 399–408.
- [13] M. Olfson, Building the mental health workforce capacity needed to treat adults with serious mental illnesses, *Health Affairs* 35 (2016) 983–990.
- [14] T. Matthews, K. O’Leary, A. Turner, M. Sleeper, J. P. Woelfer, M. Shelton, C. Manthorne, E. F. Churchill, S. Consolvo, Stories from survivors: Privacy & security practices when coping with intimate partner abuse, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2189–2201.
- [15] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, D. V. Jeste, Artificial intelligence for mental health and mental illnesses: an overview, *Current psychiatry reports* 21 (2019) 1–18.

- [16] R. Iniesta, D. Stahl, P. McGuffin, Machine learning, statistical learning and the future of biological research in psychiatry, *Psychological medicine* 46 (2016) 2455–2465.
- [17] I. Y. Chen, P. Szolovits, M. Ghassemi, Can ai help reduce disparities in general medical and mental health care?, *AMA journal of ethics* 21 (2019) 167–179.
- [18] S. Carr, ‘ai gone mental’: engagement and ethics in data-driven technology for mental health, 2020.
- [19] B. Lin, D. Bouneffouf, G. Cecchi, Predicting human decision making in psychological tasks with recurrent neural networks, *PloS one* 17 (2022) e0267907.
- [20] B. Lin, D. Bouneffouf, G. Cecchi, Predicting human decision making with lstm, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022.
- [21] B. Lin, G. Cecchi, D. Bouneffouf, J. Reinen, I. Rish, A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry, in: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, 2020, pp. 744–752.
- [22] B. Lin, G. Cecchi, D. Bouneffouf, J. Reinen, I. Rish, Unified models of human behavioral agents in bandits, contextual bandits and rl, *arXiv preprint arXiv:2005.04544* (2020).
- [23] B. Lin, D. Bouneffouf, G. Cecchi, Split Q Learning: Reinforcement Learning with Two-Stream Rewards, in: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, AAAI Press, International Joint Conferences on Artificial Intelligence Organization, 2019, pp. 6448–6449. URL: <https://doi.org/10.24963/ijcai.2019/913>. doi:10.24963/ijcai.2019/913.
- [24] B. Lin, G. Cecchi, D. Bouneffouf, J. Reinen, I. Rish, Models of human behavioral agents in bandits, contextual bandits and rl, in: *International Workshop on Human Brain and Artificial Intelligence*, Springer, 2021, pp. 14–33.
- [25] B. Lin, G. Cecchi, D. Bouneffouf, Deep annotation of therapeutic working alliance in psychotherapy, in: *International Workshop on Health Intelligence*, Springer, 2023.
- [26] B. Lin, G. Cecchi, D. Bouneffouf, Working alliance transformer for psychotherapy dialogue classification, *arXiv preprint arXiv:2210.15603* (2022).
- [27] B. Lin, Personality effect on psychotherapy outcome: A predictive natural language processing framework, *arXiv preprint* (2022).
- [28] B. Lin, Knowledge management system with nlp-assisted annotations: A brief survey and outlook, in: *CIKM Workshops*, 2022.
- [29] B. Lin, G. Cecchi, D. Bouneffouf, Supervisorbot: Nlp-annotated real-time recommendations of psychotherapy treatment strategies with deep reinforcement learning, *arXiv preprint arXiv:2208.13077* (2022).
- [30] B. Lin, G. Cecchi, D. Bouneffouf, Helping therapists with nlp-annotated recommendation, in: *Joint Proceedings of the ACM IUI Workshops*, 2023.