SupervisorBot: NLP-Annotated Real-Time Recommendations of Psychotherapy Treatment Strategies with Deep Reinforcement Learning

Baihan Lin[1*], **Guillermo Cecchi**[2], **Djallel Bouneffouf**[2]
[1] Columbia University, New York, NY 10027, USA
[2] IBM TJ Watson Research Center, Yorktown Heights, NY 10598, USA
baihan.lin@columbia.edu, gcecchi@us.ibm.com, djallel.bouneffouf@ibm.com

Abstract

We present a novel recommendation system designed to provide real-time treatment strategies to therapists during psychotherapy sessions. Our system utilizes a turn-level rating mechanism that forecasts the therapeutic outcome by calculating a similarity score between the profound representation of a scoring inventory and the patient's current spoken sentence. By transcribing and segmenting the continuous audio stream into patient and therapist turns, our system conducts immediate evaluation of their therapeutic working alliance. The resulting dialogue pairs, along with their computed working alliance ratings, are then utilized in a deep reinforcement learning recommendation system. In this system, the sessions are treated as users, while the topics are treated as items. To showcase the system's effectiveness, we not only evaluate its performance using an existing dataset of psychotherapy sessions but also demonstrate its practicality through a web app. Through this demo, we aim to provide a tangible and engaging experience of our recommendation system in action.

1 Introduction

The prevalence of mental illness poses a significant healthcare challenge not only in the United States, with an estimated 1 in 5 individuals affected according to the National Institute of Mental Health (NIMH) [Patel *et al.*, 2018], but also globally. However, the existing education systems and training programs struggle to keep up with this growing trend. Becoming a licensed therapist requires years of continuous learning and supervised training, making it difficult for these systems to adapt and produce an adequate number of qualified professionals in a timely manner. Even after therapists have completed their training and are ready for independent practice, many still seek regular supervision from more experienced therapists known as "supervisors." These supervisors play a crucial role in the development of junior therapists, as they have accumulated extensive experience working with patients over the years. They serve as essential mentors, providing guidance and periodic feedback to help address the challenges and obstacles that beginner therapists encounter in their training [Watkins Jr, 2013]. This supervision is essential in shaping the psychotherapist identities of junior therapists and assisting them in overcoming the hurdles they face.

In this research endeavor, we introduce SupervisorBot, an innovative virtual AI companion designed to provide therapists with real-time feedback and treatment recommendations during their psychotherapy sessions. Similar to a human supervisor, SupervisorBot offers case-specific guidance and feedback. Leveraging its access to an extensive database of historical therapy sessions and case studies conducted by experienced therapists, SupervisorBot draws on this wealth of knowledge to provide valuable insights. The foundation of our recommendation system is a rating system that assesses the effectiveness of different treatment strategies. Recognizing the complexities involved in characterizing a patient's mental state, we focus on well-defined clinical outcomes. One such concept is the working alliance, a psychological construct that has showed high predictive value for the success of psychotherapy in clinical settings [Wampold, 2015].

Introducing the Reinforced Recommendation model for Dialogue topics in psychiatric Disorders (R2D2), we present the first-ever recommendation system of dialogue topics tailored specifically for the psychotherapy domain. R2D2 operates by transcribing the therapy session in real-time, predicting the therapeutic outcome at a turn-by-turn level, and recommending the most suitable treatment strategies based on the current context and the state of the ongoing session.

2 Analytical Framework

The analytic framework used in this study is outlined in Figure 1. Initially, the continuous audio stream is processed through speaker diarization, employing real-time solutions such as [Lin and Zhang, 2021; Lin and Zhang, 2020b; Lin and Zhang, 2020a]. This step separates the audio into doctor-patient dyads, which are then transcribed into natural language turns for subsequent real-time analyses. In the transcription and real-time rating assessment phase, the diarized audio stream is transcribed using a standard automatic speech recognition module [Adorf, 2013]. The conversation is then evaluated based on the Working Alliance Inventory (WAI), a 36-statement questionnaire that measures the therapeutic bond, task agreement, and goal agreement [Horvath, 1981; Tracey and Kokotovic, 1989; Martin *et al.*, 2000]. By embedding the dialogue turns and WAI items using deep

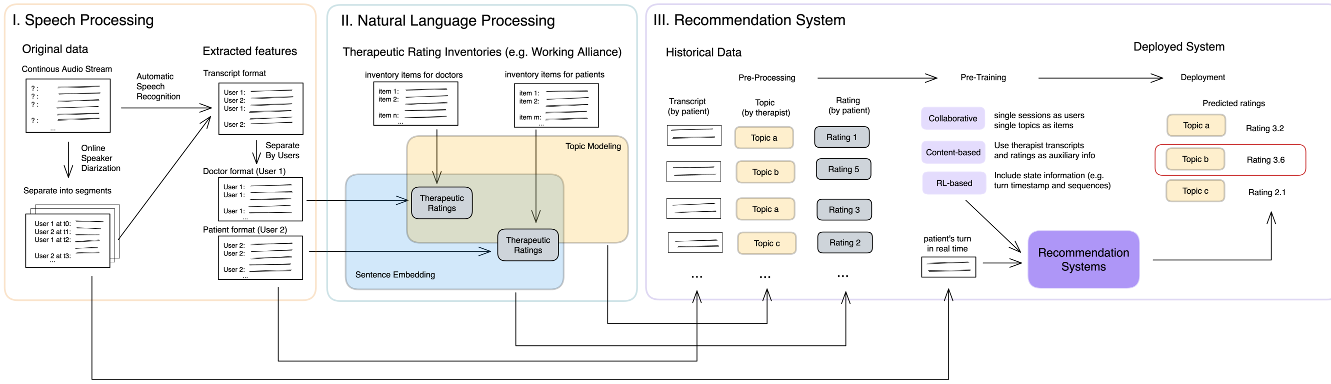

Figure 1: Analytical Framework of SupervisorBot.

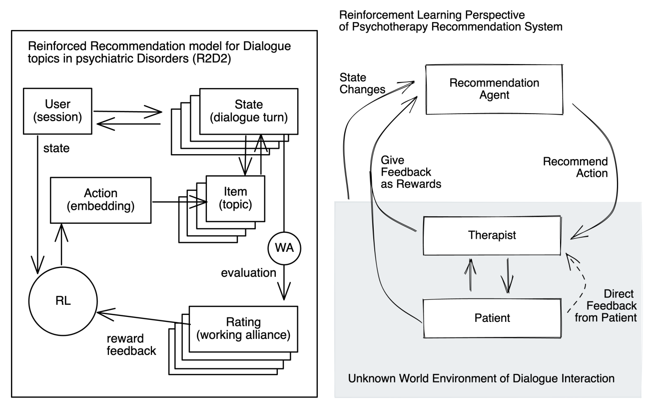

Figure 2: The R2D2 RL framework of psychotherapy recommender.

sentence or paragraph embeddings (e.g., Doc2Vec [Le and Mikolov, 2014]), and computing cosine similarity between the turn embeddings and corresponding inventory vectors, a 36-dimension working alliance score is obtained for each turn [Lin *et al.*, 2023b; Lin *et al.*, 2022b]. Next, topic modeling is employed to identify treatment strategy recommendations. Using the Embedded Topic Model (ETM) [Wang *et al.*, 2020], the main concepts discussed in psychotherapy sessions are extracted [Lin *et al.*, 2023a]. In this study, each turn is annotated with its most likely topic, resulting in seven unique topics such as self-discovery, play, and emotions. To establish the recommendation system, the treatment strategies (topics) are paired with users, contents, and ratings. In this case, the session IDs are assigned as users. Content-based and collaborative filtering approaches, as well as reinforcement learning (RL) and session-based approaches, are employed to create the recommendation engine [Pazzani and Billsus, 2007; Basu *et al.*, 1998; Aggarwal and others, 2016; Sarwar *et al.*, 2001; He *et al.*, 2017; Koren *et al.*, 2022; Su and Khoshgoftaar, 2009; Zheng *et al.*, 2018; Wang *et al.*, 2014; Zou *et al.*, 2020; Li *et al.*, 2017; Wu *et al.*, 2019; Ludewig and Jannach, 2018]. The system treats each session as a new "user" during deployment. Deep reinforcement learning techniques are applied to the recommendation system [Lin, 2022c], with three popular algorithms evaluated: Deep Deterministic Policy Gradients (DDPG) [Lillicrap *et al.*, 2015], Twin Delayed DDPG (TD3) [Fujimoto *et al.*, 2018], and Batch Constrained Q-Learning (BCQ) [Fujimoto *et al.*, 2019]. These algorithms enable the recommendation agent to make informed suggestions based on dialogue interactions and quality evaluations. Combining all the components, the proposed model is named Reinforced Recommendation model for Dialogue topics in psychiatric Disorders (R2D2). Each session is treated as a user, and the dialogue turns serve as states with real-time topic labels and working alliance ratings. The core of the model is a deep reinforcement learning framework that predicts the best action (topic) based on embeddings, which can be pre-computed using dimension reduction techniques. The dialogue continues based on the selected action, and the reward is computed using the working alliance rating. Figure 2 illustrates the R2D2 model.

	Anxi	Depr	Schi	Suic	All
R2D2-DDPG-TASK	**0.3796**	0.3376	0.1556	0.3292	0.0578
R2D2-DDPG-BOND	0.2417	0.3838	0.1539	0.0873	0.1455
R2D2-DDPG-GOAL	0.0761	0.3682	0.4589	-0.0210	0.2243
R2D2-TD3-TASK	0.0707	0.1310	0.0443	0.3188	0.3357
R2D2-TD3-BOND	0.2018	0.3363	0.0908	0.3070	0.1101
R2D2-TD3-GOAL	0.0984	0.2222	**0.4599**	0.2044	**0.3765**
R2D2-BCQ-TASK	0.1128	**0.4042**	0.1401	0.1422	0.0825
R2D2-BCQ-BOND	0.0778	0.0876	0.0987	**0.4152**	0.0885
R2D2-BCQ-GOAL	0.0810	0.1231	0.0833	0.0788	0.0780

Table 1: Pearson's r of the actual actions and predicted actions

Figure 3: State screenshots of SupervisorBot web app: inventory inputs, diarization, annotation, strategy recommendation.

3 Performance Evaluation

The performance evaluation of the recommendation systems is conducted on the Alex Street psychotherapy dataset [Street, 2023], consisting of over 950 therapy sessions across various psychiatric conditions. The dataset is preprocessed into a recommendation system format, with train-test splits, and utilized to train the R2D2 model using different reinforcement learning agents (DDPG, TD3, and BCQ) for 50 epochs. The action spaces (topics to recommend) are represented using different embeddings: Doc2Vec, PCA, and UMAP. The evaluation focuses on Pearson's correlation coefficient between the recommended actions and ground truth actions in test set. Results show that the best performing models vary based on the psychiatric condition and the working alliance scale used for ratings. For depression sessions, R2D2-DDPG-TASK and R2D2-BCQ-TASK achieve correlation coefficients of 0.3796 and 0.4042, respectively. R2D2-TD3-GOAL performs best for schizophrenia sessions with a correlation coefficient of 0.4599, while R2D2-BCQ-BOND performs best for suicidal sessions with a correlation coefficient of 0.4152. When considering all four psychiatric conditions together, R2D2-TD3-GOAL emerges as the top-performing model with a correlation coefficient of 0.3765 [Lin, 2022b]. Comparing the three RL algorithms, DDPG and TD3 consistently yield similar rankings across the three working alliance scales, while BCQ exhibits some differences. The choice of the working alliance scale impacts the recommendation predictions, with the goal scale showing more advantageous results in schizophrenia cases. However, this effect is less pronounced in R2D2-BCQ. In specific disorders, R2D2-DDPG performs well for anxiety, depression, and schizophrenia, while R2D2-TD3 is effective for suicidal cases. It is important to note that the limited amount of data for suicidal cases requires cautious interpretation. This evaluation serves as a proof of concept.

4 Important Ethical Considerations

Developing and deploying recommendation systems for psychotherapy sessions necessitates careful attention to ethical considerations. Following the discussions in [Lin, 2022a], we considered the following additional key points: (1) *Privacy and confidentiality*: Safeguard patient information throughout storage, transmission, and access, adhering to relevant data protection regulations; (2) *Informed consent and autonomy*: Obtain patient consent and ensure they are well-informed about the system's purpose, benefits, and risks, allowing them to choose participation and withdrawal; (3) *Algorithmic bias and fairness*: Mitigate biases in data and algorithms to prevent unequal treatment, discrimination, or perpetuation of stereotypes; (4) *Transparency and explainability*: Maintain transparency in system operations, providing explanations for recommendations to enhance understanding; (5) *Professional judgment and human oversight*: Emphasize the doctors' role, using the system's recommendations as augmentation rather than replacement; (6) *Continuous evaluation and improvement*: Regularly assess and enhance the system's performance, addressing concerns, biases, and limitations through feedback and research. As part of our future endeavor, we will further explore these ethical topics in recommendation systems in healthcare settings. By adhering to these ethical considerations, we can maximize the benefits of recommendation systems while prioritizing patient well-being, privacy, and autonomy.

5 Demonstration System

SupervisorBot is an interactive web-based system (Figure 3). To begin, users are provided with clear instructions on how to utilize the system. They are then guided to input their own inventory, which is used to evaluate the quality of the dialogue. In our case, we provide a default inventory utilizing the working alliance inventory. Users are prompted to input the corresponding score scale for each item in the inventory and can finalize by clicking on the "Analyze" button. Moving to the speaker diarization component, we employ a sliding window approach to compute and visualize the Mel Frequency Cepstral Coefficients (MFCC) in real-time, using microphone audio input. The MFCC bands are color-coded on the web page, allowing for a visual representation. The annotation panel is where the therapist can access the displayed transcript, indicating the speaker for each dialogue turn. Additionally, the computed alliance score, in three different scales, is dynamically presented in real-time, providing valuable information to assist the therapist. In the final panel, our recommendation guidance is provided. The available topics are ranked, and the top N recommendations are displayed. The therapist can utilize this information as a hint and initiate their response based on the top recommendation. The system transcribes the therapist's response and highlights the most likely chosen topic from the previous round. This information is saved as part of the historical data. At the end of the session, the system refreshes its parameters. Overall, the interactive web-based system of *SupervisorBot* offers therapists a comprehensive platform to enhance their psychotherapy sessions.

6 Conclusions

In this study, we offer a practical demonstration of the potential of a real-time recommendation system to enhance the effectiveness of psychotherapy sessions by providing therapists with informative clinical annotations and treatment strategy recommendations using deep reinforcement learning techniques. While our example focuses on recommending topics for therapists to initiate or continue, the same approach can be extended to encompass more intricate and nuanced treatment suggestions. As a next step, we aim to predict these inference anchors as states [Lin *et al.*, 2023a; Lin *et al.*, 2022a] and train chatbots as RL agents based on these states, building upon research such as [Lin *et al.*, 2020a; Lin *et al.*, 2021; Lin *et al.*, 2020b]. Additionally, we plan to equip clinicians with real-time visualizations of NLP-annotated insights [Lin, 2022d; Lin *et al.*, 2023e]. Moreover, we strive to devise disorder-specific policies accompanied by interpretable insights [Lin *et al.*, 2023d; Lin *et al.*, 2023c]. Lastly, we seek to generate a knowledge management system that incorporates these on-the-fly analytics [Lin, 2022b]. By pursuing these avenues of research, we aim to create a comprehensive framework that empowers therapists with cutting-edge tools and resources to enhance psychotherapy practice.

References

[Adorf, 2013] Julius Adorf. Web speech api. *KTH Royal Institute of Technology*, pages 1–11, 2013.

[Aggarwal and others, 2016] Charu C Aggarwal et al. *Recommender systems*, volume 1. Springer, 2016.

[Basu *et al.*, 1998] Chumki Basu, Haym Hirsh, William Cohen, et al. Recommendation as classification: Using social and content-based information in recommendation. In *Aaai/iaai*, pages 714–720, 1998.

[Fujimoto *et al.*, 2018] Scott Fujimoto, Herke Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, pages 1587–1596. PMLR, 2018.

[Fujimoto *et al.*, 2019] Scott Fujimoto, David Meger, and Doina Precup. Off-policy deep reinforcement learning without exploration. In *International conference on machine learning*, pages 2052–2062. PMLR, 2019.

[He *et al.*, 2017] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.

[Horvath, 1981] Adam O Horvath. *An exploratory study of the working alliance: Its measurement and relationship to therapy outcome*. PhD thesis, University of British Columbia, 1981.

[Koren *et al.*, 2022] Yehuda Koren, Steffen Rendle, and Robert Bell. Advances in collaborative filtering. *Recommender systems handbook*, pages 91–142, 2022.

[Le and Mikolov, 2014] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR, 2014.

[Li *et al.*, 2017] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1419–1428, 2017.

[Lillicrap *et al.*, 2015] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[Lin and Zhang, 2020a] Baihan Lin and Xinxin Zhang. Speaker diarization as a fully online learning problem in minivox. *arXiv preprint arXiv:2006.04376*, 2020.

[Lin and Zhang, 2020b] Baihan Lin and Xinxin Zhang. Voiceid on the fly: A speaker recognition system that learns from scratch. In *INTERSPEECH*, 2020.

[Lin and Zhang, 2021] Baihan Lin and Xinxin Zhang. Speaker diarization as a fully online bandit learning problem in minivox. In *ACML*, pages 1660–1674, 2021.

[Lin *et al.*, 2020a] Baihan Lin, Guillermo Cecchi, Djallel Bouneffouf, Jenna Reinen, and Irina 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*, pages 744–752, 2020.

[Lin *et al.*, 2020b] Baihan Lin, Guillermo Cecchi, Djallel Bouneffouf, Jenna Reinen, and Irina Rish. Unified models of human behavioral agents in bandits, contextual bandits and rl. *arXiv preprint arXiv:2005.04544*, 2020.

[Lin *et al.*, 2021] Baihan Lin, Guillermo Cecchi, Djallel Bouneffouf, Jenna Reinen, and Irina Rish. Models of human behavioral agents in bandits, contextual bandits and rl. In *International Workshop on Human Brain and Artificial Intelligence*, pages 14–33. Springer, 2021.

[Lin *et al.*, 2022a] Baihan Lin, Djallel Bouneffouf, and Guillermo Cecchi. Predicting human decision making in psychological tasks with recurrent neural networks. *PloS one*, 2022.

[Lin *et al.*, 2022b] Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. Working alliance transformer for psychotherapy dialogue classification. *arXiv preprint arXiv:2210.15603*, 2022.

[Lin *et al.*, 2023a] Baihan Lin, Djallel Bouneffouf, Guillermo Cecchi, and Ravi Tejwani. Neural topic modeling of psychotherapy sessions. In *International Workshop on Health Intelligence*. Springer, 2023.

[Lin *et al.*, 2023b] Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. Deep annotation of therapeutic working alliance in psychotherapy. In *International Workshop on Health Intelligence*. Springer, 2023.

[Lin *et al.*, 2023c] Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. Helping therapists with nlp-annotated recommendation. In *Joint Proceedings of the ACM IUI Workshops*, 2023.

[Lin *et al.*, 2023d] Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. Psychotherapy AI companion with reinforcement learning recommendations and interpretable policy dynamics. In *Proceedings of the Web Conference 2023*, 2023.

[Lin *et al.*, 2023e] Baihan Lin, Stefan Zecevic, Djallel Bouneffouf, and Guillermo Cecchi. Therapyview: Visualizing therapy sessions with temporal topic modeling and ai-generated arts. *arXiv preprint arXiv:2302.10845*, 2023.

[Lin, 2022a] Baihan Lin. Computational inference in cognitive science: Operational, societal and ethical considerations. *arXiv preprint arXiv:2210.13526*, 2022.

[Lin, 2022b] Baihan Lin. Knowledge management system with nlp-assisted annotations: A brief survey and outlook. In *CIKM Workshops*, 2022.

[Lin, 2022c] Baihan Lin. Reinforcement learning and bandits for speech and language processing: Tutorial, review and outlook. *arXiv preprint arXiv:2210.13623*, 2022.

[Lin, 2022d] Baihan Lin. Voice2Alliance: automatic speaker diarization and quality assurance of conversational alignment. In *INTERSPEECH*, 2022.

[Ludewig and Jannach, 2018] Malte Ludewig and Dietmar Jannach. Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, 28(4):331–390, 2018.

[Martin *et al.*, 2000] Daniel J Martin, John P Garske, and M Katherine Davis. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of consulting and clinical psychology*, 68(3):438, 2000.

[Patel *et al.*, 2018] Vikram Patel, Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y Collins, Janice L Cooper, Julian Eaton, et al. The lancet commission on global mental health and sustainable development. *The lancet*, 392(10157):1553–1598, 2018.

[Pazzani and Billsus, 2007] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[Sarwar *et al.*, 2001] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of international conference on World Wide Web*, pages 285–295, 2001.

[Street, 2023] Alexander Street. counseling and psychotherapy transcripts series. *Alexander Street Publishing*, 2023.

[Su and Khoshgoftaar, 2009] Xiaoyuan Su and Taghi M Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009.

[Tracey and Kokotovic, 1989] Terence J Tracey and Anna M Kokotovic. Factor structure of the working alliance inventory. *Psychological Assessment: A journal of consulting and clinical psychology*, 1(3):207, 1989.

[Wampold, 2015] Bruce E Wampold. How important are the common factors in psychotherapy? an update. *World Psych*, 14:270–277, 2015.

[Wang *et al.*, 2014] Xinxi Wang, Yi Wang, David Hsu, and Ye Wang. Exploration in interactive personalized music recommendation: a reinforcement learning approach. *ACM TOMM*, 11(1):1–22, 2014.

[Wang *et al.*, 2020] Rui Wang, Xuemeng Hu, Deyu Zhou, Yulan He, Yuxuan Xiong, Chenchen Ye, and Haiyang Xu. Neural topic modeling with bidirectional adversarial training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 340–350, 2020.

[Watkins Jr, 2013] C Edward Watkins Jr. Being and becoming a psychotherapy supervisor: The crucial triad of learning difficulties. *American Journal of Psychotherapy*, 67(2):134–150, 2013.

[Wu *et al.*, 2019] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. In *Proceedings of AAAI*, volume 33, pages 346–353, 2019.

[Zheng *et al.*, 2018] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. Drn: A deep reinforcement learning framework for news recommendation. In *Proceedings of the 2018 world wide web conference*, pages 167–176, 2018.

[Zou *et al.*, 2020] Lixin Zou, Long Xia, Pan Du, Zhuo Zhang, Ting Bai, Weidong Liu, Jian-Yun Nie, and Dawei Yin. Pseudo dyna-q: A reinforcement learning framework for interactive recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 816–824, 2020.