Research Proposal

Learning To Generate Effective Health Advertisements

Yuval Saadaty
Under the supervision of **Prof. Elad Yom-Tov**

August 2024

Department of Computer Science

Contents

1 Introduction

Public health issues such as vaccine hesitancy, smoking, obesity, and diabetes pose significant challenges to public health authorities. Despite numerous efforts to educate the public about these topics, many individuals remain resistant to positive behavior change. Health advertisements have been shown to be an effective tool in promoting healthier behaviors. For example, Wakefield et al. (2010)[1]found that health campaigns significantly improved health-related knowledge and behaviors. Thus, effective health advertising is crucial for increasing awareness and promoting healthier behaviors, thereby addressing these public health challenges.

The advent of advanced large language models offers new opportunities to enhance the effectiveness of health advertisements. We hypothesize that by leveraging these technologies, we can create more engaging and personalized ads that resonate with specific audience segments. Our goal is to do this by identifying the key features that make health ads more effective and use these insights to generate advertisements that create positive behavior change.

Studies have examined different facets of advertising strategies. Chai et al. (2022)[2] conducted an offline evaluation of nine common strategies used for generating search ads. Murakami et al. (2022)[3] investigated how different advertising appeals influence ad performance, providing valuable insights for crafting more resonant health-related advertisements. However, these studies primarily focused on general mechanisms for ad generation, without tailoring the specific psychological mechanisms to the health behavior that was promoted.

We propose to investigate which aspects of health ads make them more effective for different health-related purposes. We will then use these aspects to generate effective ads using a Large Language Model specifically tasked with creating ads that contain these aspects. This approach will create personalized health ads tailored to the promoted health behavior and to specific audience segments. By conducting comprehensive offline and online experiments, we aim to identify validate our approach.

2 Related Work

In recent years, significant research has been conducted on strategies for effective online advertising, with a particular focus on optimizing advertisement text generation (ATG). These studies explore various methods and techniques aimed at creating texts that can improve CTR.

2.1 Strategies for Effective Online Advertising

In the domain of online advertising, understanding the various factors that contribute to the effectiveness of advertisements is crucial. This section explores four main strategies derived from the literature: Emotional and Psychological Engagement, Persuasive Techniques and Principles, Call to Action (CTA) and Involvement, and Writing Style and Content Strategy.

2.1.1 Emotional and Psychological Engagement

Emotions play a critical role in enhancing the impact of advertisements. Kensinger (2004)[4] explored the dimensions of arousal and valence, showing that emotional engagement can make ads more memorable and impactful. Arousal heightens attention, while valence influences the overall emotional response, driving consumer behavior. Wang et al. (2013)[5] differentiated between thought-based and feeling-based effects in ads. Thought-based effects appeal to rational considerations such as price and product features, while feeling-based effects target emotional responses such as brand loyalty and trustworthiness. These dual aspects address both the cognitive and the emotional processes of decision-making. Additionally, Horne (2024)[6] defines generic ads as sentiment-driven with emotional appeal and specific ad content as customized, amenity-based content with little emotional appeal.Compared the effectiveness of generic emotional appeals to those of specific information in ads. His findings suggest that while generic ads with emotional appeal have a lower cost per click, ads tailored to particular amenities result in higher click-through rates, highlighting the balance between emotional engagement and content.

2.1.2 Persuasive Techniques

Effective advertisements often utilize established persuasive principles. Yang et al. (2019)[7]identified heuristic principles such as trust in a product, citing expert opinions, creating limited offers to enhance exclusivity, and offering freebies, which can enhance ad effectiveness by tapping into human psychology. Yuan et al. (2021)[8] expanded on these principles, integrating Social Identity, Social Proof, Credibility, Commitment and Consistency, Pricing, Product Description, and Motive into a context-aware persuasion model. By employing a multi-task learning model with context-aware attention mechanisms, they optimized ad content to specific consumer contexts, thereby increasing persuasive impact.

2.1.3 CTA and Involvement

Call-To-Action (CTA) is essential for converting passive readers into active participants. Rettie et al. (2005)[9] emphasized the importance of clear, concise CTAs that motivate immediate action, significantly improving engagement and conversion rates. This aligns with the concept of Involvement discussed by Pryzant et al. (2018)[10], where ads that actively engage the audience and prompt interaction lead to higher effectiveness.

2.1.4 Writing Style and Content Strategy

The effectiveness of an ad is also influenced by its writing style and content strategy. Murakami et al. (2022)[3] conducted an aspect-based analysis of advertising appeals (A3) and identified various effective appeals such as special deals, quality, problem-solving, speed, user-friendliness, limited offers, product lineup, and trends. These appeals target different psychological components, ensuring ads attract attention and lead to higher engagement and conversion rates. Pryzant et al. (2018)[10] introduces neural network architectures that not only predict ad performance but also provide insights into how specific stylistic elements influence performance. They identified semantic classes like Involvement, Authority, and Logos, which align with industry best practices. Authority, similar to the trust and credibility discussed by Yang et al. (2019)[7], enhances ad performance by establishing trust and reliability.

2.2 Advertisement Text Generation

In the field of Advertisement Text Generation (ATG), various approaches have been developed to optimize the creation of effective and engaging ad texts. These approaches can be broadly categorized into Template-Based, Extractive, and Abstractive methods. Additionally, research has expanded beyond generating individual text ads to encompass the creation of entire ad campaigns. Bulut et al. (2008)[11] addresses the automation of comprehensive ad campaign planning and execution. This work utilizes generative AI to streamline the orchestration of multiple ads within a cohesive campaign strategy, demonstrating the potential to enhance both individual ad generation and overall campaign management.

2.2.1 Template-Based Approaches

Early ad text generation systems mainly used template-based methods, where relevant keywords were inserted into predefined templates to create ad texts (Bartz et al., 2008)[12]. While this approach ensured that the ads were grammatically correct, it had limitations in terms of diversity and scalability. The range of possible variations was constrained by the number of templates, which were costly to produce. To overcome these limitations,new methods have been explored, such as reusing existing promotional texts (Fujita et al., 2010[13]) and pulling keywords from landing pages to fill the template slots (Thomaidou et al., 2013[14]).

2.2.2 Extractive Approaches

The extractive approach in ad text generation involves selecting important sentences from the input, such as an LP, and outputting them as ad texts. This method ensures consistency with the intended message of the advertiser because it extracts text directly from the source. For instance, the extractive document summarization method can generate ad text that is consistent with the input LPs, as shown in the state-of-the-art NLG models that address practical issues such as factuality and inference speed (Golobokov et al., 2022)[15].

2.2.3 Abstractive Approaches

The abstractive approach involves generating new and unique ad text that accurately captures the content of a given input. This approach is often formulated as a text-to-text generation task, including document summarization and sentence rewriting inspired by various ad-creation processes. Within the abstractive approach, further subgroups emerge based on the model used to generate the text (e.g., encoder-decoder models, reinforcement learning (RL) models, pre-trained language models), and considerations of ad performance metrics like diversity, faithfulness, fluency, and relevance.

Encoder-Decoder Models demonstrated their utility in NLG tasks such as machine translation and summarization and have been applied to automatic ATG Murakami et al. (2023)[16]. Mishra et al. (2020)[17] used ad text as input to develop generation and ranking approaches to refine the text of the ads. Hughes et al. (2019)[18] utilized the content of LPs to generate better textual advertisements for search engines using deep reinforcement learning. Kamigaito et al. (2021)[19] confirmed that generating ad-text using RL with rewards is effective in actual advertisement creation. In this method, a Seq2Seq model is trained using RL to capture useful features for generating effective ad-texts. To explicitly capture the characteristics of effective ad-text, three rewards—fluency, relevance, and ad quality—are used. These rewards are summed and incorporated into the loss function to enhance the effectiveness of the ad.

Reinforcement Learning (RL) Models have been utilized to enhance the generation of ad texts by incorporating rewards that capture useful features of the ads. Kamigaito et al. (2021)[19] confirmed the effectiveness of RL-based encoder-decoder models in ad creation by using rewards for fluency, relevance, and ad quality. Another model-based RL framework, developed for text ad generation, uses a Masked-Sequence Policy Gradient algorithm to integrate efficiently with pretrained models and explore the action space effectively. The process uses a Markov Decision Process (MDP) framework to model the ad generation task. The state represents the input, the action represents the next token to be generated, the reward function approximates the expected CTR using off-policy data to avoid disrupting user experience with real-time feedback (Wang et al., 2021)[20].

Pre-Trained Language Models have become a mainstream approach in ATG, allowing for more fluent and diverse ad text generation. These models are fine-tuned on specific tasks to enhance their performance in generating relevant and engaging ad texts. For example, Honcharenko et al. (2023)[21] use GPT for optimizing advertising texts by fine-tuning it on a dataset of existing ad texts. Honcharenko et al. (2023)[21] describe an optimization process using reinforcement learning to maximize CTR and conversion rates. The generated ad texts are published online, and performance data is used to refine and improve the model's effectiveness in real-world campaigns. The CREATER approach uses a CTR-driven method to generate ad texts based on high-quality user reviews and online A/B test data, leveraging contrastive learning to produce ad texts that achieve higher CTR (Wei et al., 2022)[22]. The FAST model designed to improve the controllability of text generation models. Controllability in text generation refers to the ability to influence and direct the generated content according to specific attributes or guidelines. The FAST model can be applied to a wide range of text generation tasks where controllability is crucial. These include generating ad texts that align with specific marketing messages or brand guidelines (Chai et al., 2022)[2].

2.3 Offline and Online Metrics

Offline evaluation methods are intrinsic assessments conducted without real-time user interaction. These methods include automatic evaluations, which are divided into reference-based and reference-free approaches. Reference-based methods measure the overlap of word n-grams between the generated text and reference texts, using metrics such as BLEU[23] (Papineni et al., 2002), ROUGE[24] (Lin, 2004), METEOR[25] (Banerjee and Lavie, 2005), and CIDEr[26] (Vedantam et al., 2015). Reference-free methods assess aspects like diversity and perplexity, with metrics such as distinct[27] (Li et al., 2016), Self-BLEU[28] (Zhu et al., 2018), and Pairwise-BLEU[29] (Shen et al., 2019) evaluating the uniqueness of the generated text. Additionally, domain-specific metrics, such as predicted click-through rate (pCTR), keyword coverage, keyword position, and predicted ad auction win rate[30] (Shuai et al., 2020), are used to simulate aspects of online evaluation (Mishra et al., 2020[17]; Kamigaito et al., 2021[19]). Human subjective evaluations provide further insights, focusing on general NLG criteria like fluency and coherence[31] (Howcroft et al., 2020), as well as domain-specific criteria such as attractiveness. These evaluations use scoring methods like five-point Likert scales (Likert, 1932)[32] or comparative rankings such as TrueSkill™ (Herbrich et al., 2007)[33] . Purely online approaches, focusing on live feedback and real-time data for continuous improvement, are demonstrated in the research by Kamigaito et al. (2021)[19] and Hughes et al. (2019)[18], which optimize ad texts based on actual user interactions and CTR. The combination of these methods ensures that generated ad texts are both contextually relevant and highly effective in engaging users.

2.4 Click Through Rate (CTR) Prediction

Predicting Click-Through Rates (pCTR) for online advertisements is a critical aspect of optimizing ad performance (see surveys[34,35]). The following sections will explore several of the models described in these surveys.

One notable approach involves the use of Logistic Regression (LR) models to estimate the probability of a click, based on historical data and various features of the ads and user interactions (Richardson et al., 2007)[36]. To overcome the limitations of Logistic Regression models, which struggle to capture high-order feature interactions, Factorization Machines (FM) were introduced by Rendle et el. (2010)[37] as a solution. Incorporating more sophisticated techniques, He et al. (2014)[33] applied gradient boosting decision trees to improve the accuracy of CTR predictions by capturing complex feature interactions. Additionally, with the advent of deep learning, researchers like Zhou et al. (2018)[38] have leveraged deep neural networks to model high-dimensional and sparse ad data, significantly enhancing prediction performance. Wide & Deep Learning combines the strengths of linear models and deep neural networks to handle both memorization and generalization (Cheng et al., 2016)[39]. DeepFM integrates factorization machines for low-order feature interactions with deep neural networks for high-order interactions, making it particularly suitable for CTR tasks (Guo et al., 2017)[40]. Murakami et al. (2022)[3] introduces a BERT-based CTR prediction model that incorporates detected aspect labels from ad texts, enhancing prediction accuracy by considering the nuanced effects of various advertising appeals. These advancements in CTR prediction models not only improve ad targeting but also contribute to more efficient advertising strategies and better user experiences.

3 Research Goals

The primary goal of this research is to enhance the effectiveness of health advertisements to promote healthier behaviors among the public. To achieve this goal, the research will be guided by the following specific objectives:

1. **Identification of Key Features:** Identify the key features of ads that cause ads in each health category (or for each health goal) to maximize the CTR. These features may include psychological mechanisms affecting behavior, targeted messaging, personalization, highlighting key benefits, addressing pain points, using testimonials and experiences, providing educational content, and CTAs.

2. **Development of models to predict key features and CTR:** Utilize an existing dataset of health ads that includes CTR information to create models that predict the key features and the CTR. Specifically, we will tag a subset of the ads with their key features. We will then create models that predict the key features and from them, the CTR for health ads. This will help in creating an explainable model which will enable an understanding of the features most effective in driving higher CTRs for each category.

3. **Online Experimentation and Validation:** Generate ads using a large language model (LLM) based on specific keywords and item descriptions, and conduct online experiments to compare the CTR of LLM-generated ads with existing ads.

4 Proposed Research

This research consists of two phases: Offline evaluation and online experiments. In the offline phase, we will utilize a dataset of health ads to identify and tag each ad with its respective key features. We will then train a model to predict the CTR of these ads from the features. Based on the insights from the offline experiments, we will proceed to the online phase, where we will use an LLM to generate new ads using the identified keywords, product descriptions, and relevant features. The CTR of these generated ads will be compared to the CTR of existing ads to evaluate the effectiveness of the proposed features.

4.1 Offline Approach

4.1.1 Data

We will utilize an existing dataset of health ads that includes CTR information.

4.1.2 Data Labeling

We will manually tag 100-200 ads based on the features identified in the literature review (CTA, arousal valence, thought-based content, feeling-based content, amenity-based content, aspect-based analysis, incorporating principles, and persuasive tactics) to create a labeled dataset of ads.

4.1.3 Features Prediction

We will create text-based predictive models to tag the remaining ads (and any new ads) with their key features. We plan to begin by training a BERT-based classification framework, as used by Murakami[3]. The process will involve encoding ad texts into a BERT encoder to generate a vector representation of the text. For each key feature, this vector will then be passed through a Multi-Layer Perceptron (MLP), which will output the probability of the ad text being associated with that particular feature. The model will learn to recognize features and generate appropriate labels for new ads, including those that will be generated using an LLM at later stages.

We will investigate other models for predicting the key features. For example, in the case of numeric features such as arousal and valance, linear regression, boosted trees, and random forest, will be examined. The features of these models will be the tokenized text of the ads.

4.1.4 CTR Prediction

We will train a model to predict CTR for product or service ads based on their descriptions, keywords, and additional features. The model will be trained on labeled data, where the label is the actual CTR. The output, denoted as pCTR, reflects the likelihood that a user will click on an ad.

The input features to the CTR predictir will be:

- **Product Description**: Text describing the product or service.
- **Keywords**: Specific words that advertisers bid on, ensuring their ads are shown when these terms are searched.
- **Features**: Numerical attributes including CTA, valence, arousal, thought-based, feeling-based, aspect-based, persuasive-based, and amenity-based features.

Model Architecture

We propose to use a model architecture inspired by Zhu et al. (2021)[41], that is, a fine-tuned BERT model to process the concatenation of the ad title, ad description, call to action, and publisher into an embedding layer. These embeddings are then concatenated with additional numerical features to form a comprehensive input vector. This vector is passed through dense layers to capture complex interactions, culminating in an output layer that produces the pCTR as a probability value.

Model Training

The model will be trained on labeled data, with actual CTRs serving as the labels. A weighted binary cross-entropy loss function is employed.

Explainability

Explainability is an essential requirement in our models because it ensures that the decisions made by our predictive models are transparent and interpretable, which is crucial for both validation and practical application. We will obtain explainable models through imputation and randomization of feature values. Features causing significant performance degradation when permuted or imputed will be deemed critical to the model's predictions. This explainability approach allows for a clear understanding of which inputs most influence the predicted CTR, aiding in model validation and refinement.

Ads Theme	Paper Title	Number Of Ads
Anorexia	Human Inducing Behavioral Change in Seekers of Pro-Anorexia Content, Using Internet Advertisements[42]	42
Obesity	The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants[43]	61
Cancer diagnosis	Screening for cancer using a learning Internet advertising system[44]	18
Sleep	Identifying sleep disorders from search engine activity: Combining user-generated data with a clinically-validated questionnaire[45]	12
COVID19	Early detection of COVID-19 in China and the USA: summary of the implementation of a digital decision-support and disease surveillance tool[46]	8
Philadelphia	Trojan horse: An analysis of targeted advertising to reduce sexually transmitted diseases among YMSM[47]	1885
Adverse drug reactions	Unpublished	559
Algorithmic copywriting	Unpublished	142
Influenza vaccination	Unpublished	50

Table 1. *Data Sources*

4.2 Online Approach

4.2.1 Data Collection

For a given product or service, we will leverage the CTR model to identify key attributes that should or should not be featured in an advertisement. Based on these insights, we will then generate new health-related ads using an LLM, prompting it with the relevant keywords, product descriptions, and the necessary key features identified by the CTR predictor.

4.2.2 Evaluation

- We will evaluate whether the LLM-generated ads include the specific features requested using the models from the offline approach.

- We will collect feedback from users through surveys and analyze the feedback for insights on relevance, appeal, and clarity of the ads.

- By conducting an online test, publishing the ads on Google Ads, and evaluating the CTR for each ad as the ratio of the number of clicks to the number of ad impressions, we will compare the CTR of the generated ads with the CTR of existing ads to determine significant differences.

By conducting both offline and online testing, this research aims to uncover the important features in health ads and leverage an LLM to generate effective advertisements. This approach not only enhances our understanding of what makes health ads successful but also provides practical tools to improve public health communication and promote healthier behaviors.

5 Uniqueness of Our Research Approach

Our research uniquely combines predictive modeling and LLM-based ad generation to enhance CTR in health product and service advertisements. By identifying key features that influence CTR through offline analysis, we guide the creation of optimized ads using an LLM. These ads are then tested in an online environment, such as Google Ads, to validate their effectiveness. This dual approach—integrating feature analysis and real-world testing—sets our research apart.

Beyond advertising, our methodology can also be applied to generate personalized messages that drive behavior change[48], making it a powerful tool in other areas where humans can be nudged to better health.

References

[1] Wakefield, M. A., Loken, B. & Hornik, R. C. Use of mass media campaigns to change health behaviour. *The lancet* **376**, 1261–1271 (2010).

[2] Chai, J. *et al.* Fast: Improving controllability for text generation with feedback aware self-training. *arXiv preprint arXiv:2210.03167* (2022).

[3] Murakami, S. *et al.* Aspect-based analysis of advertising appeals for search engine advertising. *arXiv preprint arXiv:2204.11445* (2022).

[4] Kensinger, E. A. Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences* **15**, 241–252 (2004).

[5] Wang, T., Bian, J., Liu, S., Zhang, Y. & Liu, T.-Y. Psychological advertising: exploring user psychology for click prediction in sponsored search. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 563–571 (2013).

[6] Horne, J. Should advertisers use generic ad content or specific ad content when marketing apartments with Google Ads? (2024).

[7] Yang, D., Chen, J., Yang, Z., Jurafsky, D. & Hovy, E. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. 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)*, 3620–3630 (2019).

[8] Yuan, Y. *et al.* Persuade to click: Context-aware persuasion model for online textual advertisement. *IEEE Transactions on Knowledge and Data Engineering* **35**, 1938–1951 (2021).

[9] Rettie, R., Grandcolas, U. & Deakins, B. Text message advertising: Response rates and branding effects. *Journal of targeting, measurement and analysis for marketing* **13**, 304–312 (2005).

[10] Pryzant, R., Basu, S. & Sone, K. Interpretable neural architectures for attributing an ad's performance to its writing style. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 125–135 (2018).

[11] Bulut, A. & Arslan, B. Creating ad campaigns using generative ai. In *Applications of Generative AI*, 23–36 (Springer, 2024).

[12] Bartz, K., Barr, C. & Aijaz, A. Natural language generation for sponsored-search advertisements. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, 1–9 (2008).

[13] Fujita, A. *et al.* Automatic generation of listing ads by reusing promotional texts. In *Proceedings of the 12th international conference on electronic commerce: Roadmap for the future of electronic business*, 179–188 (2010).

[14] Thomaidou, S., Lourentzou, I., Katsivelis-Perakis, P. & Vazirgiannis, M. Automated snippet generation for online advertising. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 1841–1844 (2013).

[15] Golobokov, K. *et al.* Deepgen: Diverse search ad generation and real-time customization. *arXiv preprint arXiv:2208.03438* (2022).

[16] Mita, M., Murakami, S., Kato, A. & Zhang, P. Camera: A multimodal dataset and benchmark for ad text generation. *arXiv preprint arXiv:2309.12030* (2023).

[17] Mishra, S., Verma, M., Zhou, Y., Thadani, K. & Wang, W. Learning to create better ads: Generation and ranking approaches for ad creative refinement. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 2653–2660 (2020).

[18] Hughes, J. W., Chang, K.-h. & Zhang, R. Generating better search engine text advertisements with deep reinforcement learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2269–2277 (2019).

[19] Kamigaito, H., Zhang, P., Takamura, H. & Okumura, M. An empirical study of generating texts for search engine advertising. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, 255–262 (2021).

[20] Wang, X. *et al.* Reinforcing pretrained models for generating attractive text advertisements. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3697–3707 (2021).

[21] Lipianina-Honcharenko, K., Kit, I., Zahorodnia, D. & Osolinskyi, O. Optimization method of advertising texts based on generative models. In *ITTAP*, 126–135 (2023).

[22] Wei, P., Yang, X., Liu, S., Wang, L. & Zheng, B. Creater: Ctr-driven advertising text generation with controlled pre-training and contrastive fine-tuning. *arXiv preprint arXiv:2205.08943* (2022).

[23] Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318 (2002).

[24] Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*, vol. 8, 74–81 (2004).

[25] Banerjee, S. & Lavie, A. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65–72 (2005).

[26] Vedantam, R., Lawrence Zitnick, C. & Parikh, D. Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4566–4575 (2015).

[27] Li, J., Galley, M., Brockett, C., Gao, J. & Dolan, B. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 110–119 (2016).

[28] Zhu, Y. *et al.* Texygen: A benchmarking platform for text generation models. *arXiv preprint arXiv:1802.01886* (2018).

[29] Shen, T., Lei, T., Barzilay, R. & Jaakkola, T. Towards generating long and coherent text with multi-level latent variable models. 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)*, 2079–2089 (2019).

[30] Murakami, S., Hoshino, S. & Zhang, P. Natural language generation for advertising: A survey. *arXiv preprint arXiv:2306.12719* (2023).

[31] Howcroft, D. M. *et al.* Twenty years of confusion in human evaluation: Nlg needs evaluation sheets and standardised definitions. In *13th International Conference on Natural Language Generation 2020*, 169–182 (Association for Computational Linguistics, 2020).

[32] Likert, R. A technique for the measurement of attitudes. *Archives of psychology* (1932).

[33] He, X. *et al.* Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the eighth international workshop on data mining for online advertising*, 1–9 (2014).

[34] Wang, X. A survey of online advertising click-through rate prediction models. In *2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 1, 516–521 (IEEE, 2020).

[35] Gharibshah, Z. & Zhu, X. User response prediction in online advertising. *aCM Computing Surveys (CSUR)* **54**, 1–43 (2021).

[36] Richardson, M., Dominowska, E. & Ragno, R. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, 521–530 (2007).

[37] Rendle, S. Factorization machines. In *2010 IEEE International conference on data mining*, 995–1000 (IEEE, 2010).

[38] Zhou, G. *et al.* Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1059–1068 (2018).

[39] Cheng, H.-T. *et al.* Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, 7–10 (2016).

[40] Guo, H., Tang, R., Ye, Y., Li, Z. & He, X. Deepfm: a factorization-machine based neural network for ctr prediction. *arXiv preprint arXiv:1703.04247* (2017).

[41] Mishra, S. *et al.* Tsi: an ad text strength indicator using text-to-ctr and semantic-ad-similarity. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 4036–4045 (2021).

[42] Yom-Tov, E. *et al.* Inducing behavioral change in seekers of pro-anorexia content using internet advertisements: Randomized controlled trial. *JMIR mental health* **5**, e8212 (2018).

[43] Yom-Tov, E., Shembekar, J., Barclay, S. & Muennig, P. The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. *NPJ digital medicine* **1**, 1–6 (2018).

[44] Yom-Tov, E. Screening for cancer using a learning internet advertising system. *ACM Transactions on Computing for Healthcare* **1**, 1–13 (2020).

[45] Cohen Zion, M., Gescheit, I., Levy, N. & Yom-Tov, E. Identifying sleep disorders from search engine activity: combining user-generated data with a clinically validated questionnaire. *Journal of Medical Internet Research* **24**, e41288 (2022).

[46] Hswen, Y., Brownstein, J. S., Xu, X. & Yom-Tov, E. Early detection of covid-19 in china and the usa: summary of the implementation of a digital decision-support and disease surveillance tool. *BMJ open* **10**, e041004 (2020).

[47] Tolosa-Kline, A., Yom-Tov, E., Hoffman, C., Walker-Baban, C. & Lewis, F. M. Trojan horse: An analysis of targeted advertising to reduce sexually transmitted diseases among ymsm. *Health Education & Behavior* **48**, 637–650 (2021).

[48] Lauffenburger, J. C. *et al.* The impact of using reinforcement learning to personalize communication on medication adherence: findings from the reinforce trial. *npj Digital Medicine* **7**, 39 (2024).