

Harnessing Creativity with Foundational Models

Manisha Verma, JP Morgan

work done in collaboration with interns and colleagues while at Amazon and Yahoo! Research

Generating Advertisements: A Creative Journey

1. Introduction
2. Text Generation
3. Image Generation
4. Open Problems

Creative Opportunity in Online Advertising

- **Online advertising** is a key to increase **brand awareness**.
- Advertisers routinely run campaigns with different content that aim to capture current market trends, user interests and requirements.
- **Popular brands** that have multiple products often **run several campaigns** designed manually by creative editors.
- **Churn rate or revision rate** of these campaigns is **extremely high** i.e. creatives are updated with new images, text or taglines very frequently to attract users.

Creative Opportunity in Online Advertising

Different creatives for some brands as shown in images on the right. Advertisers use different elements (time of the year, market trends or different discounts) to attract users.

The objective is to design a system that aims to **reduce the time spent in creative design** for brands across several categories.

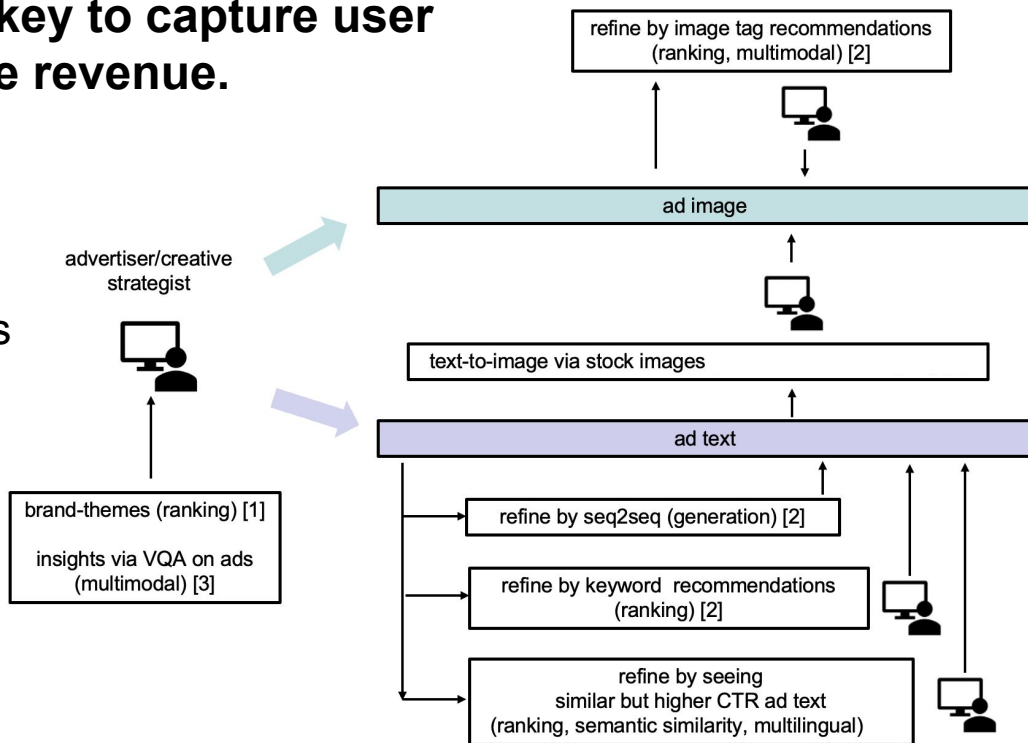


Advertisement Design Workflow

In advertising, content is the key to capture user interest and generate revenue.

1. Engaging advertisements require domain knowledge, market trend awareness, copy writing skills.
2. Ad creative (text + image) design is iterative and tedious.
3. Limited amount of content can be produced manually.

Can we reduce the time spent on designing ads?



Means to support Creative Strategists

- Recommendations [RecSys'19, WWW'20]
 - Given a brand, recommend keywords or advertising concepts to spark innovation.
 - Train models on past advertisements to recommend phrases.
 - Use Campaign Data [CIKM'20, CIKM'21]
 - Use campaigns (good and bad) to show which concepts work with different audience.
 - Train models on past campaign CTR to predict idea quality.

Can we generate ads?

- ***Ad text is different*** from regular text, both in terms of ***sentence structure*** and ***formulation novelty***.
- Native advertisements are ***short sentences***, often encouraging user to take some action. For ex. 'Free gift with new MYBRAND TV!'
- You ***do not want to generate spurious advertisements!*** For ex. Cannot generate '50% discount on new TV' if the advertiser is only giving 10% discount!

Ads at Amazon



Portable Jelly Quiet Book

[Shop Sank >](#)



Sank Portable Jelly
Quiet Book, Toddler...

★★★★★ 9
✓prime



Sank Portable Jelly
Quiet Book, Toddler...

★★★★★ 9
✓prime



Sank Portable Jelly
Quiet Book, Toddler...

★★★★★ 9
✓prime

- We support 3 types of advertisements where we can show one single product or multiple products with a custom image.
- We usually show these along with a headline to attract user's attention.



Sponsored ⓘ

deMoca Quiet Book Montessori
Toys, Toddlers Travel Toy,
Preschool Learning Activities –
Educational Toy, 9 Sensory Toddler...

★★★★★ ~ 3,759

Ad text generation launches at Amazon

Headline suggestions for Store Spotlight now available in the Amazon Ads console

December 27, 2022

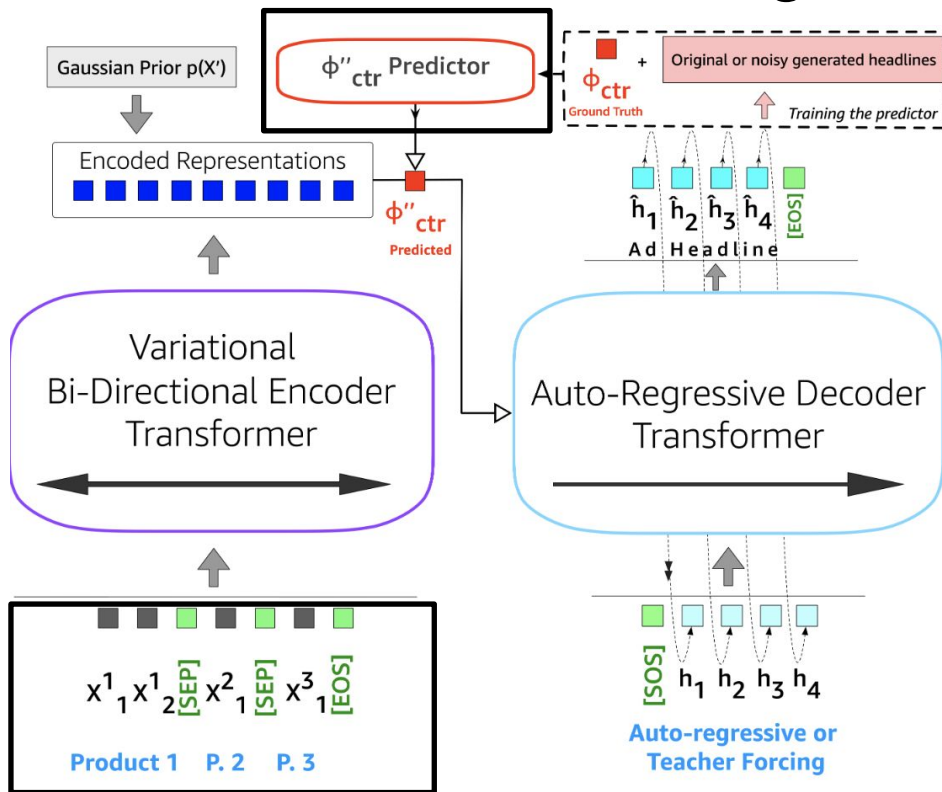
Headline suggestions for Sponsored Display now available in the Amazon Ads console

Updated on February 21, 2023

Generating ad-text with LLMs

- Finetune LLM with some clever tricks:
 - Encoder-Decoder models such as BART or decoder only models as backbone.
 - If CTR data available, train a predictor to estimate generative text quality. Train with feedback from CTR predictor.

Generating ad-text with LLMs



Inputs

3 Product titles and corresponding descriptions

CTR predictor

Trained on historical campaigns

Model Architecture

BART – encoder decoder model.

Training method

Teacher forcing

Feedback Mechanism

Reinforce Trick

Generating ad-text with LLMs

- Pretrain and Finetune LLM for generation.

Controlled Pre-Training

1 Construct Pre-Training Data from Reviews: Aspect-Controlled Masking

Source (Review): x

Freshly sliced fruit
platters are affordable.
Watermelon is delicious,
and plating is
excellent.

Masking

Control code (Aspect): c
affordable

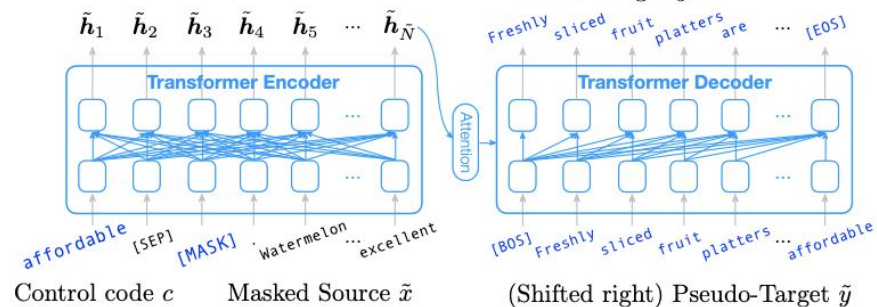
Masked Source: \tilde{x}

[MASK].
Watermelon is delicious,
and plating is
excellent.

Pseudo-Target: \tilde{y}

Freshly sliced fruit
platters are affordable

2 Aspect-Controlled Generation

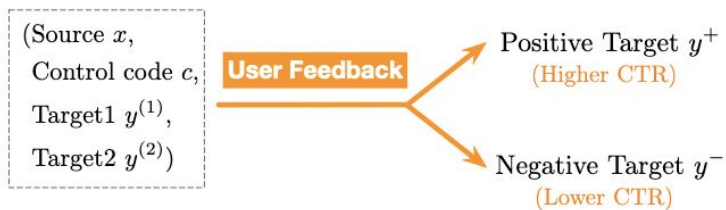


Generating ad-text with LLMs

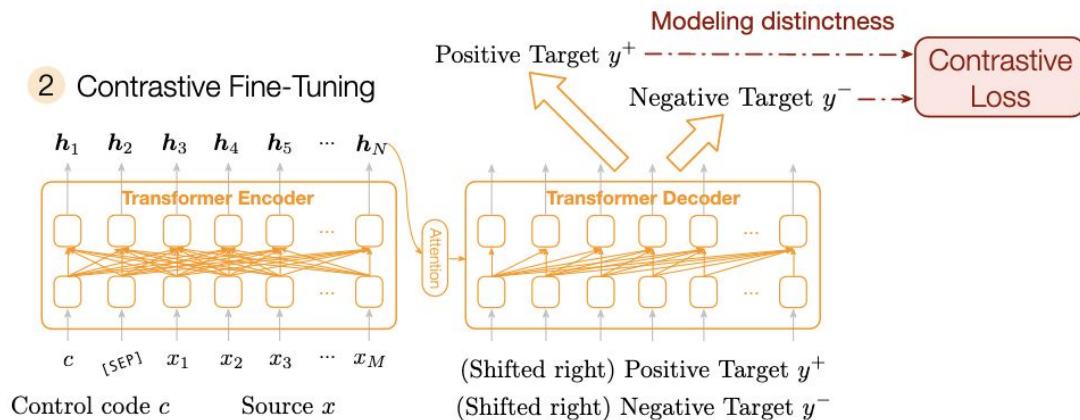
- Pretrain and Finetune LLM for generation.

Contrastive Fine-Tuning

- 1 Construct Fine-Tuning Data via Online A/B Test

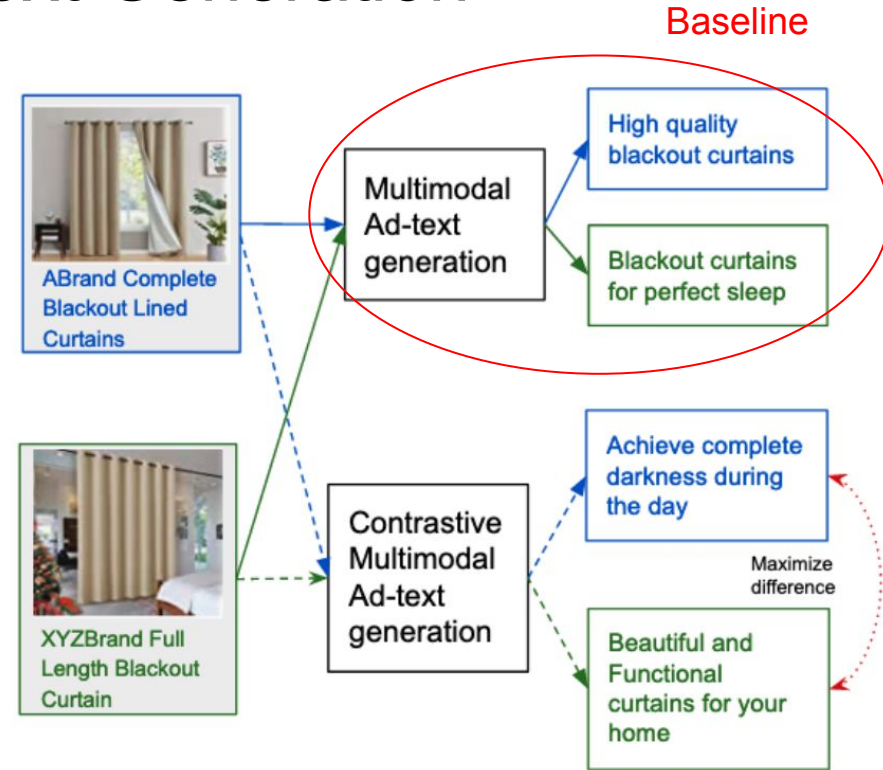


- 2 Contrastive Fine-Tuning

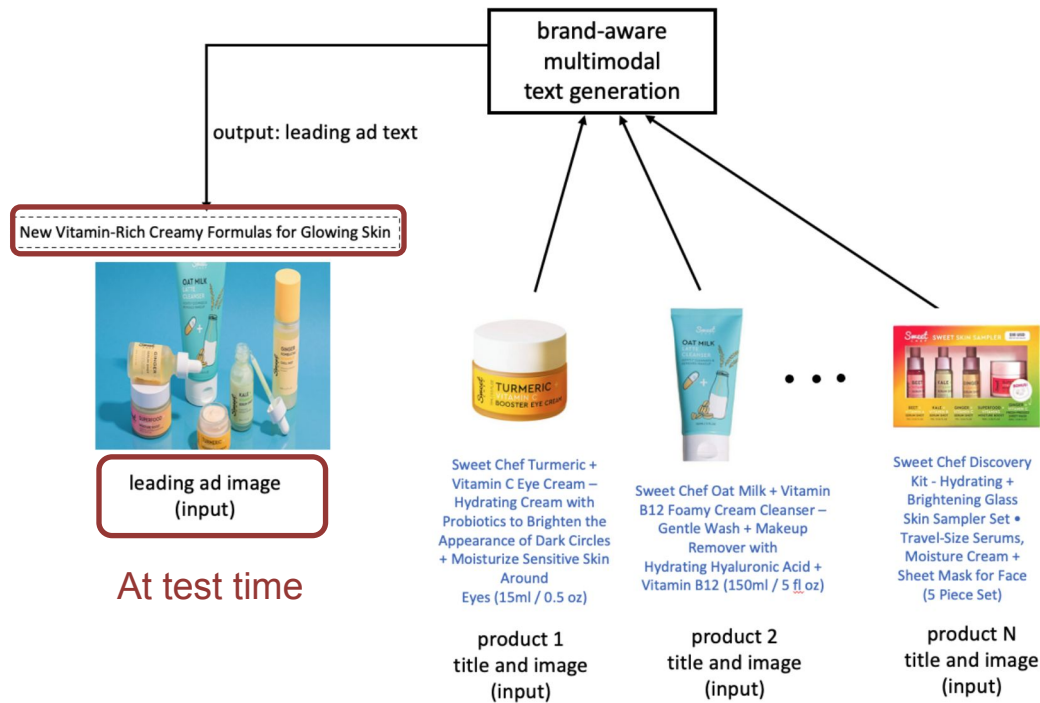


Multimodal Ad Text Generation

- Generate advertisements by exploiting brand level **textual and visual** information across categories.
- Existing methods do not exploit brand level differences.
- Contrastive multimodal model **maximizes** difference between text generated for similar products sold by two different brands.
- It generate **diverse** text ensuring that users have unique experience across brands on their shopping journey.



Multimodal Ad text Generation



Proposed Model

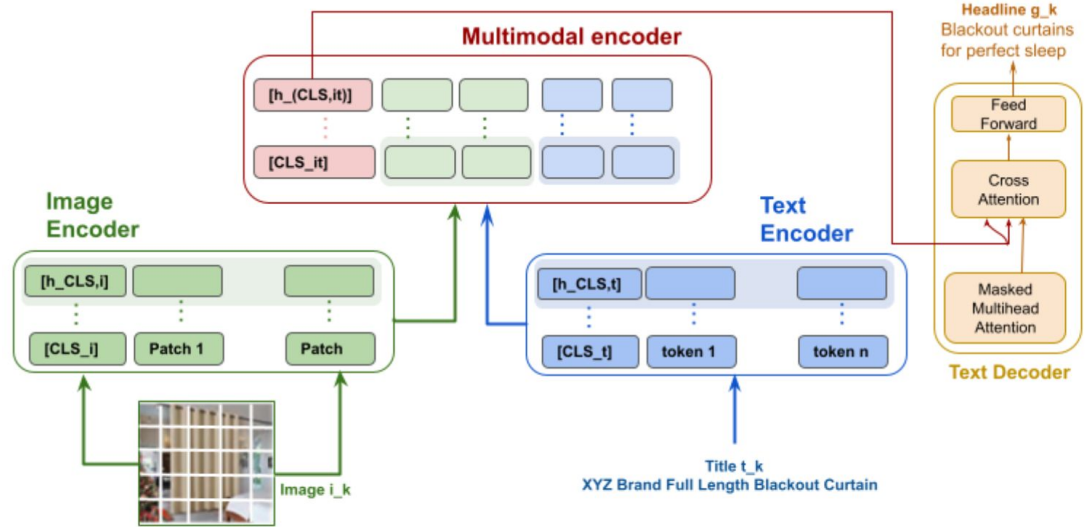
- Minimize cross entropy loss
- Minimize contrastive pairwise margin loss

$$\mathcal{L}_{CE}(\theta) = - \sum_{t=1}^T \log P_{\theta}(y_t | y_{<t}, h_{(it)})$$

$$\mathcal{L}(y^+, y^-) = \max(0, \cos(h_{g^+}, h_{y^-}) - \cos(h_{g^+}, h_{y^+}) + 1),$$

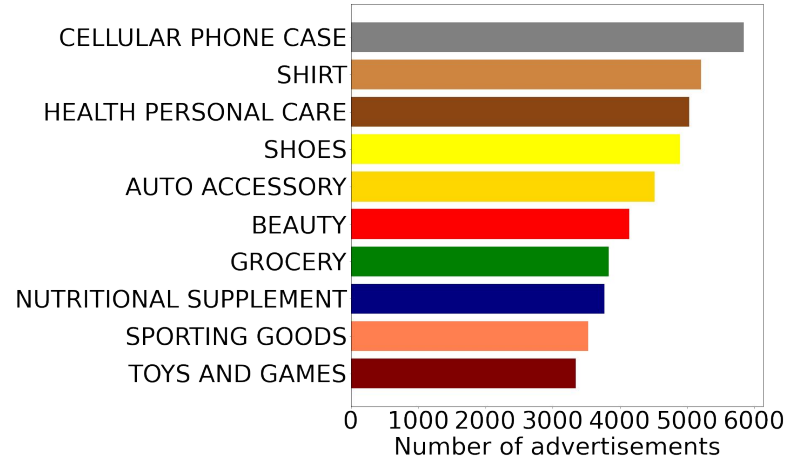
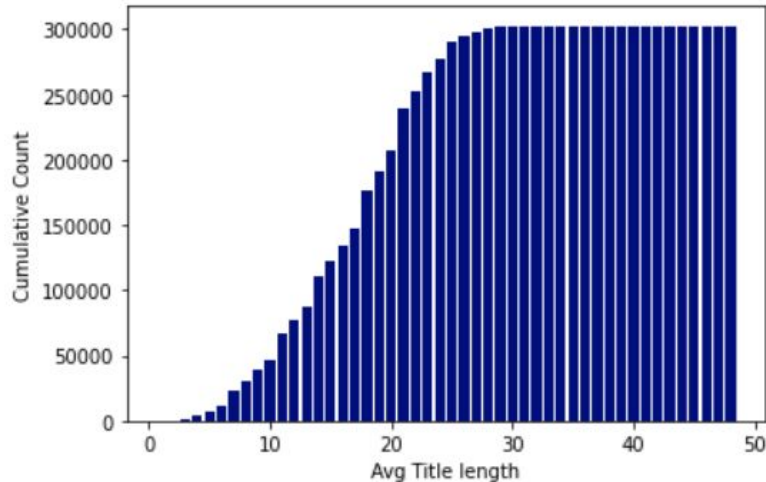
$$\mathcal{L}_{CL} = \sum_{(y^+, y^-) \in P} \mathcal{L}(y^+, y^-)$$

$$\mathcal{L} = \lambda \mathcal{L}_{CE} + (1 - \lambda) \mathcal{L}_{CL}$$



Data

- Data with 335,165 advertisements from 42400 brands. Each advertisement can contain upto three products.
- 20% of titles are shorter than 10 words. Text-only generation models perform poorly on this segment.



Results

- % improvement over GPT2 baseline.
- MATG is trained with one product image.
- MATG-3 is trained with 3 product images.

Text baselines: GPT2, T5, BART, SCMLM, COBART

Image baselines: ClipCap, Mantis

Model	BLEU	R1	R2	RL	BertS	SBLEU
BART[11]	-23.61	-7.98	-9.64	-9.09	-3.41	-4.19
T5[20]	-13.89	-1.84	-5.42	-9.74	-2.27	-3.52
ClipCap[16]	-71.53	-54.29	-69.28	-53.57	-2.27	-164.32
ManTis[23]	-7.64	-4.29	-6.63	-4.87	-0.68	20.70
GPT-2[1]	-	-	-	-	-	-
SCMLM[8]	52.78	21.60	53.61	23.70	1.59	29.52
COBART[7]	80.56	29.14	63.25	32.14	1.14	36.12
MATG ($\lambda = 1$)	76.39	34.66	69.28	37.34	2.73	40.53
MATG	84.72	38.04	75.90	40.91	2.95	44.93
MATG-3 ($\lambda = 1$)	88.89	36.50	78.31	39.61	2.95	40.31
MATG-3	90.28	38.96	86.75	40.26	3.41	41.63

Cold Start evaluation

Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-18.75	-3.37	-1.47	-3.26	0.24	-11.65
COBART	-	-	-	-	-	-
MATG-3	57.81	38.55	89.71	39.13	2.24	21.12

Short Title evaluation (len < 10 words)

Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-8.20	-3.90	-8.16	-6.21	-0.57	21.43
GPT2	-	-	-	-	-	-
MATG	95.90	36.69	80.95	40.00	2.85	42.86
MATG-3	90.98	30.52	75.51	34.14	2.62	37.62

Results

- % improvement over GPT2 baseline.
- MATG is trained with one product image.
- MATG-3 is trained with 3 product images.

Ablation baselines
No Contrastive loss

Text baselines: GPT2, T5, BART, SCMLM, COBART

Image baselines: ClipCap, Mantis

Model	BLEU	R1	R2	RL	BertS	SBLEU
BART[11]	-23.61	-7.98	-9.64	-9.09	-3.41	-4.19
T5[20]	-13.89	-1.84	-5.42	-9.74	-2.27	-3.52
ClipCap[16]	-71.53	-54.29	-69.28	-53.57	-2.27	-164.32
ManTis[23]	-7.64	-4.29	-6.63	-4.87	-0.68	20.70
GPT-2[1]	-	-	-	-	-	-
SCMLM[8]	52.78	21.60	53.61	23.70	1.59	29.52
COBART[7]	80.56	29.14	63.25	32.14	1.14	36.12
MATG ($\lambda = 1$)	76.39	34.66	69.28	37.34	2.73	40.53
MATG	84.72	38.04	75.90	40.91	2.95	44.93
MATG-3 ($\lambda = 1$)	88.89	36.50	78.31	39.61	2.95	40.31
MATG-3	90.28	38.96	86.75	40.26	3.41	41.63

Cold Start evaluation

Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-18.75	-3.37	-1.47	-3.26	0.24	-11.65
COBART	-	-	-	-	-	-
MATG-3	57.81	38.55	89.71	39.13	2.24	21.12

Short Title evaluation (len < 10 words)

Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-8.20	-3.90	-8.16	-6.21	-0.57	21.43
GPT2	-	-	-	-	-	-
MATG	95.90	36.69	80.95	40.00	2.85	42.86
MATG-3	90.98	30.52	75.51	34.14	2.62	37.62

Results

- % improvement over GPT2 baseline.
- MATG is trained with one product image.
- MATG-3 is trained with 3 product images.

Text baselines: GPT2, T5, BART, SCMLM, COBART

Image baselines: ClipCap, Mantis

Model	BLEU	R1	R2	RL	BertS	SBLEU
BART[11]	-23.61	-7.98	-9.64	-9.09	-3.41	-4.19
T5[20]	-13.89	-1.84	-5.42	-9.74	-2.27	-3.52
ClipCap[16]	-71.53	-54.29	-69.28	-53.57	-2.27	-164.32
ManTis[23]	-7.64	-4.29	-6.63	-4.87	-0.68	20.70
GPT-2[1]	-	-	-	-	-	-
SCMLM[8]	52.78	21.60	53.61	23.70	1.59	29.52
COBART[7]	80.56	29.14	63.25	32.14	1.14	36.12
MATG ($\lambda = 1$)	76.39	34.66	69.28	37.34	2.73	40.53
MATG	84.72	38.04	75.90	40.91	2.95	44.93
MATG-3 ($\lambda = 1$)	88.89	36.50	78.31	39.61	2.95	40.31
MATG-3	90.28	38.96	86.75	40.26	3.41	41.63

Cold Start evaluation

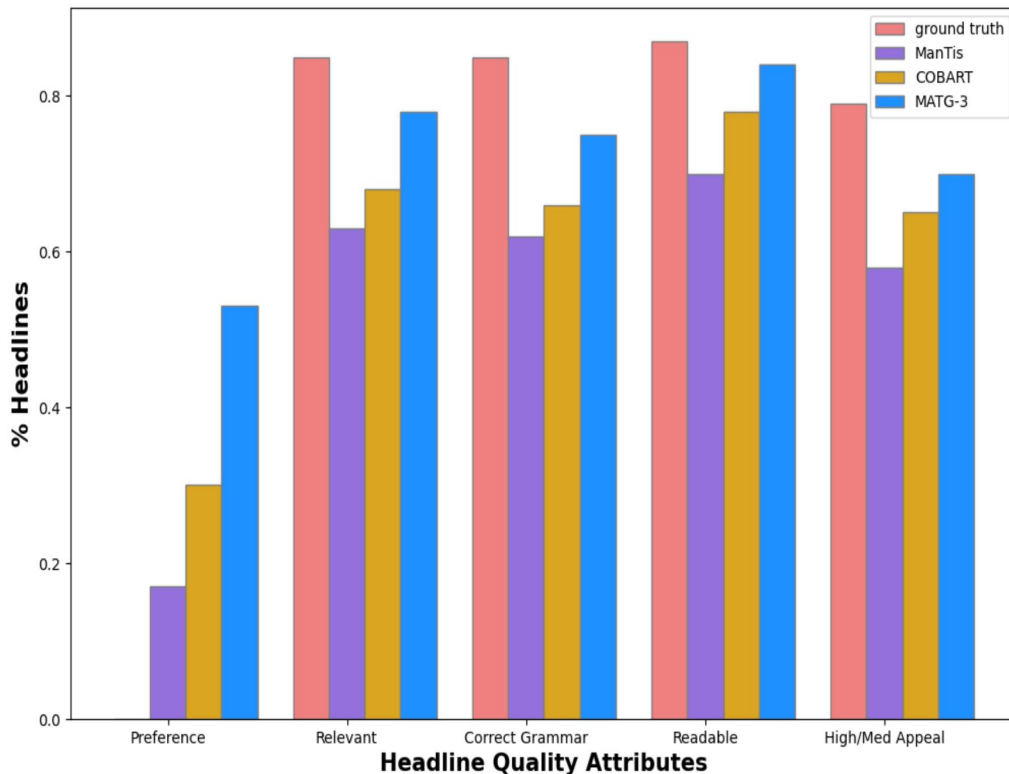
Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-18.75	-3.37	-1.47	-3.26	0.24	-11.65
COBART	-	-	-	-	-	-
MATG-3	57.81	38.55	89.71	39.13	2.24	21.12

Short Title evaluation (len < 10 words)













Model	BLEU	R1	R2	RL	BScore	SBLEU
ManTis	-8.20	-3.90	-8.16	-6.21	-0.57	21.43
GPT2	-	-	-	-	-	-
MATG	95.90	36.69	80.95	40.00	2.85	42.86
MATG-3	90.98	30.52	75.51	34.14	2.62	37.62

Qualitative Analysis


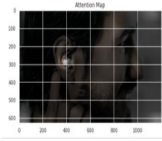

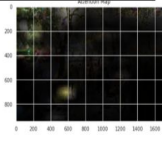

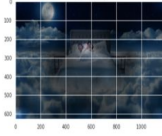

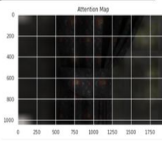
- Token level diversity
- Embedding level diversity
- Manual evaluation on relevance, grammar, readability and advertising appeal.
- Annotators tend to prefer contrastive model MATG-3.
- MATG-3 gets higher score for relevance and advertising appeal.
- Advertiser submitted ad-text does not get perfect scores.



Results

Product 1	Product 2	Product 3	Reference	Generated
 <p>Sam Leather Crossbody Bag</p>	 <p>Rowan Bucket Bag in Teal</p>	 <p>Cassie Convertible Crossbody Bag</p>	Shop Anabaglish Handmade Quality Leather Bags	Sturdy & Stylish Crossbody Phone Bag
 <p>80pcs Waterproof Nature Dinosaur Stickers for Laptop Water Bottle Scrapbook Sticker Pack</p>	 <p>80pcs Waterproof Neon Light Vinyl Stickers for Laptop Water Bottle</p>	 <p>80pcs Waterproof Easter Stickers for Water Bottle Envelopes Cards</p>	Best Stickers for Kids Teens	Cute Waterproof Stickers for Kids and Teens
 <p>PURPLE LEAF Outdoor Dining Chair Coffee</p>	 <p>PURPLE LEAF Outdoor Dining Chair Grey</p>	 <p>PURPLE LEAF Outdoor Dining Chair Dark Blue</p>	Elaborate chairs through complete handiwork	Exquisite dining chairs for every- one
 <p>Quiver Time 80+ Deck Blocks with 2 Dividers - Set of 5 Boxes - White, Black & Green</p>	 <p>Quiver Time Red Portable Game Card Carrying Case</p>	 <p>Black Bolt Quiver Card Case for Carrying Trading Card Games Like Pokemon</p>	Searching for a Better Deck Box?	Looking for a Better Way to Carry Your Cards?

Examples

Product Image	Product Title	Reference Head- line	Generated Head- line	Attention Map
	Shure AONIC 215 True Wireless Sound Isolating Earbuds, Premium Audio Sound with Deep Bass, Bluetooth 5, Secure Fit Over-the-Ear, Long Battery Life with Charging Case	Decades of Experience Supporting Music Legends	The Perfect Earbuds For Your Life	
	Ultimate Confetti Bright Multicolor Biodegradable Tissue Confetti Circles- 1" 30,000 Pieces (1lb) - Confetti Balloons	The Ultimate Confetti Superstore!	Premium Quality Sturdy Party Decorations	
	Wake-Up Single Size Pocket Queen Mattress (72x66x10-inch)	Shop For Single 72x66 Size Mattress	The Most Comfortable Mattress for you	
	Casableu Polyester Blackout Printed Set of 2 Curtains - Silo Orange (Door 7 Feet)	Home that reflects you.	Elegant Blackout Curtains for Bedroom	

Examples of generated headlines with product information and reference headlines

Model Deployment

Generating text is not enough. It needs to be:

1. Policy compliant
2. Grammatically correct
3. Have proper casing
4. Faithful to the product
5. Free of Gender biases

Related Multimodal Work

- Can Pretrained Language Models Generate Persuasive, Faithful, and Informative Ad Text for Product Descriptions? *[Koto et al. ECNLP'22]*
 - Models generate fluent advertisements, but are less faithful and informative, especially in out-of-domain settings
- CAMERA: A Multimodal Dataset and Benchmark for Ad Text Generation. *[Mita, Masato, et al. arxiv 2023]*
 - 12K pairs for training models.
 - Search ads with layout information

Takeaways

- There is ***limited work on using multiple modalities while preserving brand identity*** for creative text generation which differs from more factual tasks such as visual Q&A, caption generation and summarization.
- Our findings suggest that ***product images, along with product title*** can aid in creative text generation.
- ***Contrastive learning especially promotes diversity in headlines for brands that sell very similar products***, where titles would be very similar but images can provide more unique information about the product.
- One can ***explore image-level modeling for different brands*** to improve generation quality.

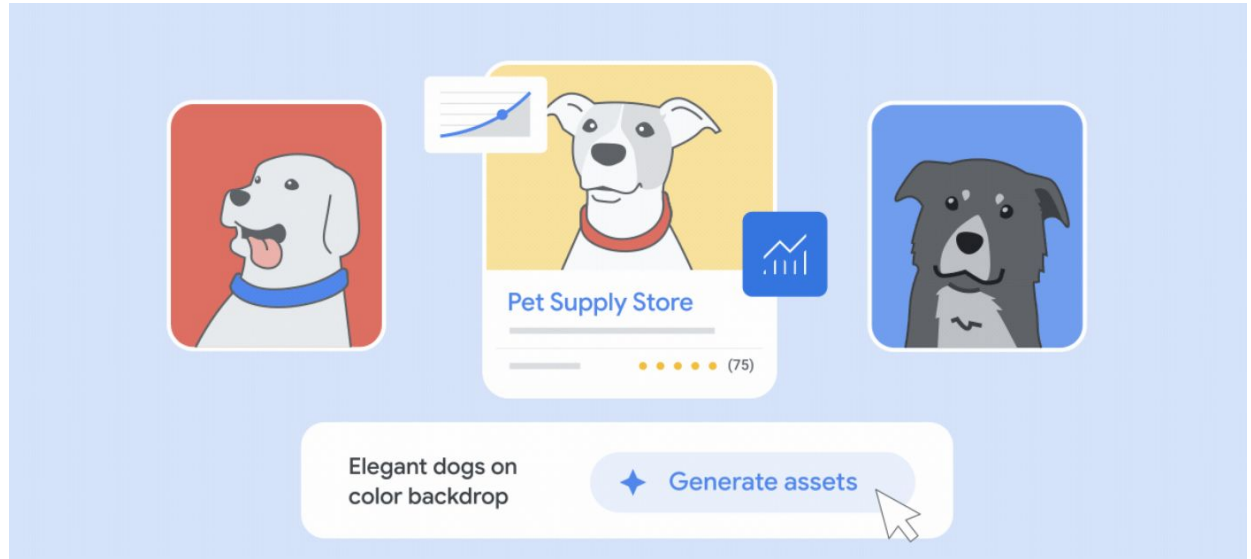
Ad Image Generation

unBoxed 2023: Amazon Ads introduces AI-powered image generation to help brands produce richer creative

Amazon Ads has launched image generation in beta—a generative AI solution designed to remove creative barriers and enable brands to produce lifestyle imagery that can help improve their ads' performance.

Amazon

Google



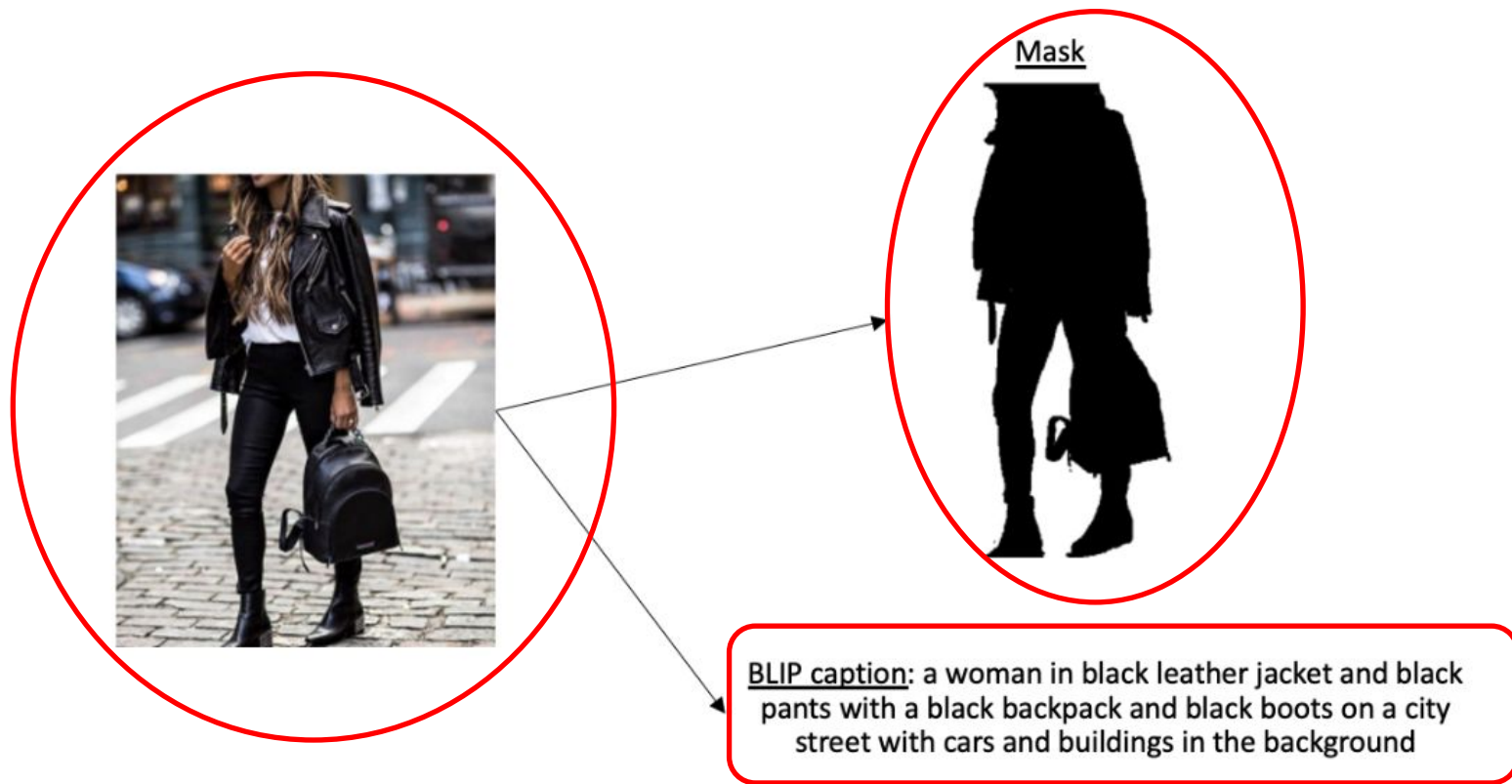
Ad Image Generation

- We can automate not just text but also image generation.
- It takes several iterations to design an image (with or without actors) to sell a product.
- What are the key components to designing such a system?

Ad Image Generation

- What data will we use?
 - Catalogue images with descriptions.
- What model do we use?
 - Diffusion models
- What if we have limited data?
 - ControlNet / Lora

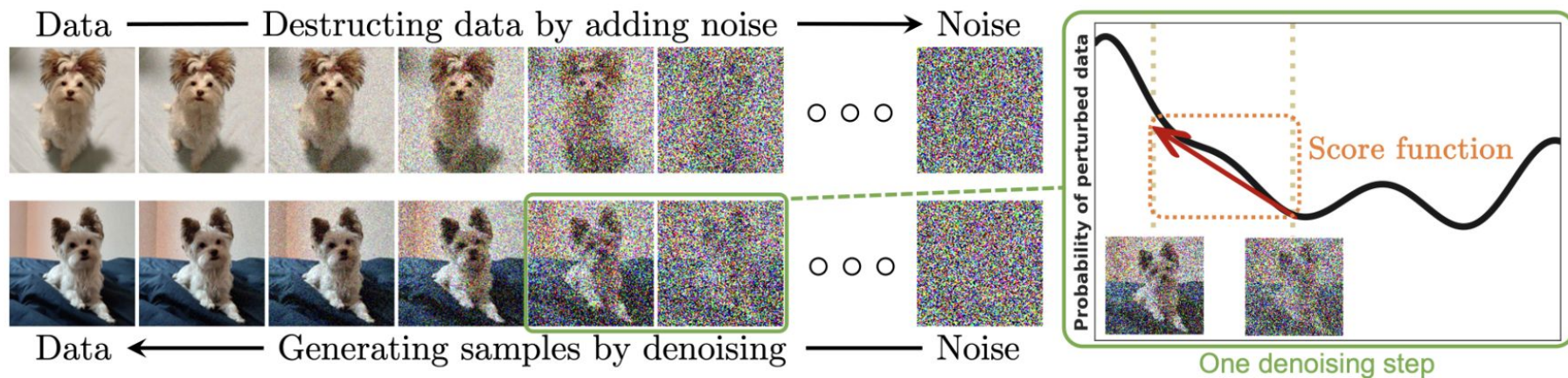
Data for Ad Image Generation



Diffusion Models

- Diffusion models smoothly perturb data by adding noise, then reverse this process to generate new data from noise.
- Each denoising step in the reverse process typically requires estimating the score function.
- The score function is a gradient pointing to the directions of data with higher likelihood and less noise.

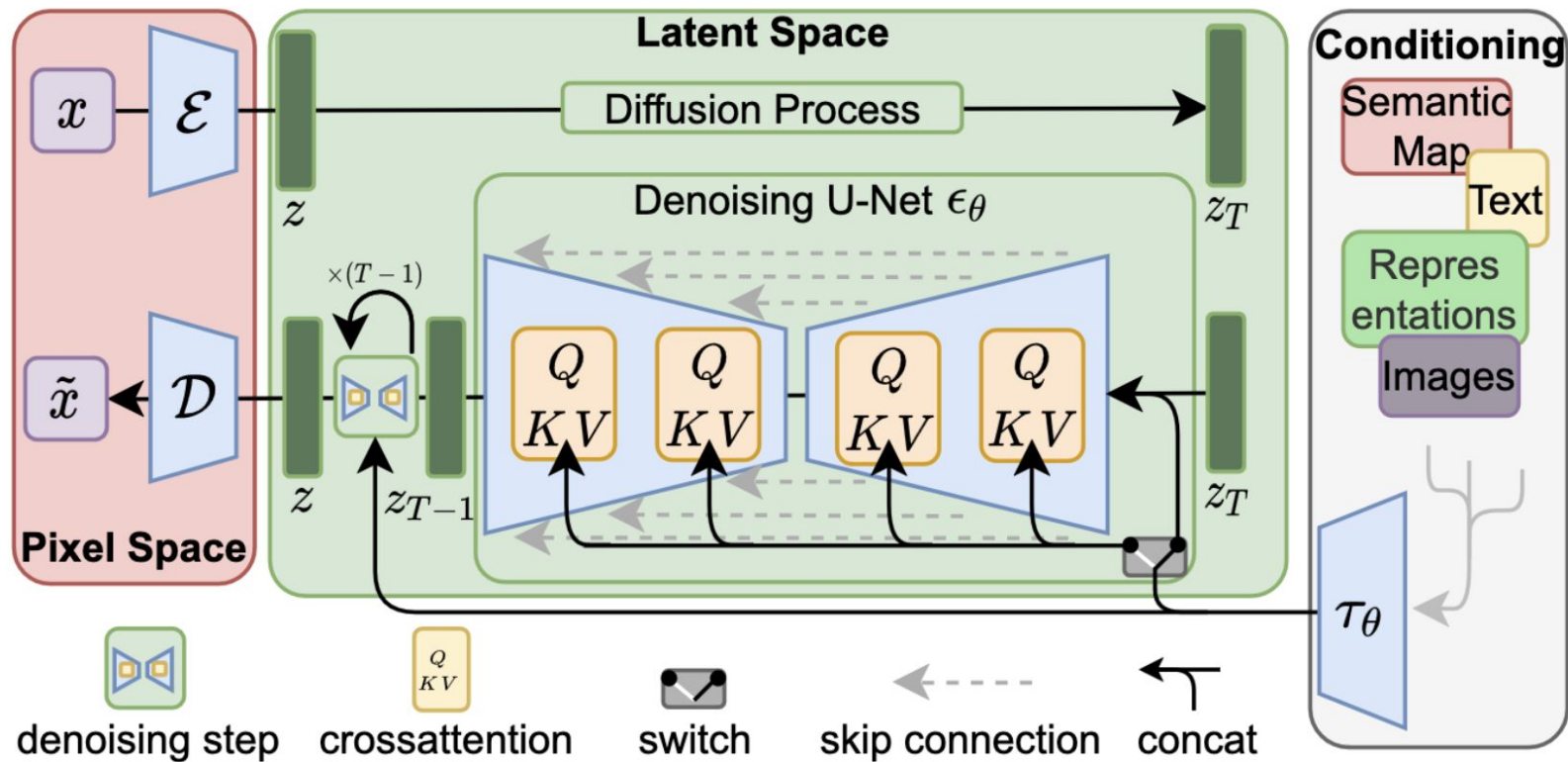
Diffusion Models



Design Choices?

1. Noise generator
2. Denoising model

Ad Image Generation



Ad Image Generation

- How do you encode text prompt:
 - CLIP, T5XXL (different models have different text encoders).
- Positive and Negative prompts are supported by most models.
 - Positive: objects or style that ***should be present*** in image
 - Negative: objects or style that ***must be absent*** from image

Ad Image Generation

- Finetune base diffusion models with
 - Input: <mask, image, text prompt>
 - Output: <target ad image>
- Data volume:
 - 1000s of images: finetune.
 - Few samples: use architectures like control net.

Output Examples

BestNeutralImage
CLIP score: 24.542



BestContextualSetting
CLIP score: 25.454



GCO (z_{QC})
CLIP score: 26.406



GCO (z_P)
CLIP score: 25.807



(a) Summer

Output Examples

BestNeutrallImage
CLIP score: 15.154



BestContextualSetting
CLIP score: 14.319



GCO (z_{QC})
CLIP score: 25.205



GCO (z_P)
CLIP score: 25.298



(b) Hiking

Challenges in Ad Image Generation

- How do we ensure product masks are accurate?
- How do we evaluate such generated images?
- What are the ethical questions associated with such models?
- Finally, such models still struggle with cardinality and special instructions, how do we ensure they are more faithful?

Ethical Evaluation

Image generation models have inherent biases that need to be addressed when running at scale.

- Naik et al '23: First study on evaluating differences in image generation across location, gender and ethnicity.
- Seshadri et al '23: Discussion on how these biases get amplified from training to inference.
- Jha et al '24: More recent work that expands on how different tokens in prompts can lead to different results for location based image generation.

Ethical Evaluation

What if you are not generating humans? What do you evaluate?

- Events based in different locations.
 - Weddings, birthdays etc.
 - There should be healthy representation of country (city) specific details in events.
- Places of work, travel or stay in different locations.
 - Hotels in Nigeria vs Hotels in USA.
 - While there is bound to be a difference, not all hotels in Nigeria are 'huts'. Visual sampling allowed us to determine the level of bias and design post-processing filters.
 - Train stations, Offices or colleges.
- Depending on how you design prompts for image generation, we need to evaluate and control for some level of bias from these models.

Open problems

We have only begun to use multi-modal models for content generation. We can do so much more!

- Holistic understanding of user segments with creative elements of ad.
- Understanding and highlighting product differences in advertisements.
- Helping advertisers build entire catalogues. [code generation]
- Helping advertisers create Videos.

Questions?

References

1. Shaunak Mishra, Manisha Verma and Jelena Gligorijevic. *Guiding Creative Design in Online Advertising*. In RecSys '19.
2. Shaunak Mishra, Manisha Verma, Yichao Zhou, Kapil Thadani and Wei Wang. *Learning to Create Better Ads: Generation and Ranking Approaches for Ad Creative Refinement*. In CIKM '20.
3. Yichao Zhou, Shaunak Mishra, Manisha Verma, Narayan Bhamidipati and Wei Wang. *Recommending Themes for Ad Creative Design via Visual-Linguistic Representations*. In WWW '20.
4. Shaunak Mishra, Changwei Hu, Manisha Verma, Kevin Yen, Yifan Hu, and Maxim Sviridenko. *TSl: an Ad Text Strength Indicator using Text-to-CTR and Semantic-Ad-Similarity*. In CIKM '21.
5. Manisha Verma and Shaunak Mishra. 2022. Recommendation Systems for Ad Creation: A View from the Trenches. In Proceedings of the 16th ACM Conference on Recommender Systems (RecSys '22).
6. Sumit Negi, Manisha Verma, Rajdeep H. Banerjee, Pooja A, Lydia Chilton, Mithun Das Gupta, Vinay P. Namboodiri, and Dinesh Garg. 2022. First Workshop on Content Understanding and Generation for E-commerce. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22).
7. Kanungo, Yashal Shakti, Sumit Negi, and Aruna Rajan. "Ad headline generation using self-critical masked language model." *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*. 2021.