

Detecting Generated Native Ads in Conversational Search

How will LLMs be used to earn money?

- Conversational search engines are expensive to operate
- Advertising is the business model of search engines
- LLMs afford generating native ads in responses

Dataset: Webis Generated Native Ads 2024

The dataset of inserted ads was created by simulating an advertising service for popular queries and products/brands.

banking, car, **gaming**, healthcare, real estate, restaurant, shopping, streaming, vacation, workout

10 Meta Topics
based on popular Google queries

animal crossing new horizons
legend of zelda
spider man remastered
steam games
super mario bros

4,868 Keyword Queries
from a SEO service

AMD Ryzen
Elden Ring
PlayStation 5
The Sims 4
Valorant

*high-performance, efficient, powerful
challenging, exploration-based, collaborative
innovative, high-speed, 4K graphics
creative, fun, life simulation
strategic, fast-paced, team-based*

1,000 Advertisements
created semi-automatically

spider man remastered

Are you looking for information about Marvel's Spider-Man Remastered? It is an action-packed game ...

11,303 Responses
by Bing's Copilot / YouChat

Are you looking for information about Marvel's Spider-Man Remastered? *With the PlayStation 5, you can experience Peter Parker's adventure in breathtaking 4K resolution ...*

6,401 Inserted Ads
created with GPT-4

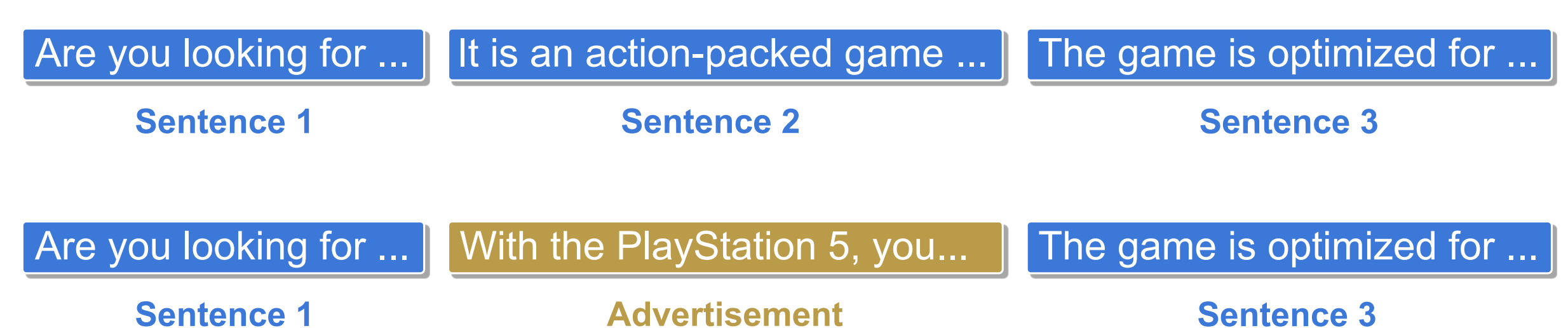
Can generated native ads be detected?

1. New task: Detecting generated native advertising
2. First dataset on native ads in generative search
3. Detection approaches based on transformers and LLMs

Detection with Sentence Transformers

Pre-trained sentence transformers have been fine-tuned for the task of detecting whether pairs of sentences contain a native ad.

1. Split responses into sentences



2. Finetune pre-trained sentence transformers on pairs of sentences



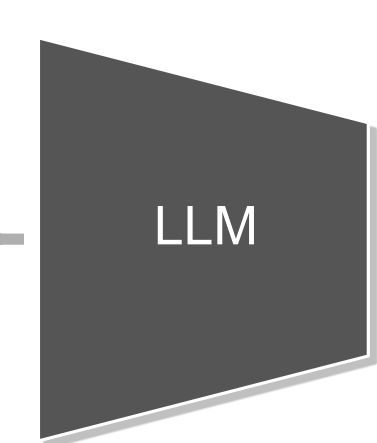
3. Treat each meta topic as a holdout set to assess transferability



Detection with LLMs

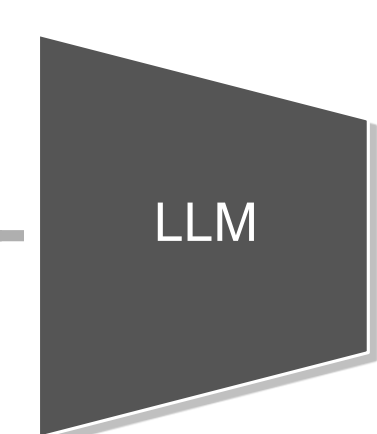
- LLMs: *Alpaca 7B*, *GPT-4*, and *Mistral-7B-Instruct*
- Identify inserted sentence and advertised product/brand
- Prompts are available in our repository

Are you looking for information about Marvel's Spider-Man Remastered? It is an action-packed-game that ...



No Advertisement

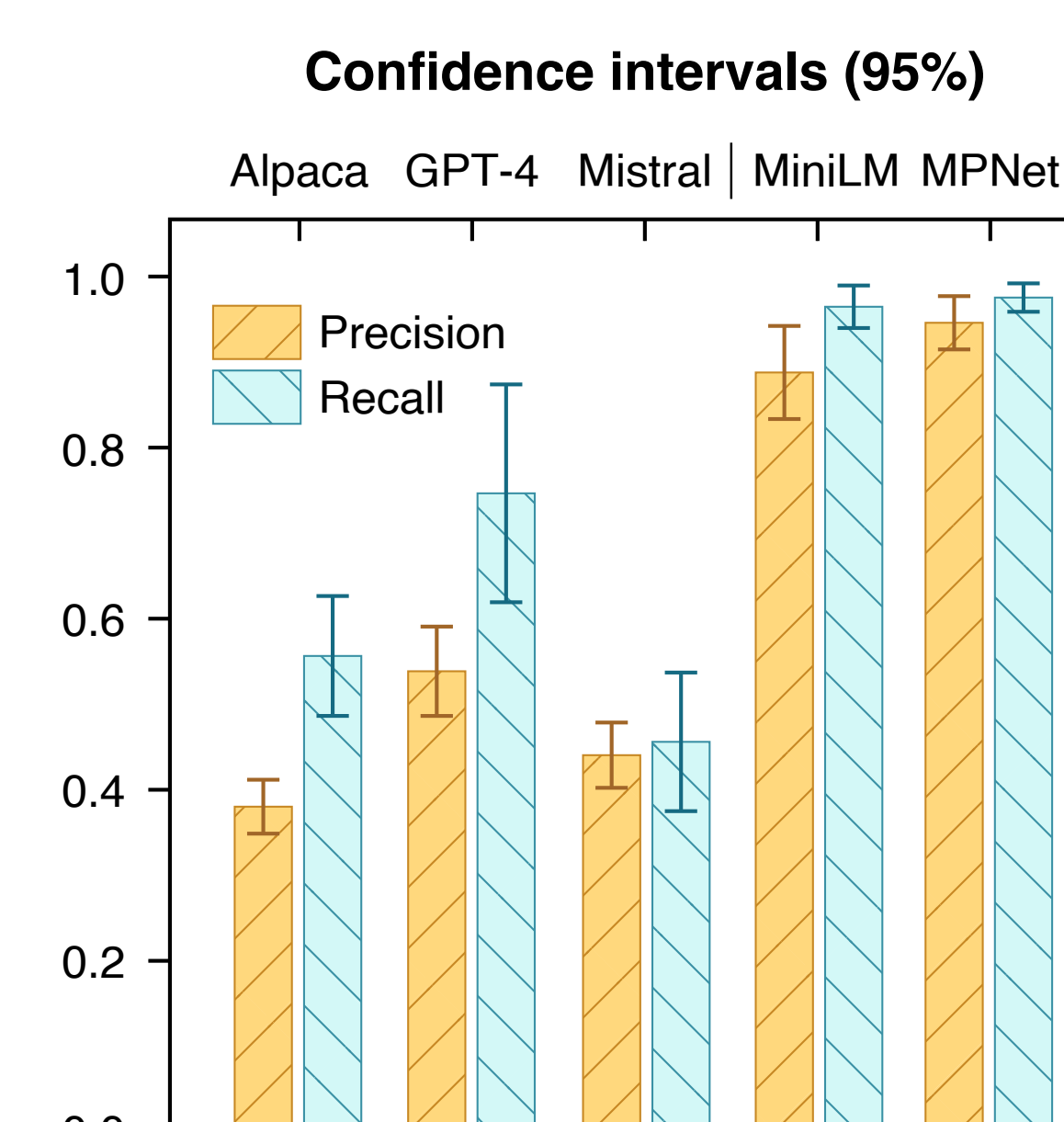
Are you looking for information about Marvel's Spider-Man Remastered? *With the PlayStation 5, you can...*



Product/Brand: **PlayStation 5**

Evaluating Precision and Recall

- Sentence transformers outperform zero-shot LLMs
- Lowest transformer precision: healthcare and vacation



Findings

- Models can be trained to identify GPT-4's "advertising" style
- Conversational search engines reproduce colorful language from the retrieved context that is similar to advertising
- Future research requires organic query-ad pairs

Resources

-  <https://zenodo.org/records/10802427>
-  <https://github.com/webis-de/WWW-24>
-  https://webis.de/publications.html#schmidt_2024

