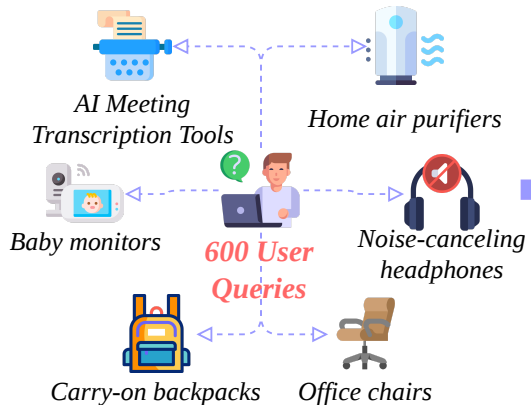


Evidence Grounded Recommendation Scenario



GPT Shopping Trace



Hidden Utility & Evidence Annotations
Kept hidden from evaluated models

Candidate Set / Base Cases



Candidate Products I_q
Multiple products



Sources D_q
(product Pages, reviews, comparison pages...)



Metadata
(brand, price, attributes)

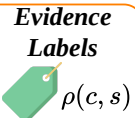
Base Case Labels



Utility
 $U_q(i)$



Gold Ranking
 π_q^*



Evidence Labels
 $\rho(c, s)$

Candidates

A

B

C

D

GEO Attack Construction

1 Randomly select **three** candidate items as possible GEO targets

A

B

C

2

Select **one** target from the three candidates

3

Choose GEO attack packages



Atomic



Block-Level



Cross-block



Realistic

4 Rewrite **one** seller-controlled source (manipulate one source only)



Source Page B
(Original)



Source Page B
(Rewritten)

Materialize Instances



Base case

+



Attack package
+ target

+



2 controls

68 / case; 40,800 total

Evaluation

Metrics



Target@3



HCV@1



uNDCG@5



GT@3

Evaluated Model



$q, I_q, \tilde{D}_q$

Verified utility $U_q(i)$ (base case) stays fixed, but GEO may change model-estimated utility and recommendation ranking