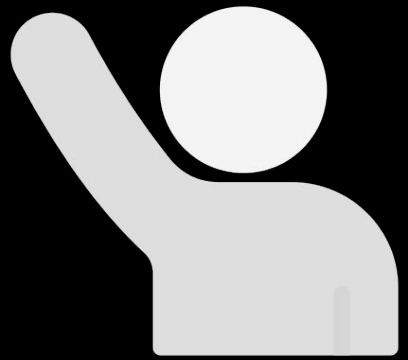
USENIX PEPR ‘24

Through the Lens of LLMs: Unveiling Differential Privacy Challenges

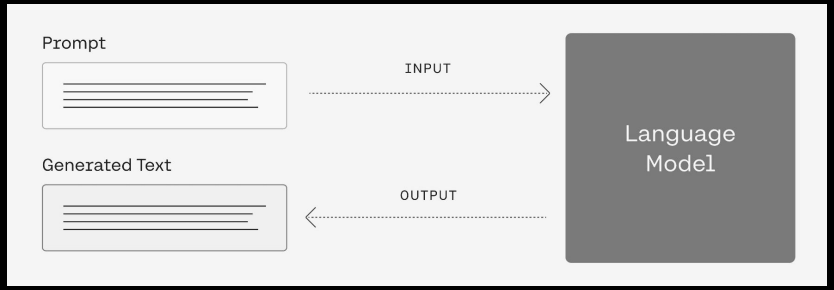
Aman Priyanshu, Yash Maurya, Vy Tran

Are *you* using DP?



Your DP system's vulnerability to LLM-based privacy attacks

Text Inputs? Text Outputs?

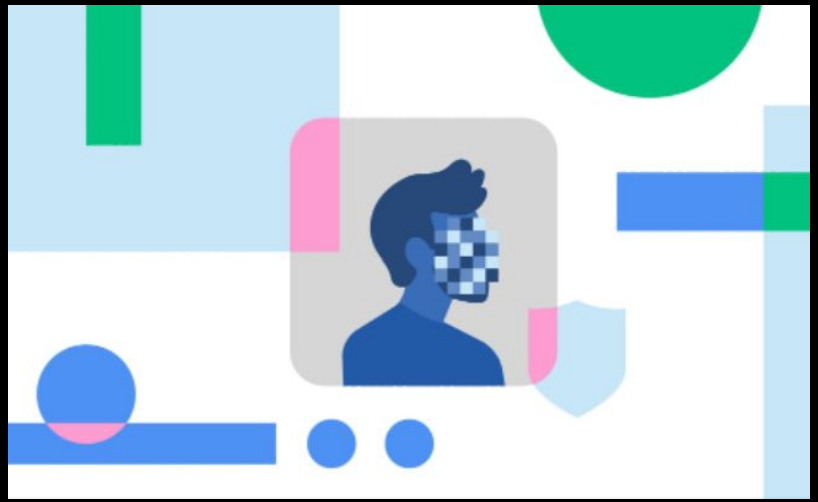


Example:

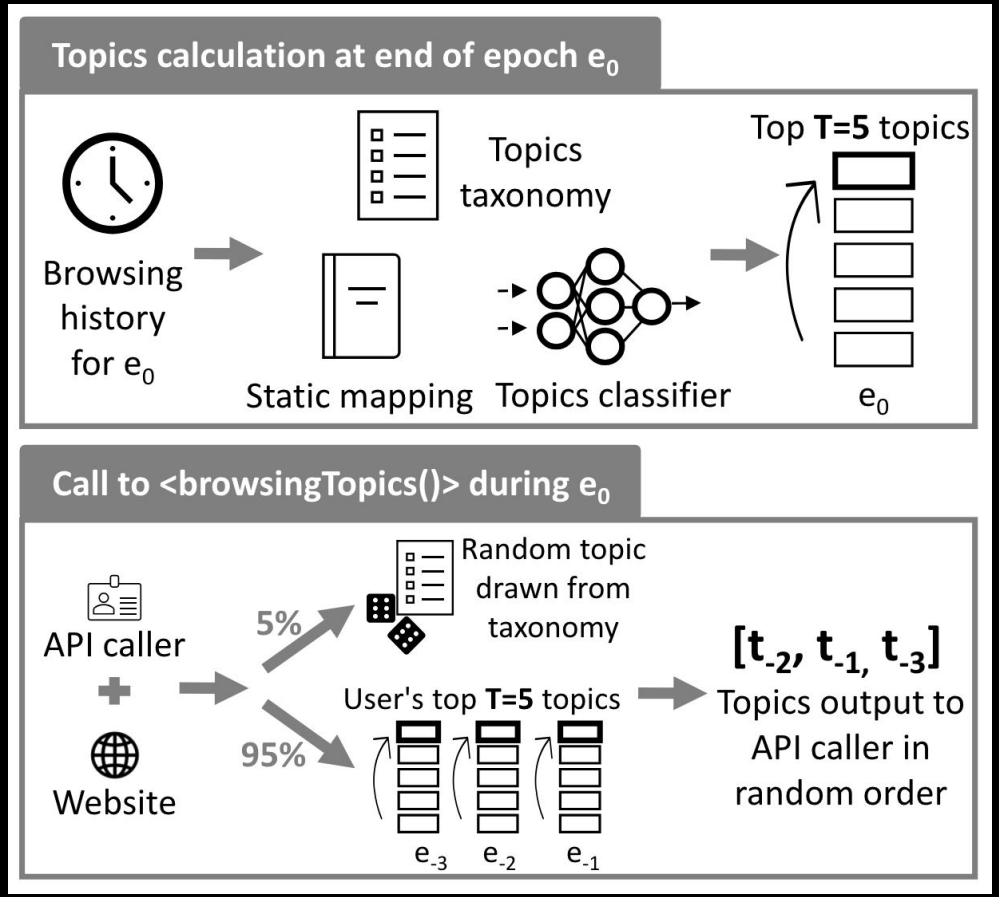
Host
developer.chrome.com
web.dev
google.com

Topics
148. Web browsers 139. Programming
139. Programming
219. Search engines

Case Study: Topics API



Google (2022). Get to know the new Topics API for Privacy Sandbox



Yohan, Beugin, McDaniel, Patrick (2024). A Public and Reproducible Assessment of the Topics API on Real Data

Threat Model: Membership Inference Attack (MIA)

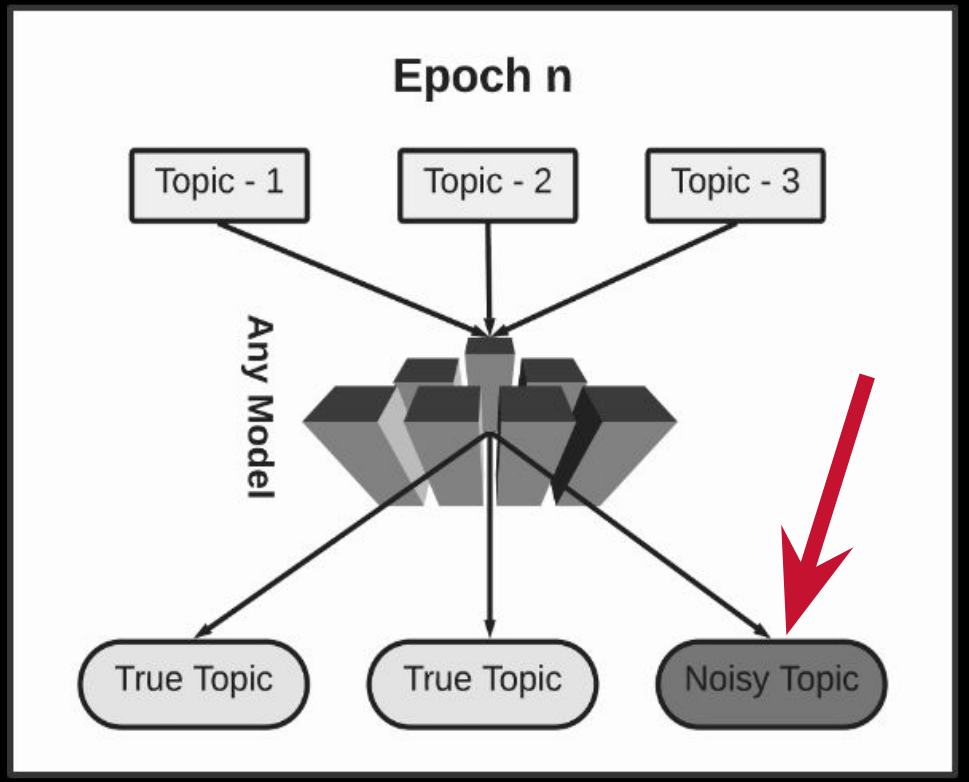
Google (2022). Get to know the new Topics API for Privacy Sandbox

MIA Template

Input: 3 topics from a user

Output: 3 binary values

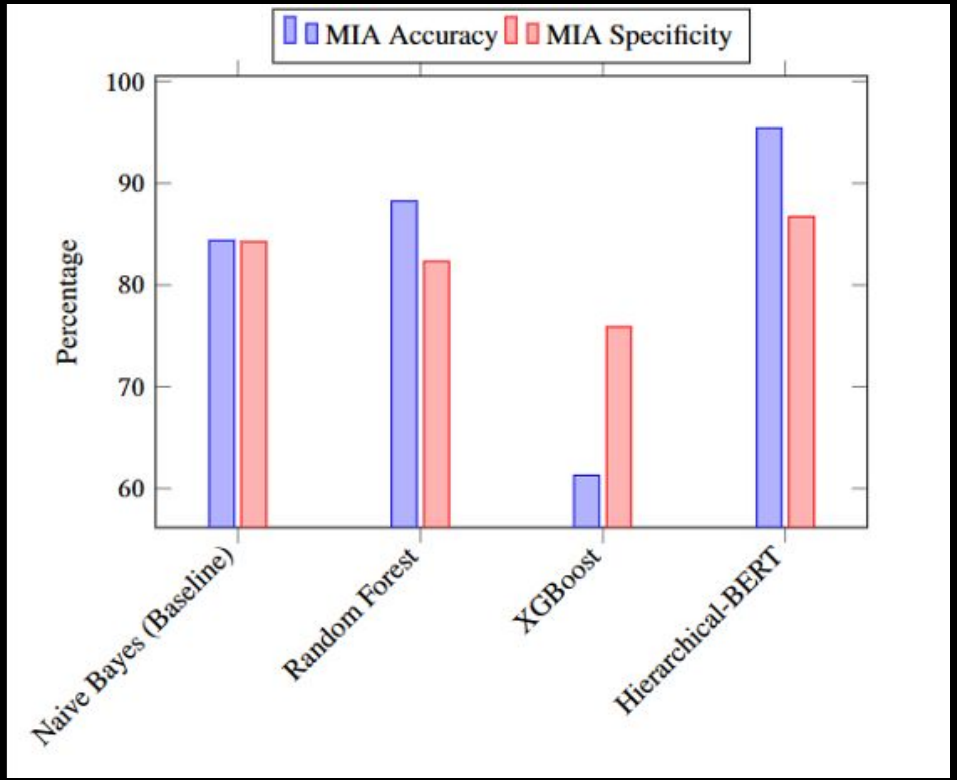
- 0: topic is “normal” i.e. is a real user topic
- 1: topic was randomly selected and does not portray true user behavior



MIA Results

Hierarchical BERT's ability to comprehend nuanced hierarchical relationships within topics

→ 95.41% accuracy



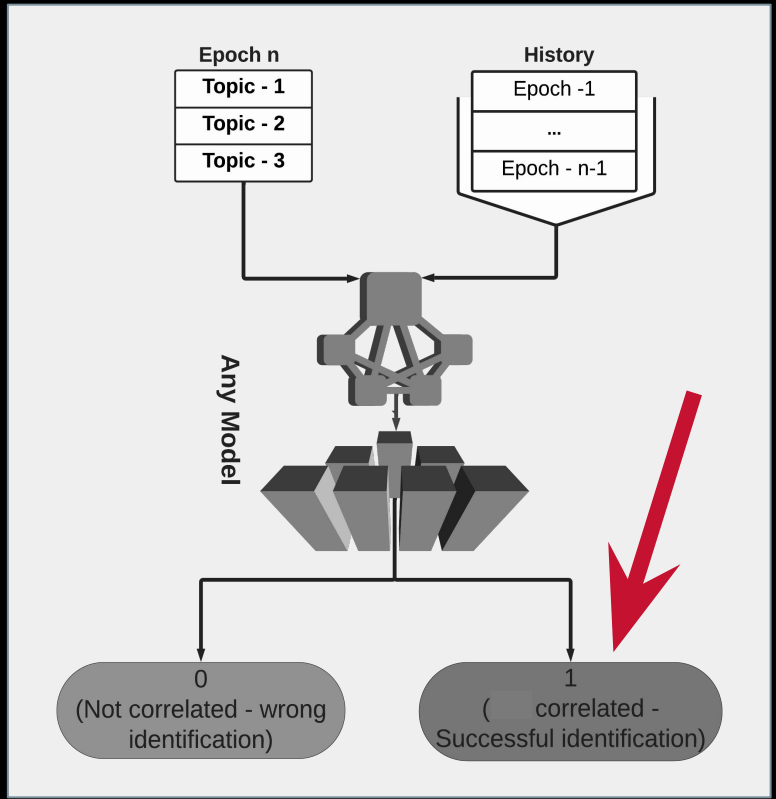
Threat Model: Re-identification Attack (RIA)

RIA Template

Input: 3 topics + history per user

Output: a binary value for each user

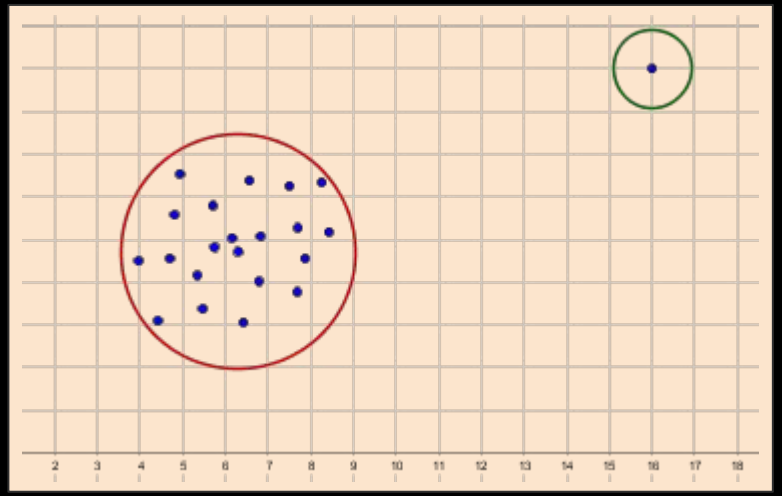
- 0: historical topic set not related to user's given current topics
- 1: strong correlation and thus indicate re-identification



RIA Results (Alongside MIA)

Model	MIA Accuracy	MIA Specificity	RIA Accuracy
Naive Bayes - baseline	84.38%	84.26%	29.33%
Random Forest	88.26%	82.32%	51.80%
XGBoost	61.30%	75.90%	37.42%
Hierarchical-BERT	95.41%	86.73%	68.19%

Addendum: Niche Topics & Edge Cases



Niche Topics

/News	/Pets & Animal
/News/Economy News	/Pets & Animals / Pets
/News/Local News	/Pets & Animals /Pet Food & Pet Care Supplies
/News/Politics	/Pets & Animals / Pets / Birds
/News/Weather	/Pets & Animals / Pets / Cats
/News/World News	/Pets & Animals / Pets / Dogs
	/Pets & Animals / Pets / Fish & Aquaria
	/Pets & Animals / Pets / Reptiles & Amphibians
	/Pets & Animals / Veterinarians

→ Formed their clusters in every embedding model. Due to uniqueness among other topics, higher tendency for word-vector models / sentence encoders to uniquely distinguish these 2 classes of topics

Edge Cases


/Games/Billiards
/Games
/Finance/Investing/Commodities & Futures Trading
/Books & Literature/Children's Literature
/Finance/Investing/Mutual Funds
/Finance/Home Financing
/Reference
/Law & Government/Legal
/Games/Roleplaying Games
/News/Politics
/News/Local News
/Law & Government
/Books & Literature/Poetry
/Law & Government/Crime & Justice
/Games/Card Games
/Finance/Investing
/Finance/Investing/Hedge Funds
/News/Mergers & Acquisitions
/Games/Drawing & Coloring
/Shopping/Flowers
/Finance/Personal Loans
/Real Estate
/Games/Table Tennis
/Finance/Investing/Currencies & Foreign Exchange
/Finance
/News/World News
/Finance/Investing/Stocks & Bonds
/Finance/Credit Cards
/News/Economy News

Remarks on Dataset

A web tracking data set of online browsing behavior of 2,148 users

Kulshrestha, Juhi[1]; Oliveira, Marcos[1]; Karacalik, Orkut[1]; Bonnay, Denis[2]; Wagner, Claudia[1]

Show affiliations

This anonymized data set consists of one month's (October 2018) web tracking data of 2,148 German users. For each user, the data contains the anonymized URL of the webpage the user visited, the domain of the webpage, category of the domain, which provides 41 distinct categories. In total, these 2,148 users made 9,151,243 URL visits, spanning 49,918 unique domains. For each user in our data set, we have self-reported information (collected via a survey) about their gender and age.

We acknowledge the support of Respondi AG, which provided the web tracking and survey data free of charge for research purposes, with special thanks to François Erner and Luc Kalaora at Respondi for their insights and help with data extraction.

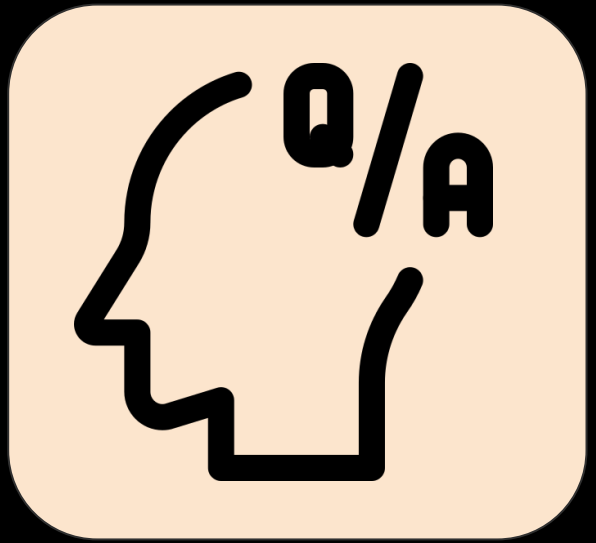
The data set is analyzed in the following paper:

- Kulshrestha, J., Oliveira, M., Karacalik, O., Bonnay, D., Wagner, C. "*Web Routineness and Limits of Predictability: Investigating Demographic and Behavioral Differences Using Web Tracking Data*." Proceedings of the International AAAI Conference on Web and Social Media. 2021. https://arxiv.org/abs/2012.15112.

The code used to analyze the data is also available at https://github.com/gesiscss/web_tracking.

Kulshrestha, J., et al. (2020). A Web Tracking Data Set of Online Browsing Behavior of 2,148 Users

Questions?



Stay in Touch!

Aman Priyanshu
apriyans@andrew.cmu.edu

Yash Maurya
ymaurya@andrew.cmu.edu

Vy Tran
vtran@andrew.cmu.edu

Suriya Ganesh
sayyampe@andrew.cmu.edu

Saranya Vijayakumar
saranyav@andrew.cmu.edu

Dr. Norman Sadeh
ns1i@andrew.cmu.edu

Dr. Hana Habib
htq@andrew.cmu.edu