

A close-up photograph of autumn leaves in shades of orange, yellow, and red, serving as the background for the slide.

CLAP: Mining Android App Descriptions for Permission Requirements Recommendation

Xueqing Liu, Yue Leng, Wei Yang, Chengxiang Zhai, Tao Xie

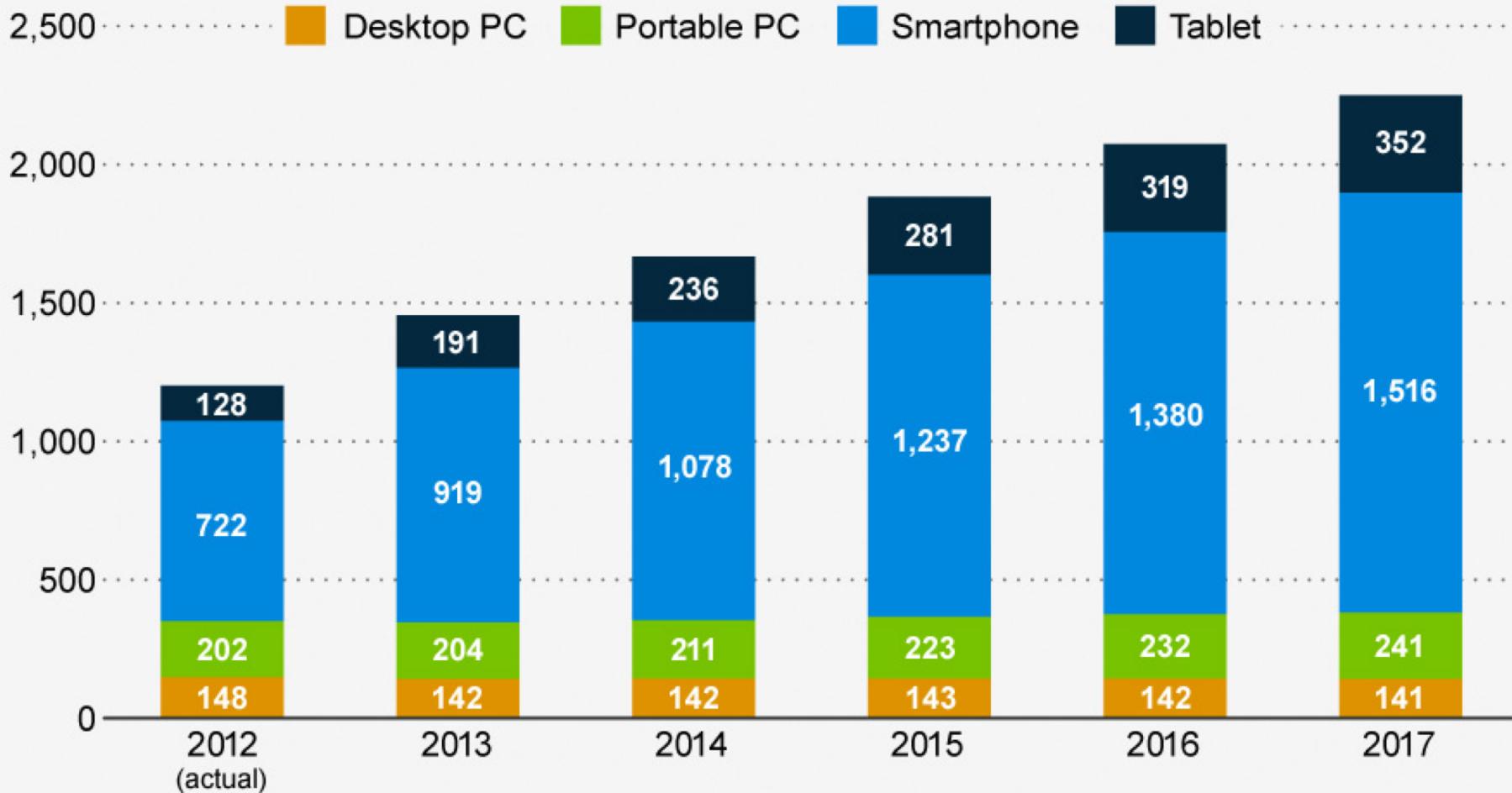
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN



illinois.edu

Growing Smartphone Market

Worldwide Smart Connected Devices Market Forecast Unit Shipments (Millions)



Growing Challenges in Android Security

Facebook Fined £500,000 Over Cambridge Analytica Scandal

California and GDPR “light”: A Match Made in Plaintiffs’ Lawyer Heaven?

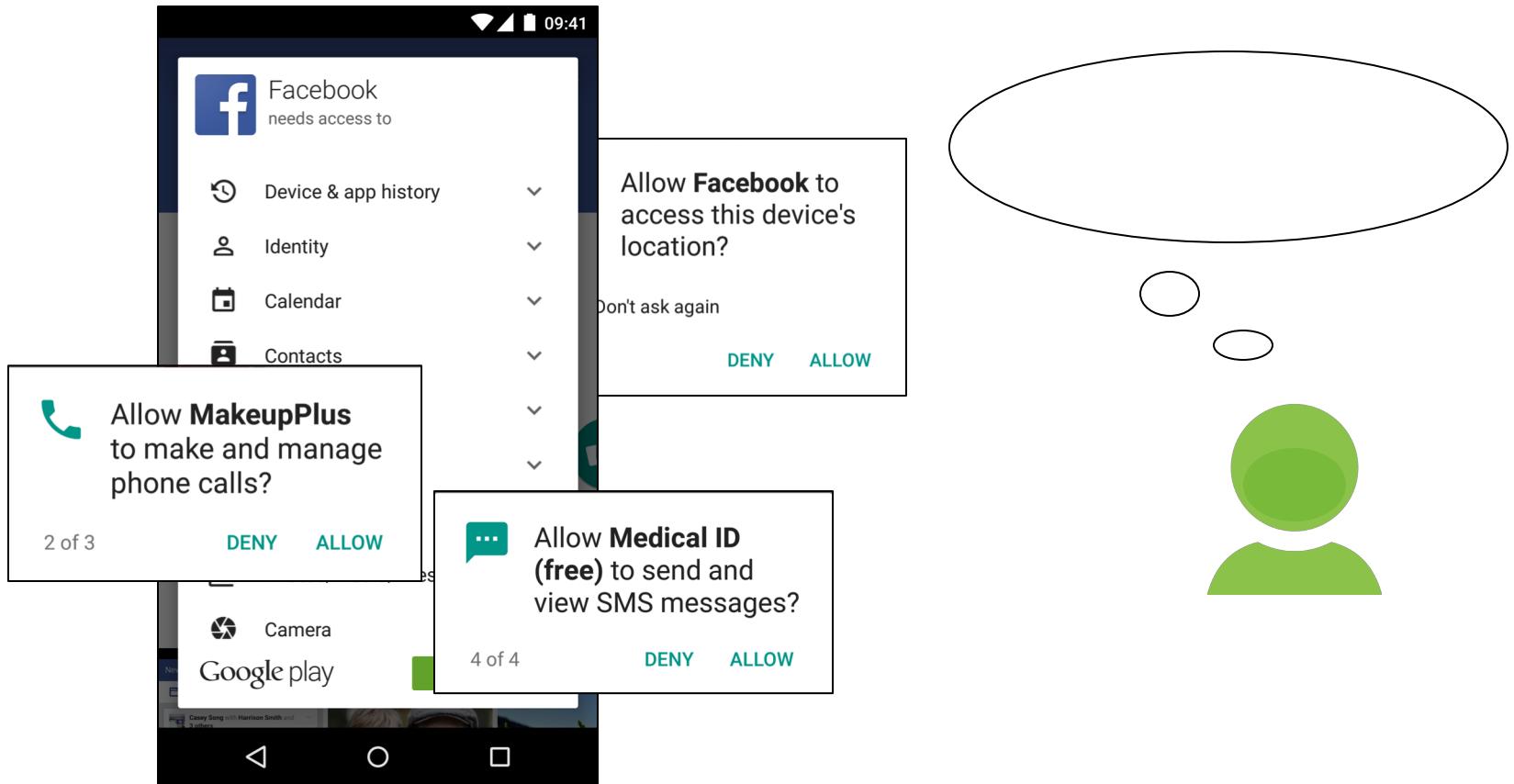
Uber data breach from 2016 affected 57 million riders and drivers

Incident Of The Week: 21M Users Affected By Recent Timehop Breach

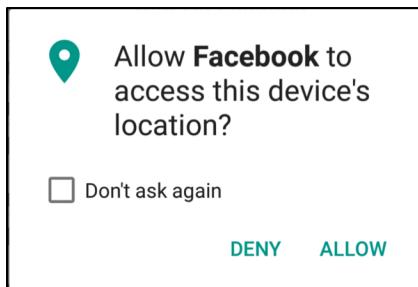
Company Says ‘Memories’ Unaffected By Hack

Tags: Community, Encryption, Hack, Data Breach, Accounts, Wi-Fi, API, Decentralized, Keys, Privacy Concerns, Cloud Environment, Multifactor Authentication, WiFi

Android Permission Requests



Understanding Permission Purposes



Why does Facebook need my location?



- User understanding toward permission [Felt et al. 12]

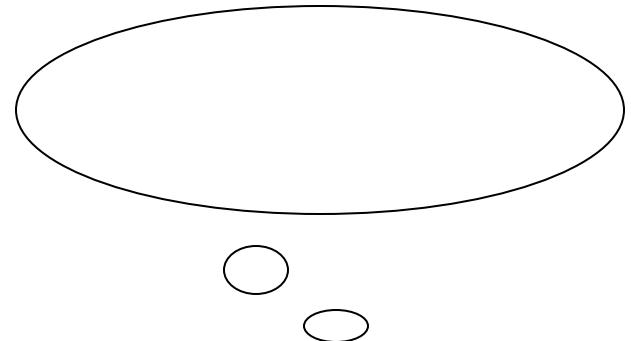


Permission: location

Explaining Permission Purposes

Facebook:

“...Find **local social events**, and make plans...”



- Challenges for apps to explain permission purposes?



Permission: location

3 Challenges in Explaining Permissions: Technical Jargon

GeoTimer: "...GeoTimer harnesses Google's **Geofence API**..."



URL: <https://androidappsapk.co/detail-geotimer-lite/>

Why does Geofence API
need my location?



??

(our anonymous reviewer)

- Challenge 1: Technical Jargons



Permission: location

3 Challenges in Explaining Permissions:

Explaining Rare Permissions

Best Barcode scanner:
 Explains **CAMERA**
Best Barcode scanner
 Skips **LOCATION**
 [Liu et al. 18]

Barcode Scanner:
 Explain **CAMERA** vs. explain **LOCATION?**



URL: **CAMERA 1.**
<https://play.google.com/store/apps/details?id=com.tn.qrbarcodescanner&hl=en>
LOCATION!

CAMERA

- Challenge: Explain **LOCATION** permission



Permission: location



I understand **CAMERA**, but
why is **LOCATION** used?

??



3 Challenges in Explaining Permissions: Deciding the Best Specificity

Orlando theme park: “park location is vital for many useful features in tripwolf , such as locating yourself on the map , routing, finding nearby sights , wifi hot spots , atms , public transport stations and much more...”



URL:

https://play.google.com/store/apps/details?id=com.tripwolf.miniapp_869403_free



- Challenge 1: Technical Jargons

- Challenge 2: Explaining Rare Permissions

Challenge 3: Specificity Requirements



Permission: location

How to Help Apps Write **User-Understandable** Explanations?

- Related work
 - WHYPER: Explaining permission purposes using app description sentences [Pandita et al. '13]
- **CLAP**: Helping Apps Write and Improve their Permission Explanations
- Main idea:
 - Recommend permission explanations in similar apps (**CoL**aborative **A**pp **P**ermission **R**ecommendation)

A Use Case of CLAP

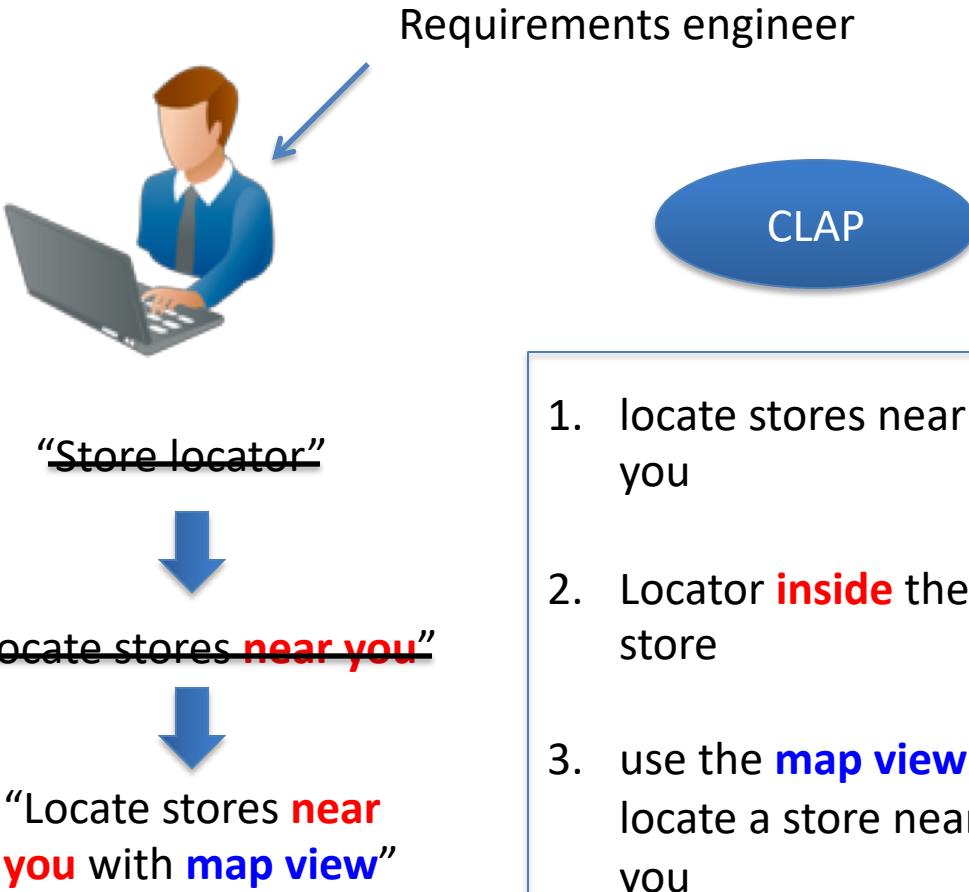


LOCATION
CAMERA
...

AndroidManifest.xml

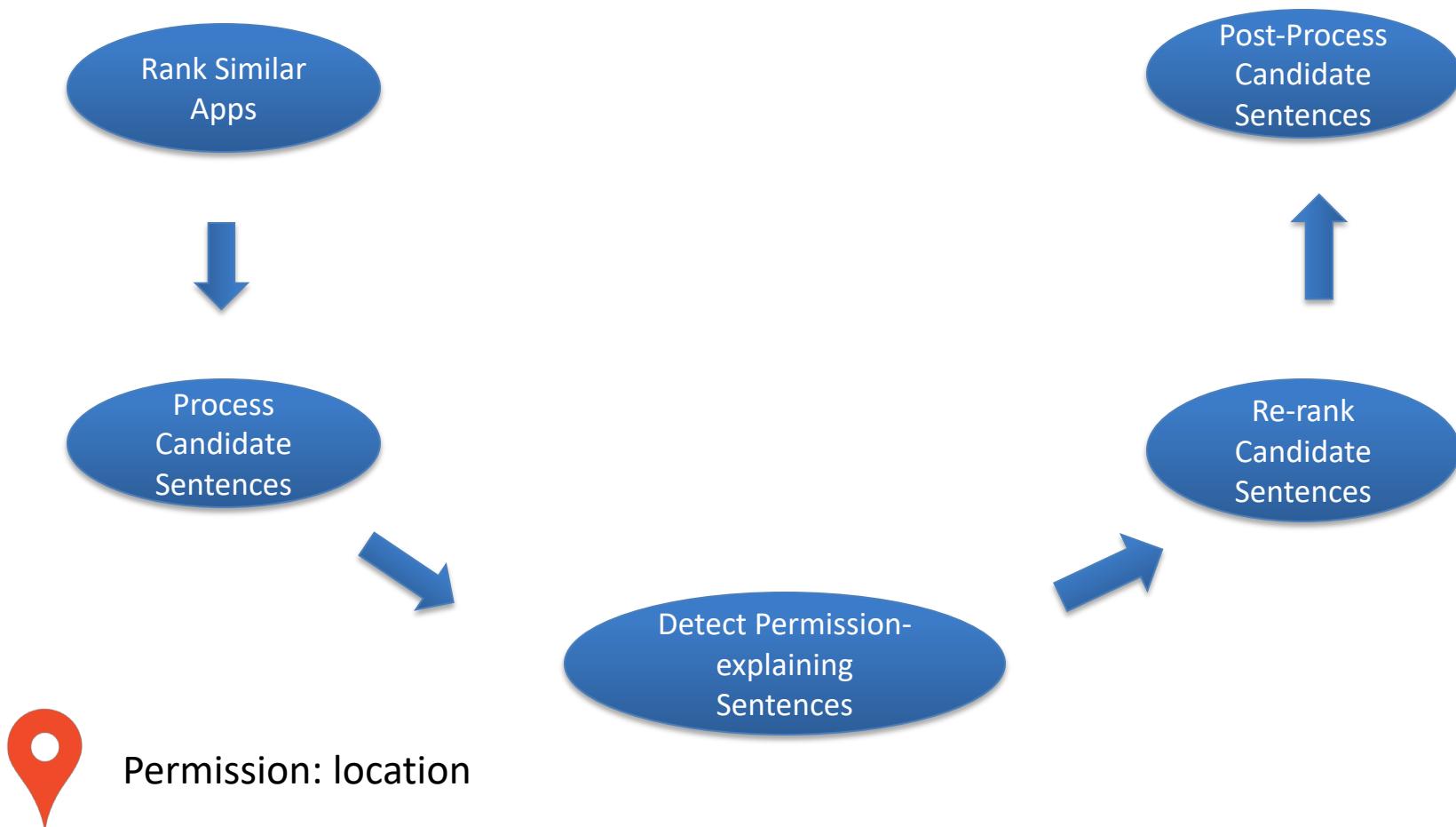


Permission: location

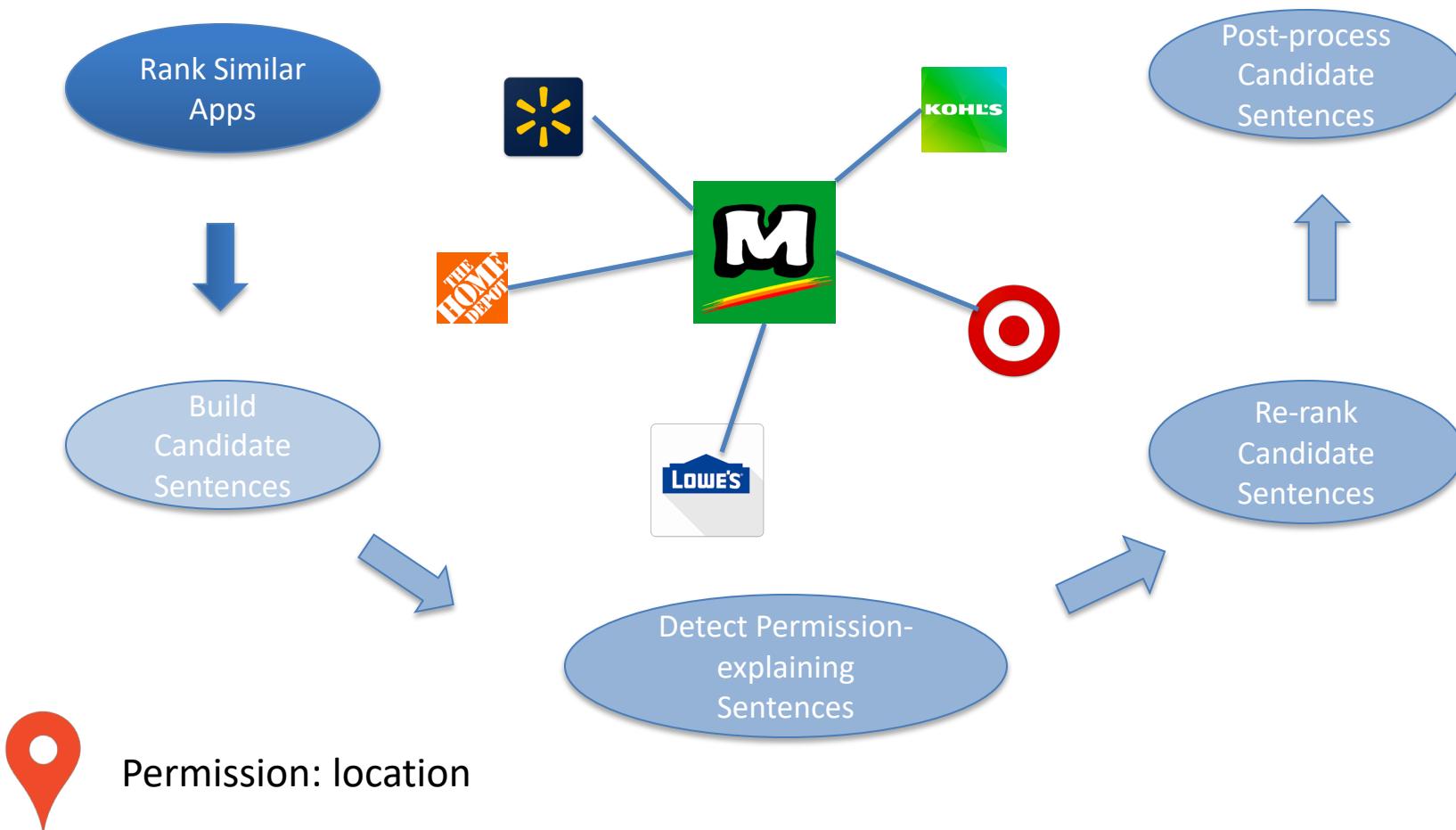


1. locate stores near you
2. Locator **inside** the store
3. use the **map view** to locate a store near you

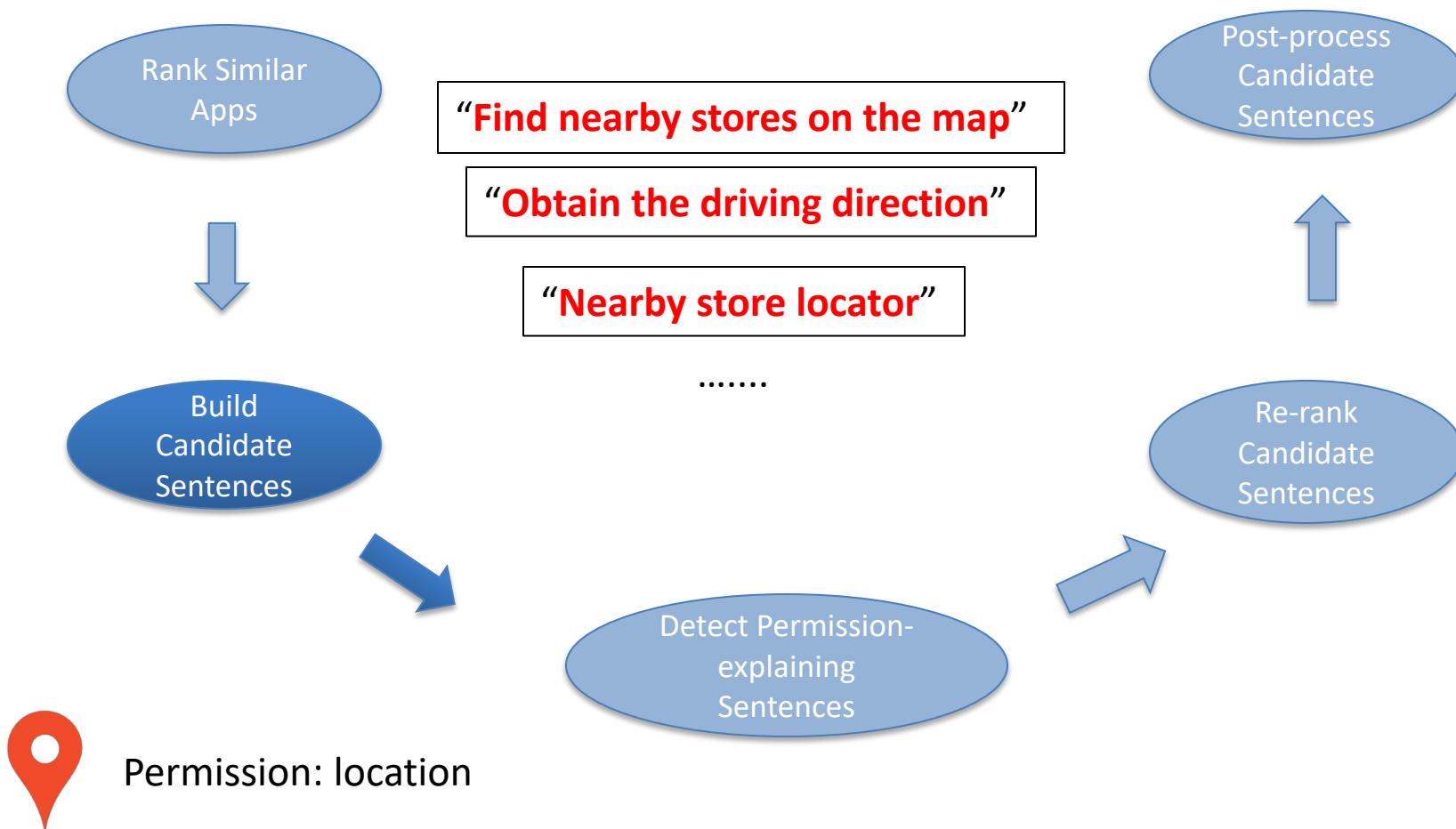
Overview of CLAP



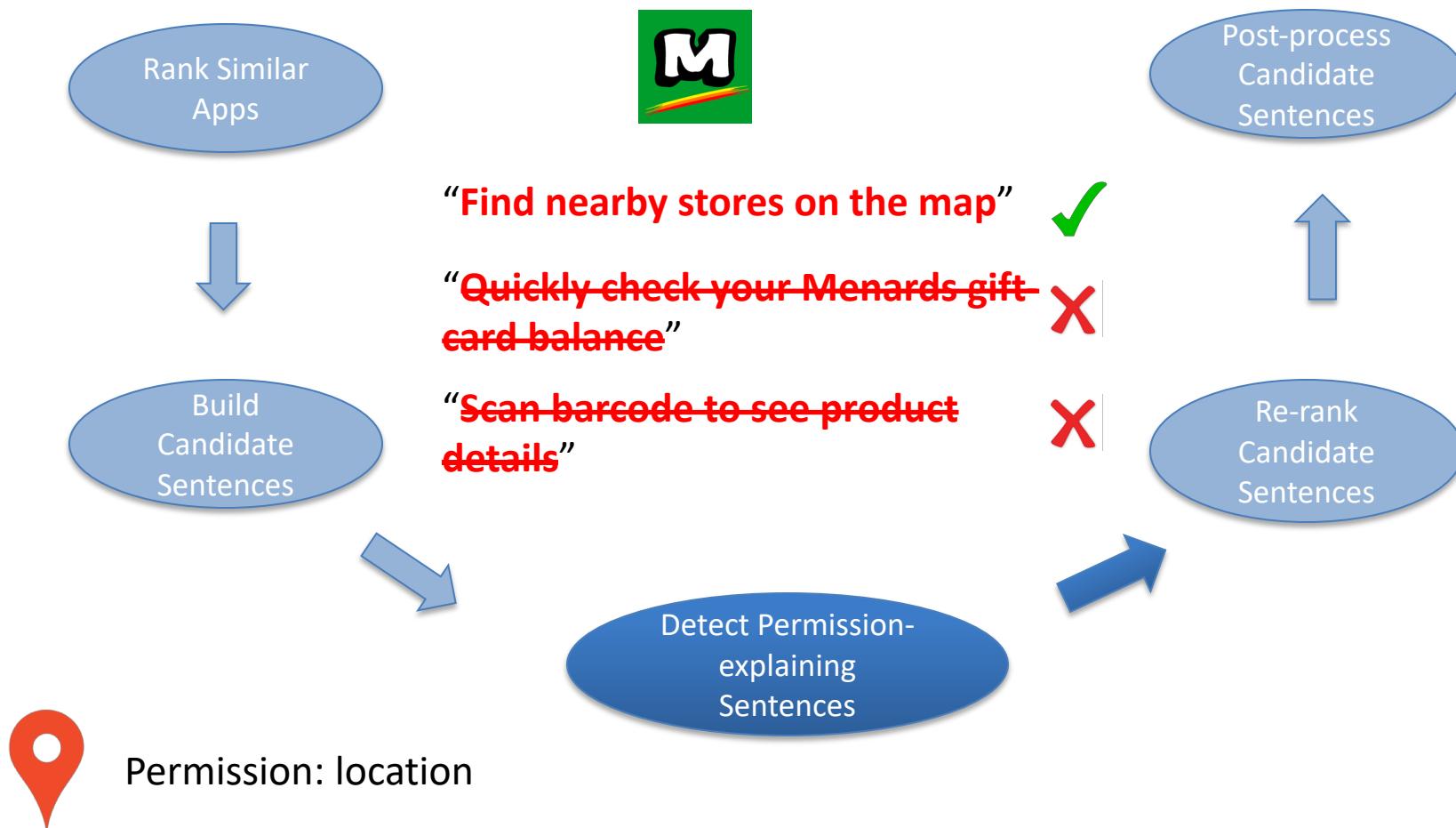
Step 1: Rank Similar Apps



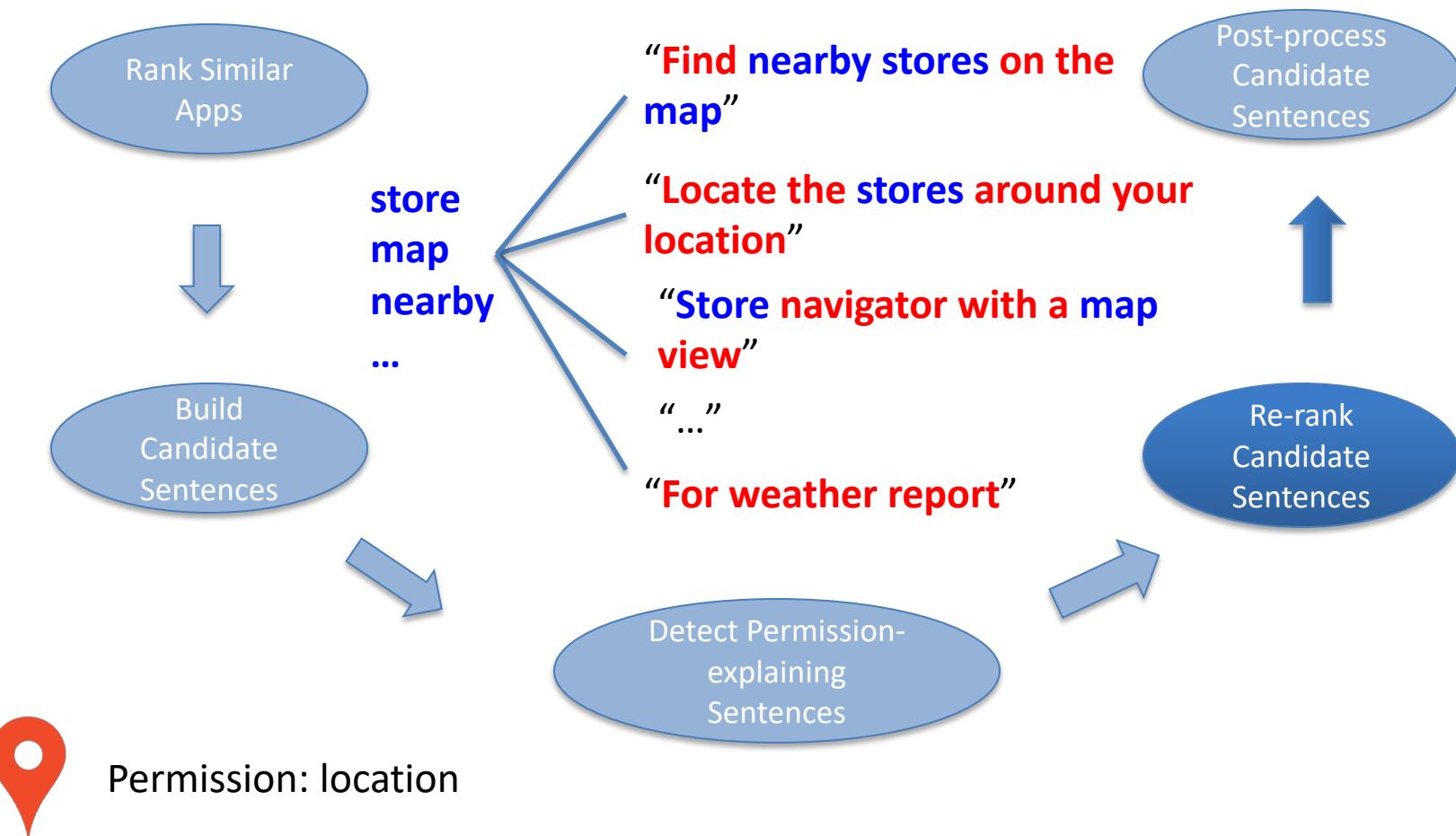
Step 2: Build Candidate Sentences



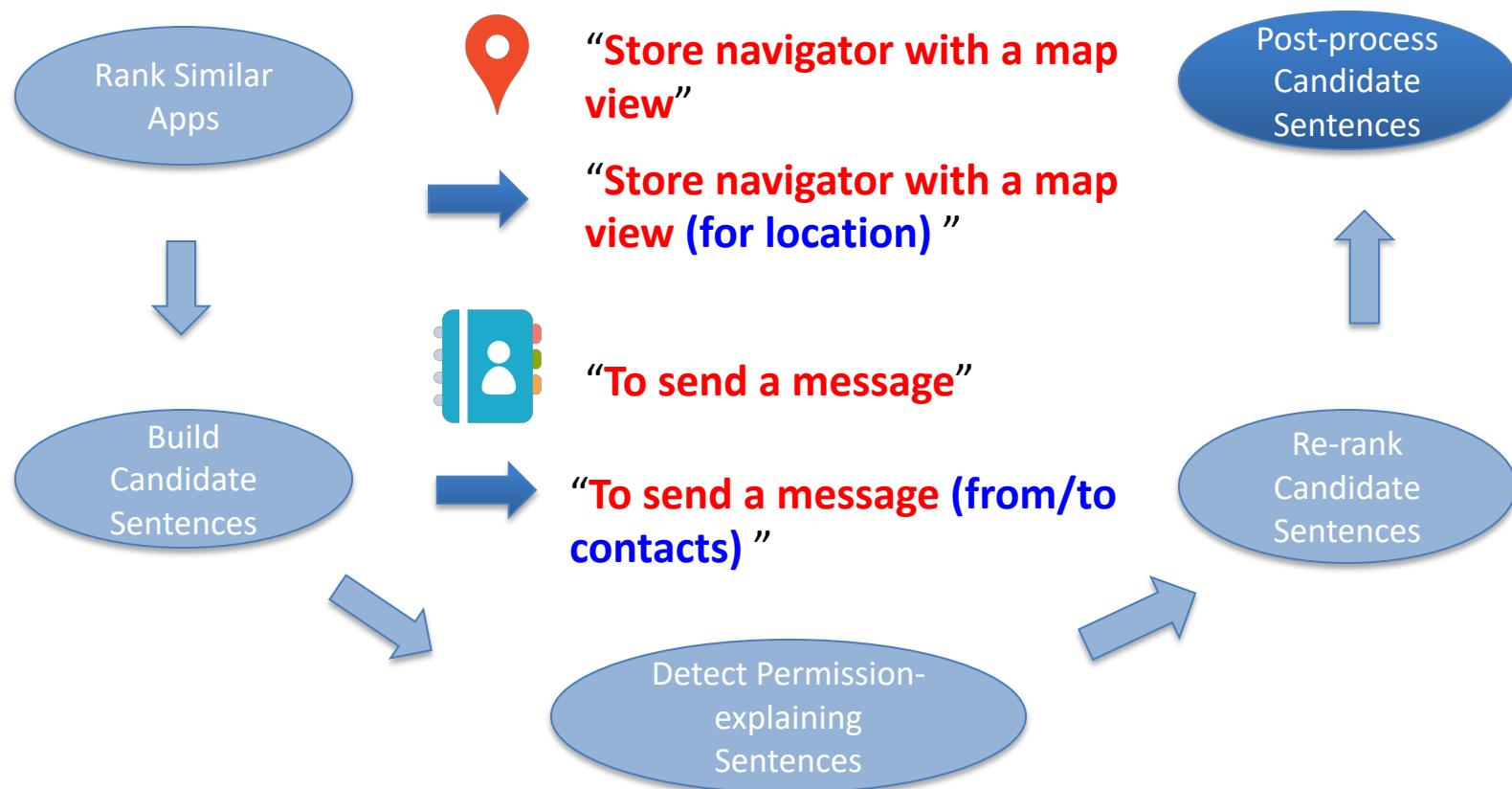
Step 3: Detect Permission-explaining Sentences



Step 4: Re-rank Candidate Sentences



Step 5: Post-process Candidate Sentences



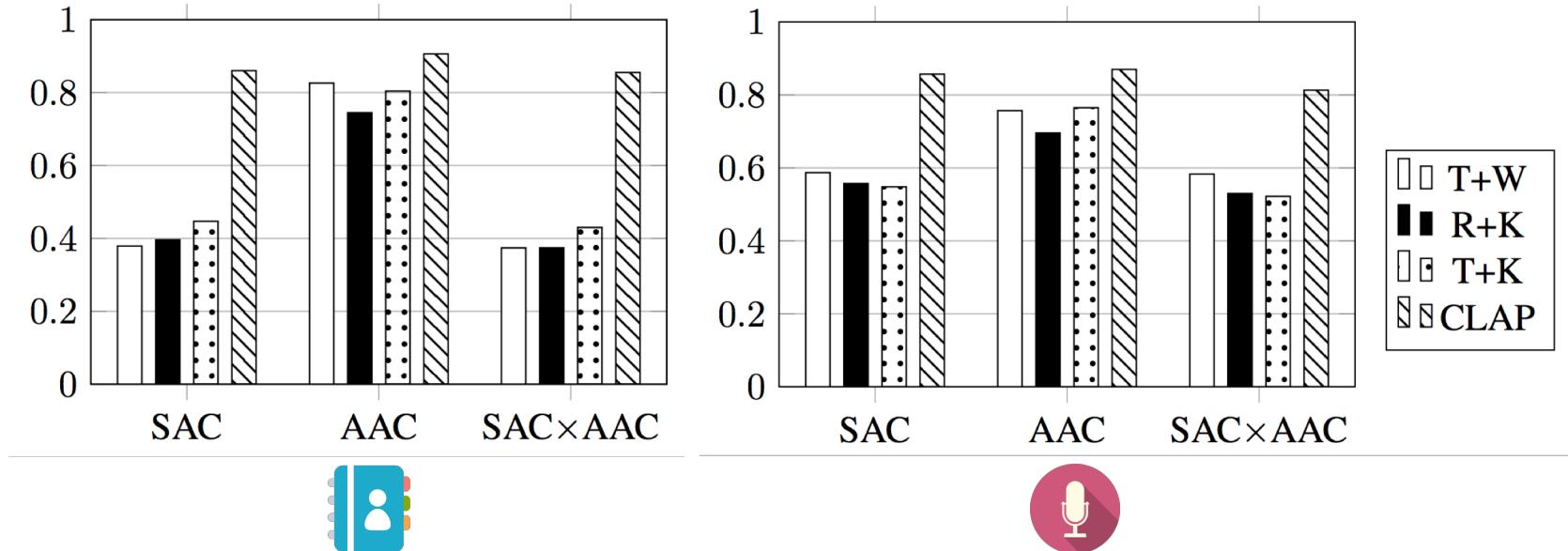
Evaluation

- RQ1: Are sentences recommended by CLAP **relevant** to the true purpose?
 - Large-scale quantitative study
- RQ2: Are sentences recommended by CLAP **interpretable**?
 - Qualitative study

RQ1: Dataset

- PlayDrone (1.4 million apps)
- True purpose sentences
 - App specified permission purposes: **827** apps, e.g., “--Permission Usage--: LOCATION: **Location is used for Geofence API**”
 - Author labeled permission purposes: **96** apps

RQ1: Relevance of Recommended Sentences



SAC: sentence accuracy



AAC: app accuracy



SAC x AAC: sentence & app accuracy

CLAP's relevance scores significantly outperform baselines!

CLAP's accuracy ~ 85%!

RQ2: Interpretability of Recommended Sentences

input app: Builddoc is an easytouse project-based photo documentation application that allows you to capture field issues and assign and manage team member's tasks....



Ground truth: to record **audio and voice notes**

URL: <https://mighty.signal.com/a/google-play/com.app.builddoc1>

CLAP outputs:

- Creating **audio notes** using the device microphone (to record voice);
- Use your own (recorded) voice to **create audio notes**
- Record **voice notes** to **explain expenses**
- Compose text notes using (recorded) speech to text and voice commands;
- Capture photo of a book and record reading it to your child



Permission: microphone

CLAP sentences match the ground truth sentence semantically!

RQ2: Interpretability of Recommended Sentences

Substantial purposes

Counter example:
“to record voice”

CLAP outputs:

- Creating audio notes using the device microphone (to record voice);
- Use your own (recorded) voice to create audio notes
- Record voice notes to explain expenses

Diverse wording choices

concise

CLAP sentences show 3 characteristics of interpretability!



Permission: microphone

RQ2: Quantitative Evaluation on Conciseness

(word number of output sentences)

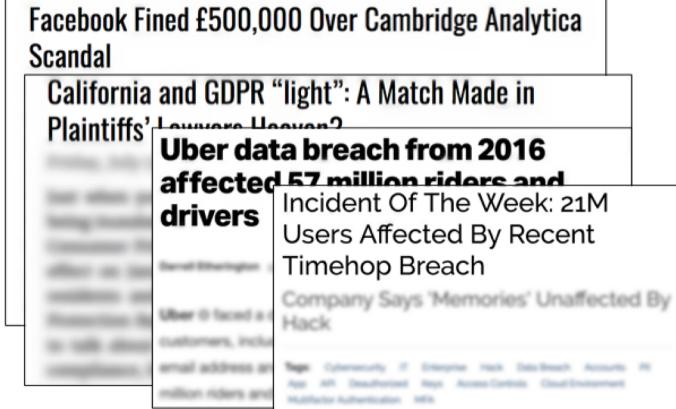
	CLAP	T+K	T+ W	R + W
average	8.1	14.6	14.3	15.6
maximum	31	174	174	86

CLAP recommended sentences are 50% shorter than baselines!

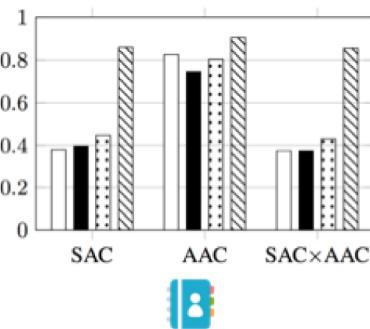
Discussions

- Real user study
- Relies on the number of similar apps

Conclusion



Challenges in Android Security



Higher relevance

GeoTimer: "...GeoTimer harnesses Google's **Geofence API...**"



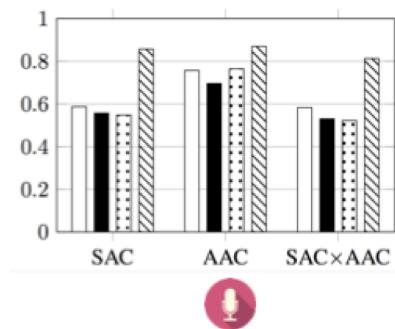
URL: <https://androidappapk.co/detail-geotimer-lite/>

Why does Geofence API need my location?



(our anonymous reviewer)

Challenges in Explaining Permissions



Interpretability of Recommended Sentences

- Diverse wording choices for improving explanations
- Deliver substantial permission purposes
- Concise sentences

Good interpretability



Cheng Zhai

Thank you!



Tao Xie

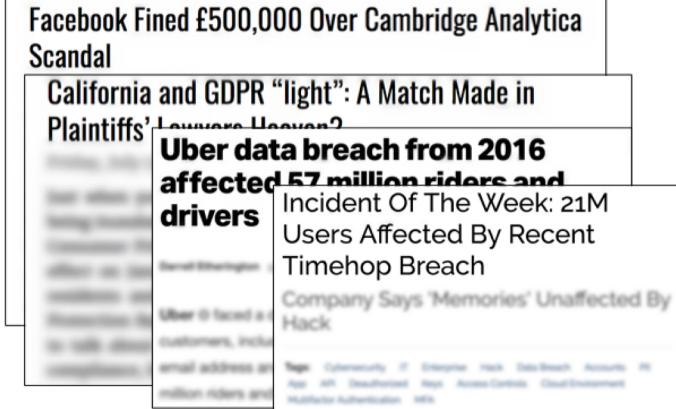
Acknowledgment: NSF CNS-1513939, NSF CNS-1408944, NSF CCF-1409423, and NSF CNS-1564274

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

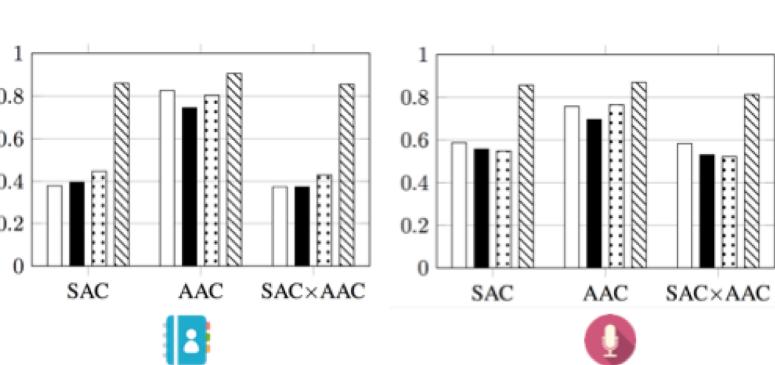


illinois.edu

Conclusion



Challenges in Android Security



Higher relevance

GeoTimer: "...GeoTimer harnesses Google's **Geofence API...**"



URL: <https://androidappsapk.co/detail-geotimer-lite/>

Why does Geofence API need my location?



(our anonymous reviewer)

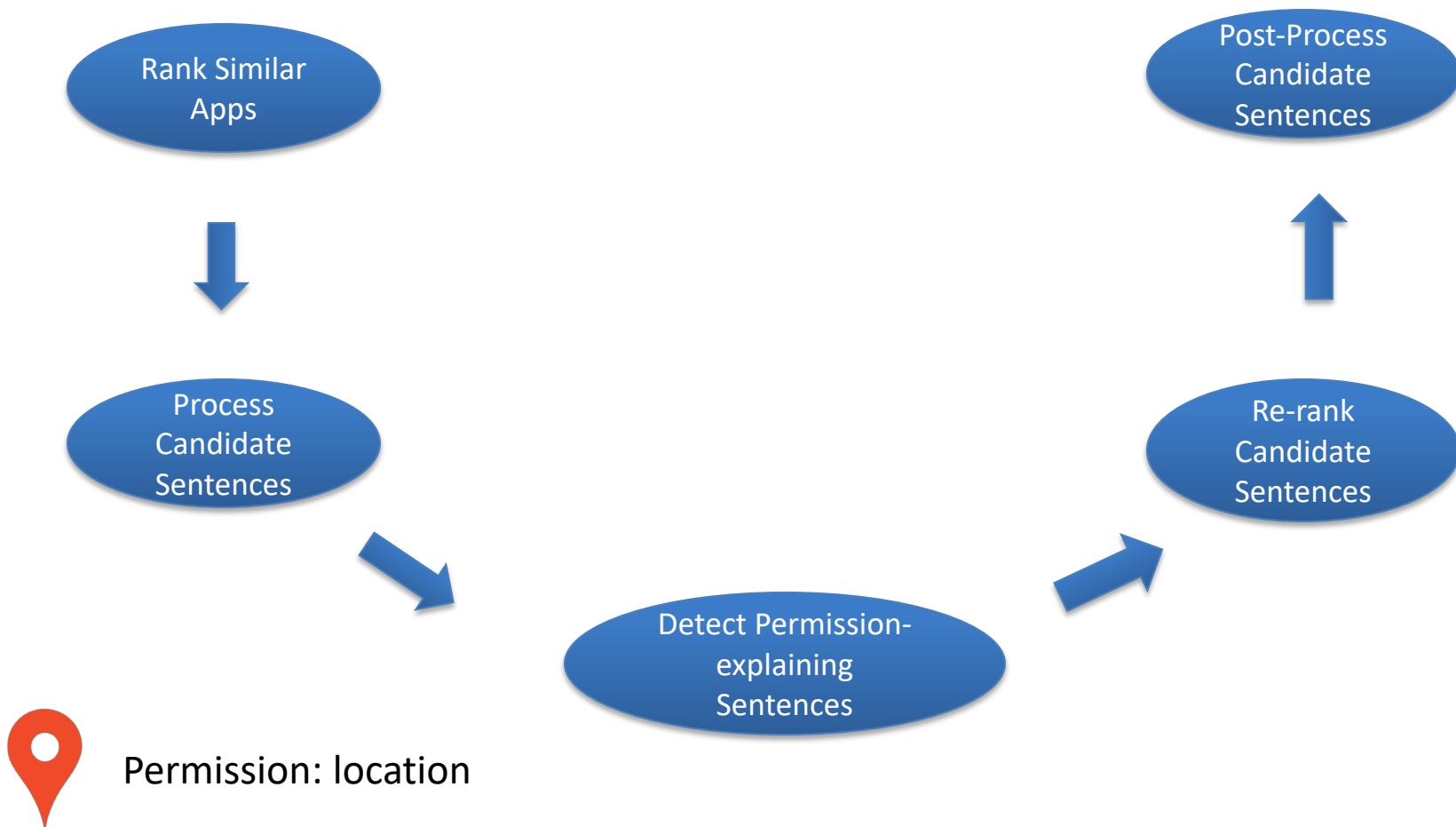
Challenges in Explaining Permissions

Interpretability of Recommended Sentences

- Diverse wording choices for improving explanations
- Deliver substantial permission purposes
- Concise sentences

Good interpretability

Overview of CLAP



RQ1: Relevance of Recommended Sentences

Jaccard Index: (exact-match)

		CONTACT _{dev}			RECORD _{dev}			LOCATION _{dev}		
		top1	top3	top5	top1	top3	top5	top1	top3	top5
JI	T+K	0.015	0.015	0.014	0.054	0.052	0.054	0.019 [†]	0.019 [†]	0.019 [†]
	T+W	0.023 [†]	0.026 [†]	0.026 [†]	0.092	0.087 [†]	0.086 [†]	\	\	\
	R+K	0.013	0.008	0.008	0.042	0.044	0.043	0.014	0.012	0.012
	CLAP	0.032	0.036	0.037	0.091 [†]	0.105	0.103	0.027	0.025	0.023
	p	0.18	0.07	0.03	\	0.16	0.15	0.04	0.03	0.03

Word embedding similarity (semantic)

WES	T+K	0.012	0.013	0.012	0.041	0.040	0.040	0.014 [†]	0.014 [†]	0.014 [†]
	T+W	0.016 [†]	0.018 [†]	0.019 [†]	0.061 [†]	0.060 [†]	0.060 [†]	\	\	\
	R+K	0.012	0.010	0.010	0.039	0.035	0.038	0.010	0.010	0.010
	CLAP	0.031	0.033	0.033	0.079	0.084	0.081	0.025	0.023	0.021
	p	3e-4	2e-4	5e-4	0.11	9e-3	9e-3	6e-5	3e-6	5e-7

T test between CLAP (bold) and the second best baseline (dagger)

CLAP's relevance scores significantly outperform baselines!