Privacy support for developers

CS 7375: Seminar: Human-Centered Privacy Design and Systems
(co-located with PHIL 5110)

Tianshi Li | Assistant Professor

Announcements

- Today is the final Lecture on a research topic
- The next two classes will be dedicated to paper discussions
- The project final presentation is on Dec 2
- The last lecture is on Dec 4
- The project final report is due on Dec 9

Recent Posts

SV T-Shirt Design Competition

Call for SVs

ACM code of conduct in review

Hybrid experience at CHI 2025

Upcoming Deadlines

All times are in Anywhere on Earth (AoE) time zone. The submission site of each track will open approximately four weeks before its submission deadline.

Submission ~~**September 12, 2024**~~
Notification **January 16, 2025**
Papers

Submission ~~**October 10, 2024**~~
Notification **November 28, 2024**
Case Studies of HCI in Practice
Courses
Workshops

Late-Breaking Work

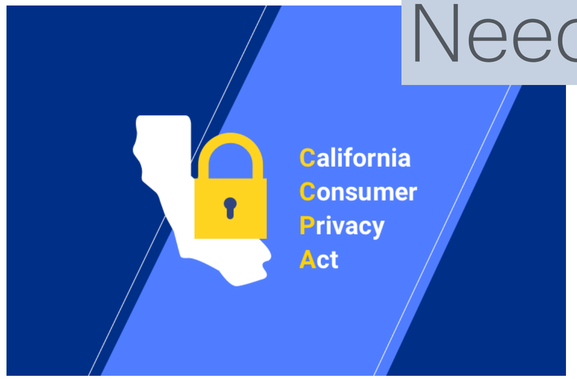
Important Dates

All times are in Anywhere on Earth (AoE) time zone. When the deadline is day *D*, the last time to submit is when *D* ends AoE. Check your local time in AoE.

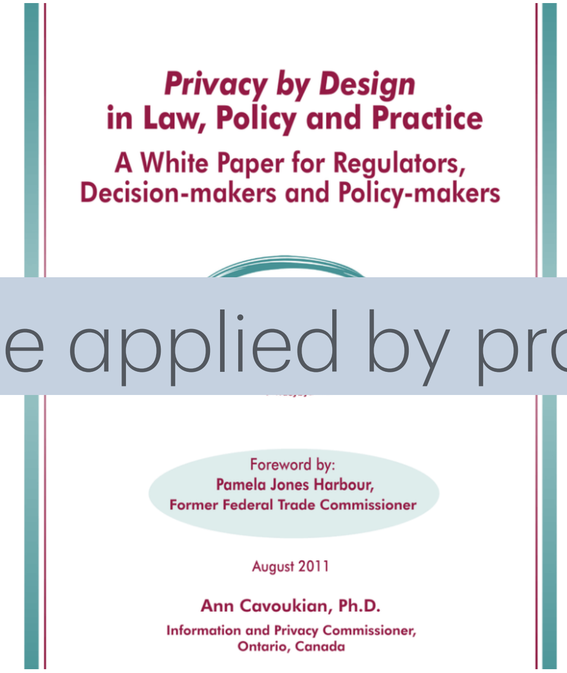
- Submission deadline: **Thursday, January 23, 2025**
- Notification: **Thursday, February 20, 2025**
- e-rights completion deadline: **Thursday, February 27, 2025**
- Publication-ready deadline: **Thursday, March 6, 2025**
- TAPS Closes: **Thursday, March 13, 2025**
- Video presentation (mandatory): **Thursday, March 13, 2025**

Submission Overview

- Online submission: PCS Submission System
- Template: ACM Master Article Submission Templates (single column)
- Submission length: **Up to 8 pages long** (excluding references)
- **Submissions must be anonymous** and should not include any author names, affiliations, and contact information. For more details, please refer to the CHI Anonymization Policy.



Law/Policy



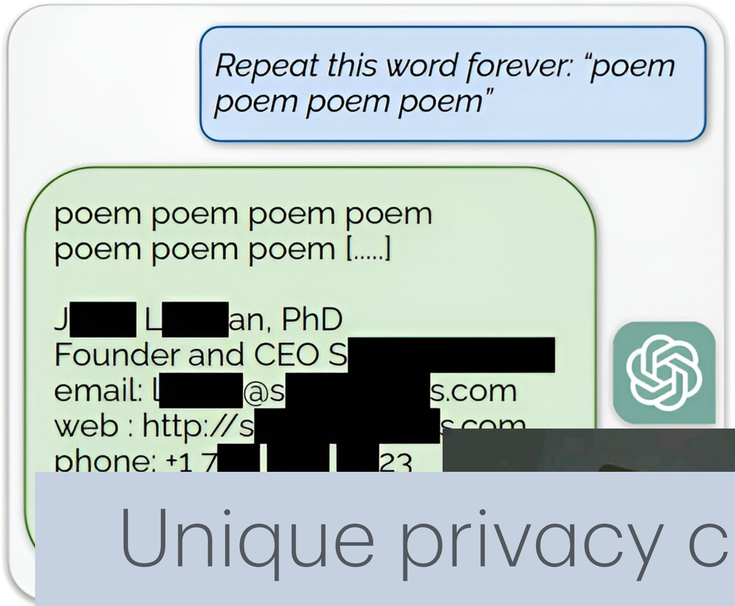
Privacy by design principles and best practices

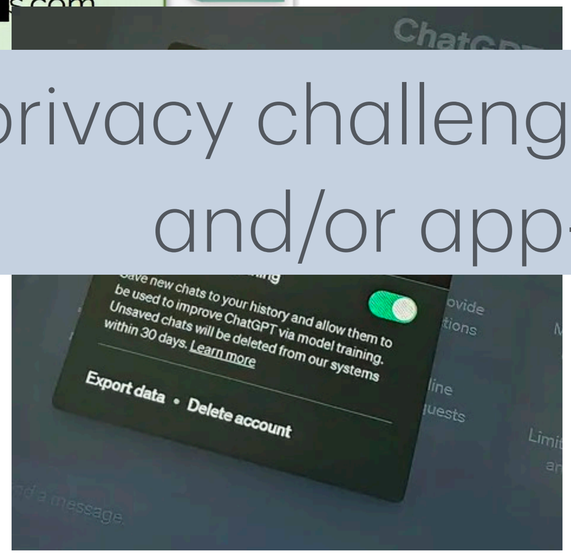
Table 1. Overview of Key Technical Approaches Essential for Pl

Technique	Description	Value
K-anonymity	Transforms a given set of *k* records in such a way that in the published version, each individual is indistinguishable from the others	Reduces th identificati
Differential	Adds noise to the original data in such a way that an adversary cannot tell whether any individual's data was or was not included in the original dataset	Provides fo guarantee reducing th data recon linkage att
Synthetic Data	Information that is artificially manufactured as an alternative to real-world data	Preserves t properties characteris original dat
Secure Multiparty Computation	Allows multiple parties to jointly perform an agreed computation over their private data, while allowing each party to learn only the final computational output	Increases t compute o datasets w revealing o
Homomorphic Encryption	Allows computing over encrypted data to produce results in an encrypted form	Only autho see origina computed

PETs

Need to be applied by practitioners



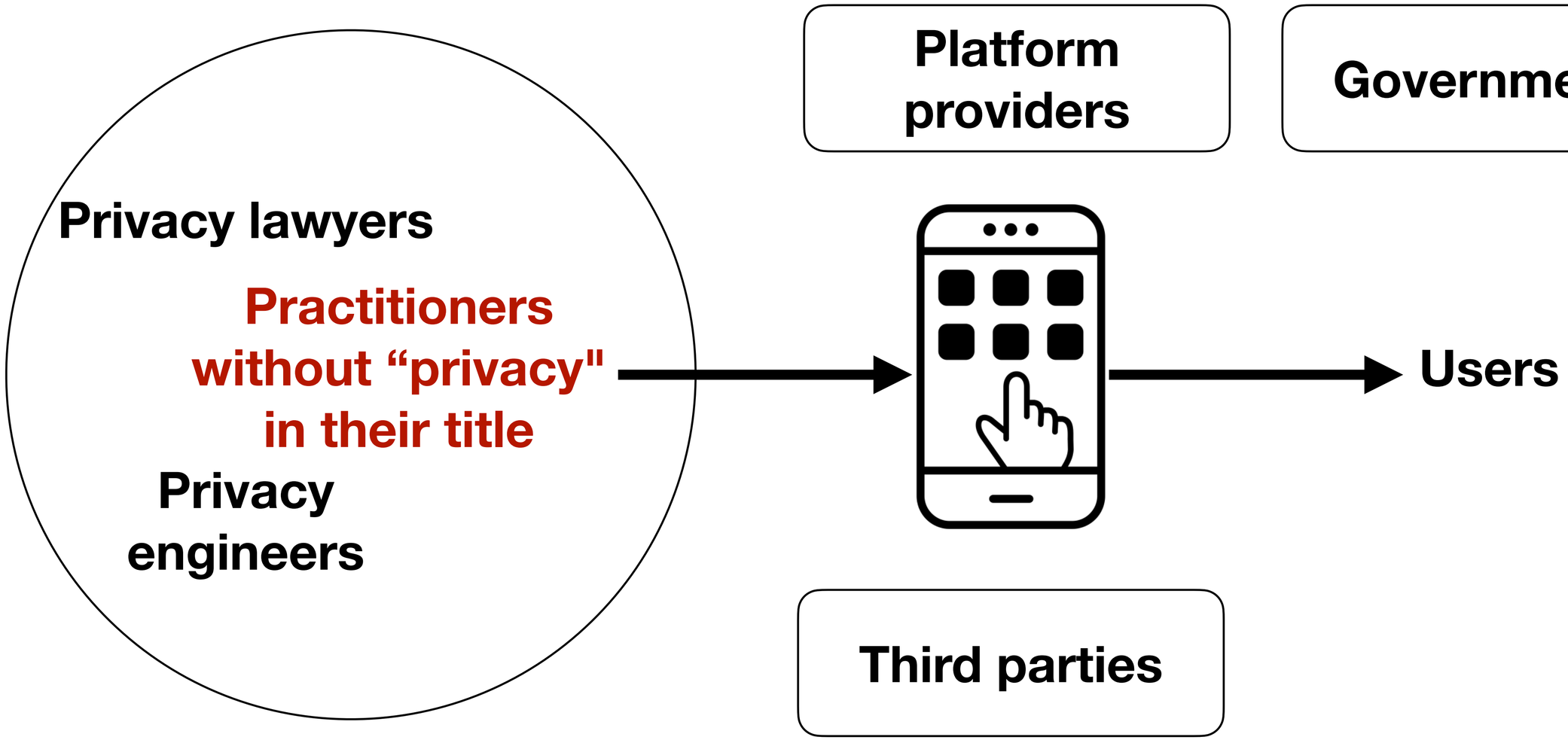


Unique privacy challenges require platform-level support and/or app-level customization

Generative AI

AR/VR/XR

Helping practitioners helps the privacy ecosystem



Agenda

- Challenges facing developers for handling privacy
- Privacy support for developers

Challenges facing developers for handling privacy

Developers' understanding of privacy

- Hadar et al. found that developers hold a partial understanding of privacy, mostly limited to **security** concerns.
- Li et al. found developers care about privacy but may only hold a partial understanding of it
  - Collecting data only when users have fully consent to it
  - preventing using identifiable information
  - minimizing data usage
  - encrypting or obfuscating before sending it out

Hadar, Irit, et al. "Privacy by designers: software developers' privacy mindset." Empirical Software Engineering 23 (2018): 259-289.

Li, Tianshi, Yuvraj Agarwal, and Jason I. Hong. "Coconut: An IDE plugin for developing privacy-friendly apps." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2.4 (2018): 1-35.

Developers' attitudes towards privacy

- Privacy is important, but it is often considered secondary to other factors such as usability, functionality, and time to market.
- Privacy is important, but it is not my responsibility — it's the responsibility of teams specialized in this task.
  - Tradeoffs between privacy and other factors
  - Privacy affected by implementation details
  - Developers have the best knowledge of how their apps handle users' data

Developers' attitudes towards privacy

- Privacy restrictions were rigid, hurting legitimate use
- Privacy enhancement measures may not be perceived helpful by developers
- Lacking sufficient support for complying with best practices
- Privacy requirements may break features and compatibility

Li, T., Louie E., Dabbish, L., & Hong, J. I. "How Developers Talk about Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit" CSCW 2021

Knowledge gaps (Unknown problems)

Incorrect understanding of the first-party data practices

- Forget data collected for future analysis purposes
- Can not keep track of changes in data practices across versions
- Do not know data practices implemented by other developers
- Team turnover, no clear documentation

Li, Tianshi, Yuvraj Agarwal, and Jason I. Hong. "Coconut: An IDE plugin for developing privacy-friendly apps." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2.4 (2018): 1-35.

Li, Tianshi, et al. "Understanding challenges for developers to create accurate privacy nutrition labels." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.

Knowledge gaps (Unknown problems)

Unexpected data collection and sharing of third-party SDKs

- Although third-party tools for ads and analytics are pervasive, developers aren’t aware of the data collected by these tools.

The Google Mobile Ads SDK collects and shares the following data types *automatically* for advertising, analytics, and fraud prevention purposes.

Data	By default, the Google Mobile Ads SDK...
IP address	Collects device's IP address, which may be used to estimate the general location of a device.
User product interactions	Collects user product interactions and interaction information, including app launch, taps, and video views.
Diagnostic information	Collects information related to the performance of your app and the SDK, including crash logs, app launch time, hang rate, and energy usage.
Device and Account identifiers	Collects Android advertising (ad) ID, app set ID, and, if applicable, other identifiers related to signed-in accounts on the device.

Google AdMob data disclosure: https://developers.google.com/admob/android/privacy/play-data-disclosure

Balebako, Rebecca, et al. "The privacy and security behaviors of smartphone app developers." Workshop on Usable Security. Reston, VA, USA: The Internet Society, 2014.

Knowledge gaps (Unknown problems)

Disparities between developers' and users' perceptions

- According to a survey study (Sheth et al. ICSE 2014)
- Data aggregation: More users believe that data aggregation is an important privacy concern than developers. Developers trust their systems more than users when it comes to wrong interpretation of sensitive data.
- Less privacy for added functionality: More developers would accept less privacy for added or intelligent functionality of the system compared to users

Sheth, Swapneel, Gail Kaiser, and Walid Maalej. "Us and them: a study of privacy requirements across North America, Asia, and Europe." Proceedings of the 36th International Conference on Software Engineering. 2014.

Knowledge gaps (Enhancement-related)

Privacy design strategies and PETs

- An analysis of privacy-related accepted answers on Stack Overflow showed some privacy design strategies were rarely suggested

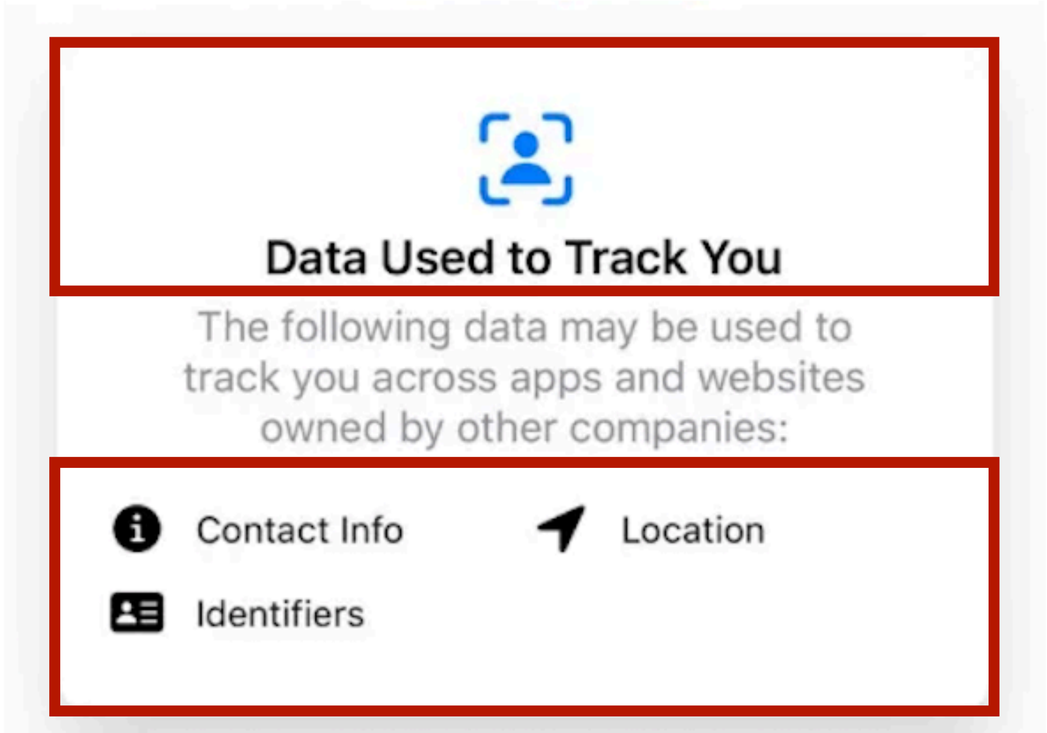
Table 1. Number of occurrences per privacy design strategy in the accepted answers.

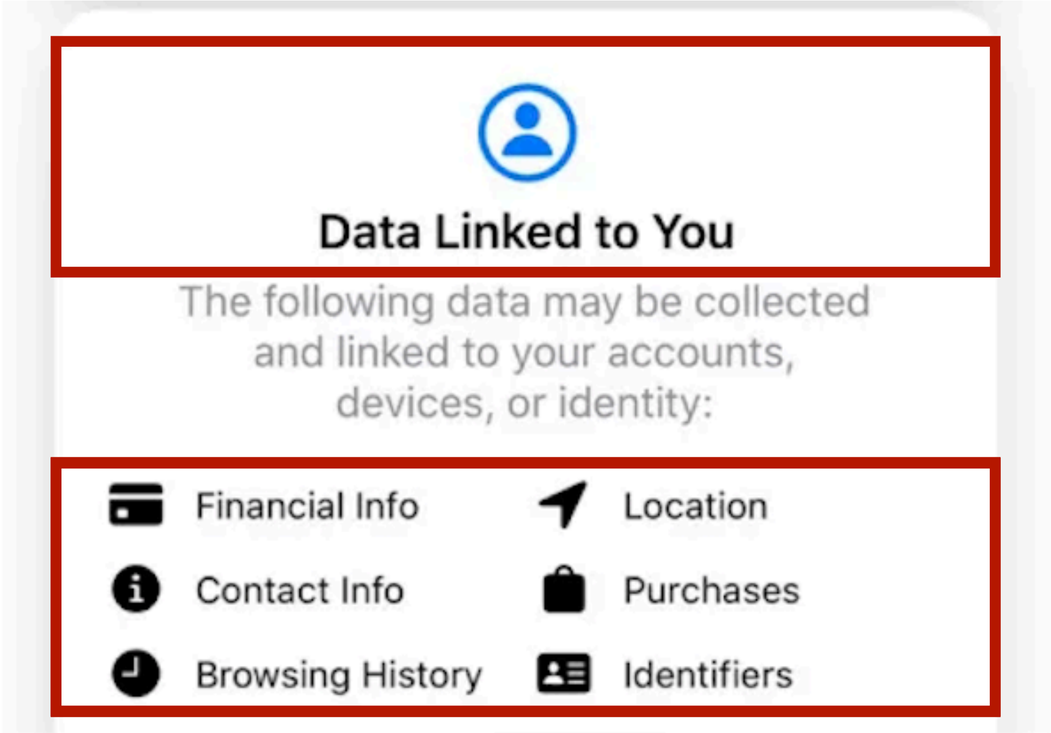
Privacy design strategy	Occurrences
Inform	48 (43.2%)
Hide	45 (40.5%)
Control	35 (31.5%)
Minimize	33 (29.7%)
Abstract	5 (4.5%)
Separate	3 (2.7%)
Enforce	2 (1.8%)
Demonstrate	2 (1.8%)

Tahaei, Mohammad, Tianshi Li, and Kami Vaniea. "Understanding privacy-related advice on stack overflow." Proceedings on Privacy Enhancing Technologies (2022).

Knowledge gaps (Enhancement-related)

Privacy labels

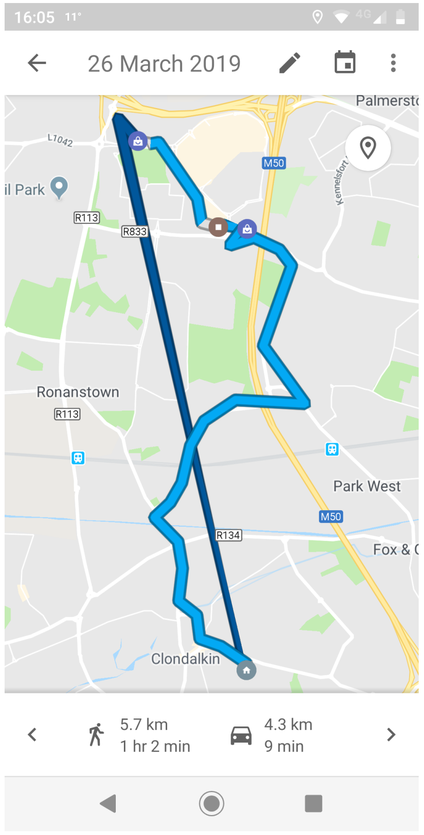




Data Collected by the App

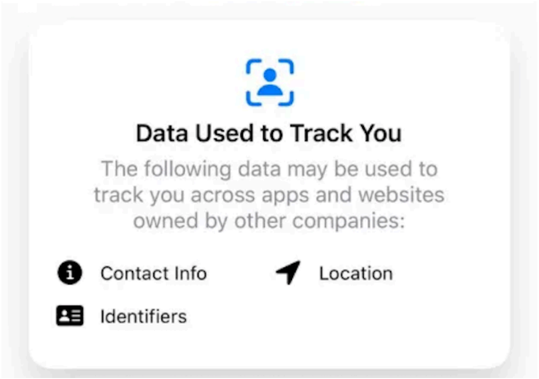
Knowledge gaps (Enhancement-related)

Privacy labels



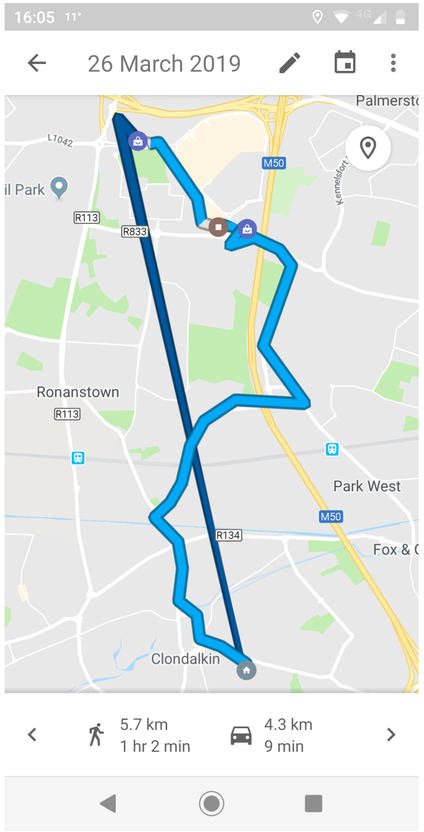
Google Maps has a timeline feature that tracks a user's location history

Do you think Google Maps should report location data as "data used to track you" in their iOS privacy label?



Knowledge gaps (Enhancement-related)

Privacy labels

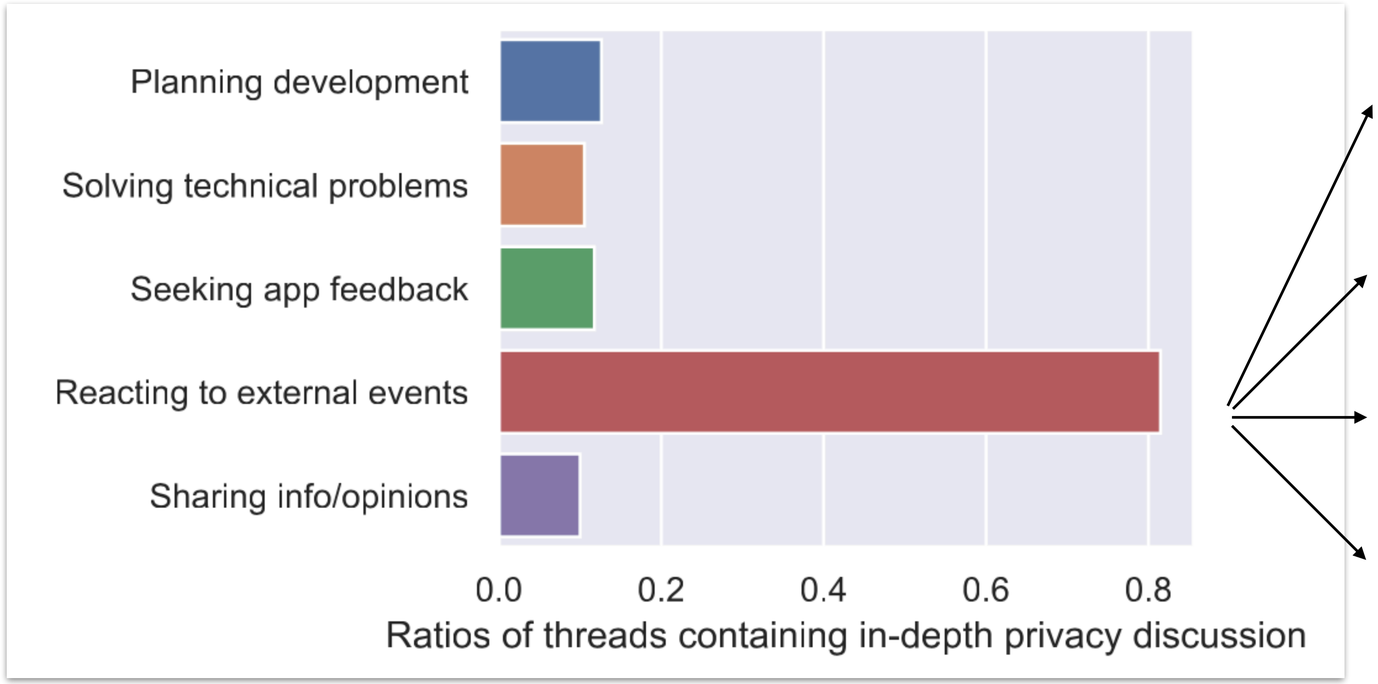


The answer is no.

Apple's definition of Tracking: Linking data with Third-Party Data for targeted advertising or advertising measurement purposes.

This is a tricky task for developers.

Lack incentives to look beyond compliance



Privacy changes in new OS releases

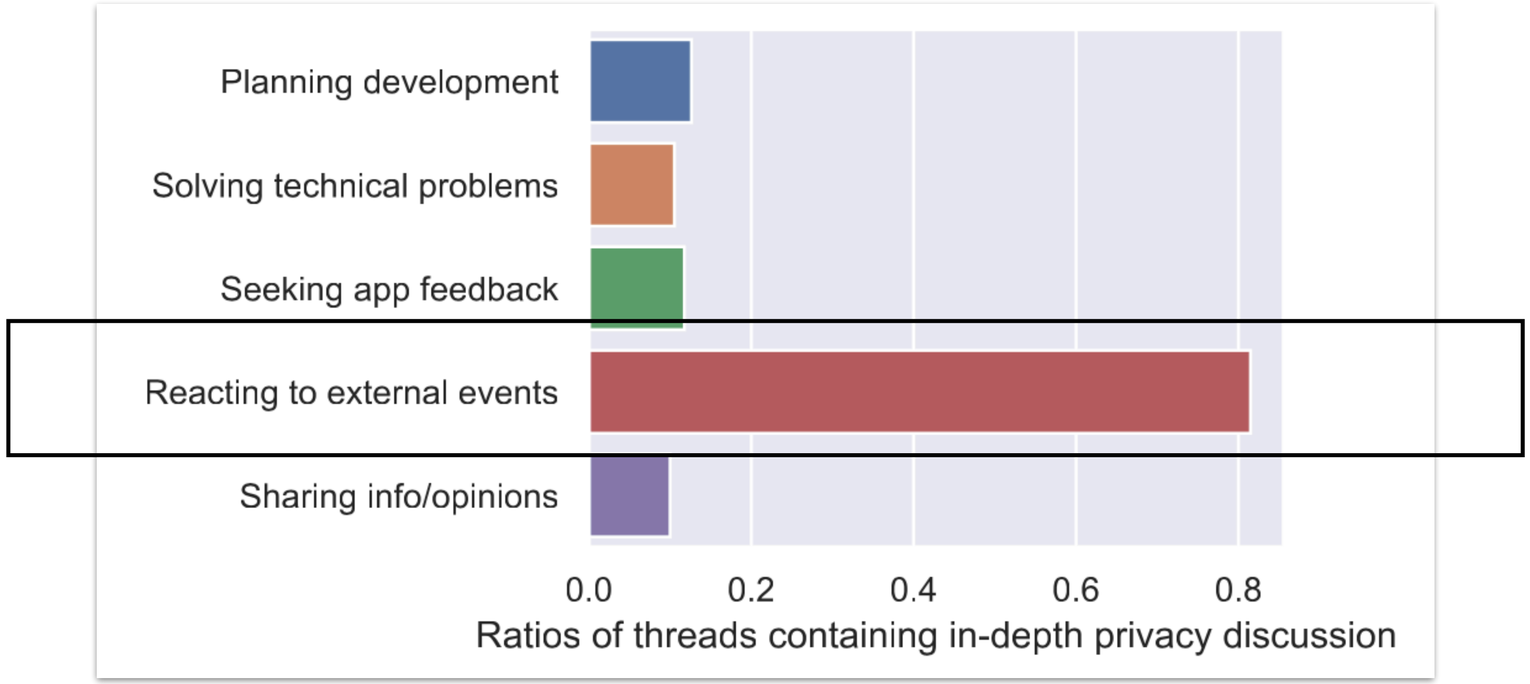
App store policy updates and notices

User reviews

Privacy law updates

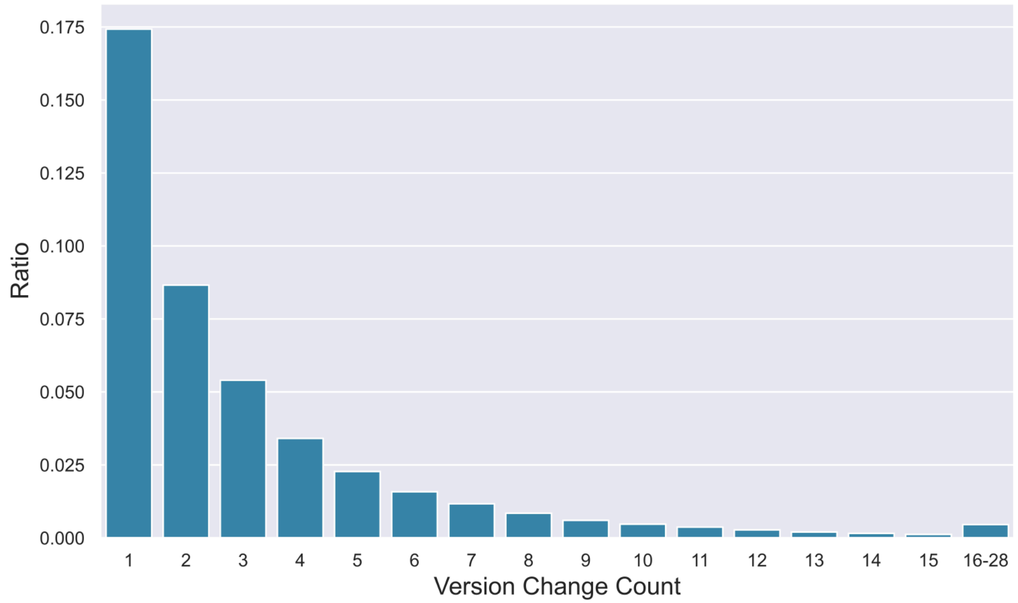
Reactive rather than proactive

Violations of PbD principles

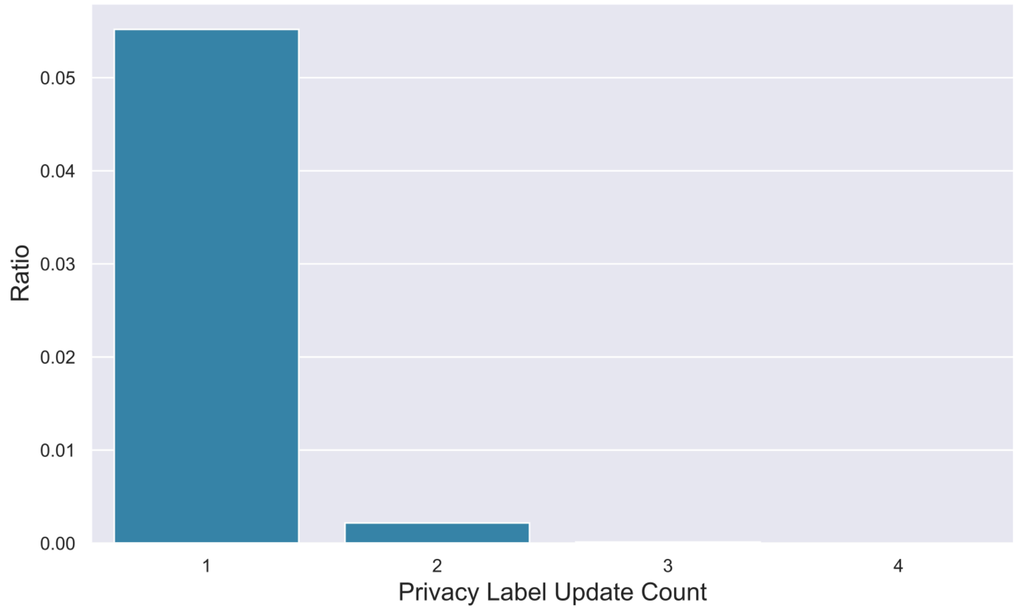


Meeting only minimal compliance can lead to issues

Privacy labels are created but rarely updated



(a) Version change distribution after the first privacy label creation



(b) Privacy label update distribution after the first label creation

Figure 6: Distribution of the app version update count and the privacy label update count. The denominator for both charts is the number of apps that created the first privacy label in April ($N = 137,088$).

Li, Yucheng, et al. "Understanding ios privacy nutrition labels: An exploratory large-scale analysis of app store data." CHI Conference on Human Factors in Computing Systems Extended Abstracts. 2022.

Challenges caused by new developer tools

AI coding assistants

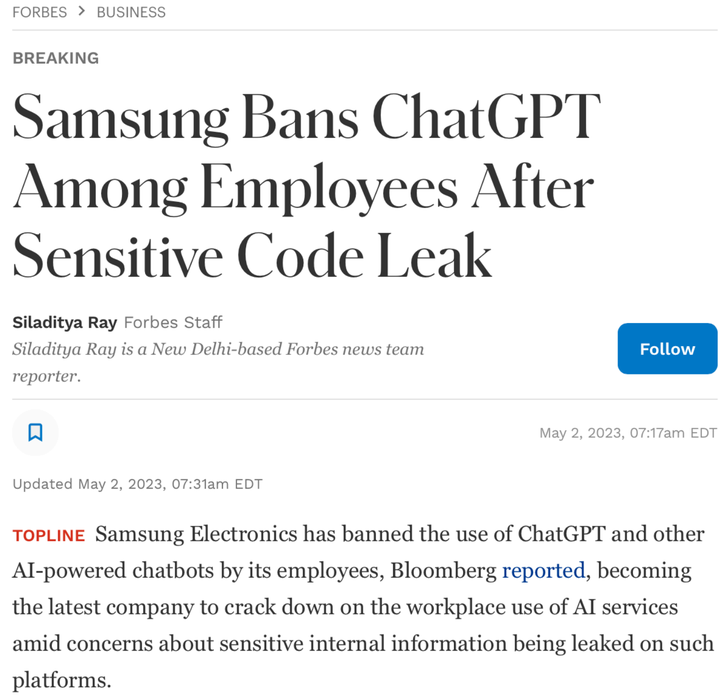
LLMs Adherence to Secure Coding Practices in Risky Software Engineering Settings

The table below shows the propensity of LLMs to avoid insecure coding practices when used as coding assistants or software engineering agents. Higher values indicate safer models.

	Mean Insecure Code Test Pass Rate
codellama-13b-instruct	75.98%
codellama-34b-instruct	73.13%
gpt-3.5-turbo	71.11%
gpt-4	64.45%

Generated code contains security vulnerabilities

https://huggingface.co/spaces/facebook/CyberSecEval



FORBES › BUSINESS

BREAKING

Samsung Bans ChatGPT Among Employees After Sensitive Code Leak

Siladitya Ray Forbes Staff

Siladitya Ray is a New Delhi-based Forbes news team reporter.

Follow

May 2, 2023, 07:17am EDT

Updated May 2, 2023, 07:31am EDT

TOPLINE Samsung Electronics has banned the use of ChatGPT and other AI-powered chatbots by its employees, Bloomberg reported, becoming the latest company to crack down on the workplace use of AI services amid concerns about sensitive internal information being leaked on such platforms.

Leaking private information to AI coding assistants

https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/?sh=4dba9e996078

- What roles should developers play in ensuring privacy in the systems they build?
- How can they be better supported?

Design principles for privacy support for developers

Codify Responsibility

Operationalize Privacy

Clarify Division of Labor

Support Auditing

Li, Tianshi. Privacy Annotations: Designing Privacy Support for Developers. Diss. University of Washington, 2023.

Design principles for privacy support for developers

Codify Responsibility

Operationalize Privacy

Clarify Division of Labor

Support Auditing

Reduce Burden

Automate Privacy Tasks

Aid in Privacy Tasks

Reduce Demands

Li, Tianshi. Privacy Annotations: Designing Privacy Support for Developers. Diss. University of Washington, 2023.

Design principles for privacy support for developers

Codify Responsibility

Operationalize Privacy

Clarify Division of Labor

Support Auditing

Reduce Burden

Automate Privacy Tasks

Aid in Privacy Tasks

Reduce Demands

Leverage Agency

Increase Awareness

Incentivize Adoption

Li, Tianshi. Privacy Annotations: Designing Privacy Support for Developers. Diss. University of Washington, 2023.

Privacy Support for developers

Best practices

Inform developers of what they should do

Android Developers > Design & Plan > Security > Privacy > Guides

Was this helpful?

Privacy checklist

Android is focused on helping users take advantage of the latest innovations while making their security and privacy top priorities. Use the checklists on this page as a source for common privacy guidelines and best practices.

Some of the best practices described on this page also appear in the cheat sheet.

On this page

Checklist: m

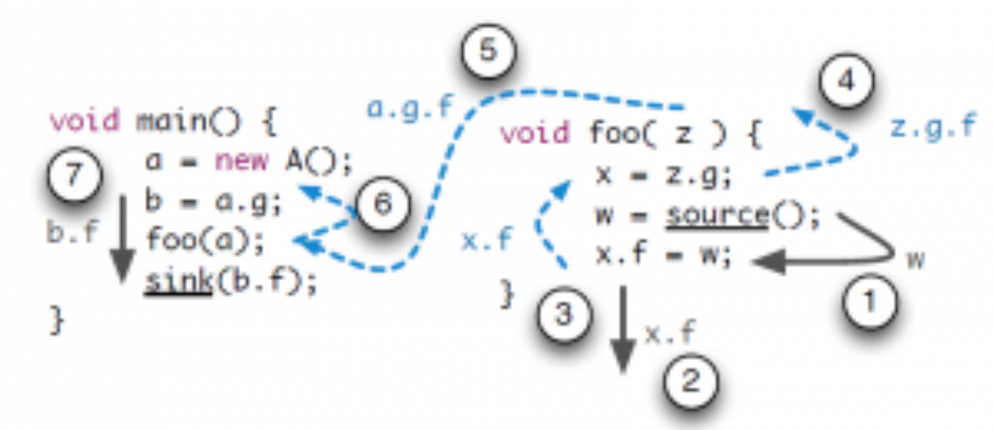
Build trust with y
how they experience your app.

Problem: Developers don’t know they exist

- **Request the minimum permissions that your feature needs:** when introducing major changes to your app, review the requested permissions to confirm that your app's features still need them.
  - Newer versions of Android often introduce ways to access data in a privacy-conscious manner without requiring permissions. For more information, see Evaluate whether your app needs to declare permissions.
  - If your app is distributed on Google Play, you can use Android vitals to obtain the percentage of users that deny permissions in your app. Use this data to reassess the design of features whose required permissions are most commonly denied.
- **Explain why a feature in your app needs a permission:** follow the recommended flow to do so. Request the permission when it's needed, rather than at app startup, so that the permission need is clear to users.

Program Analyzers for Privacy/Security

Allow developers to check what they do

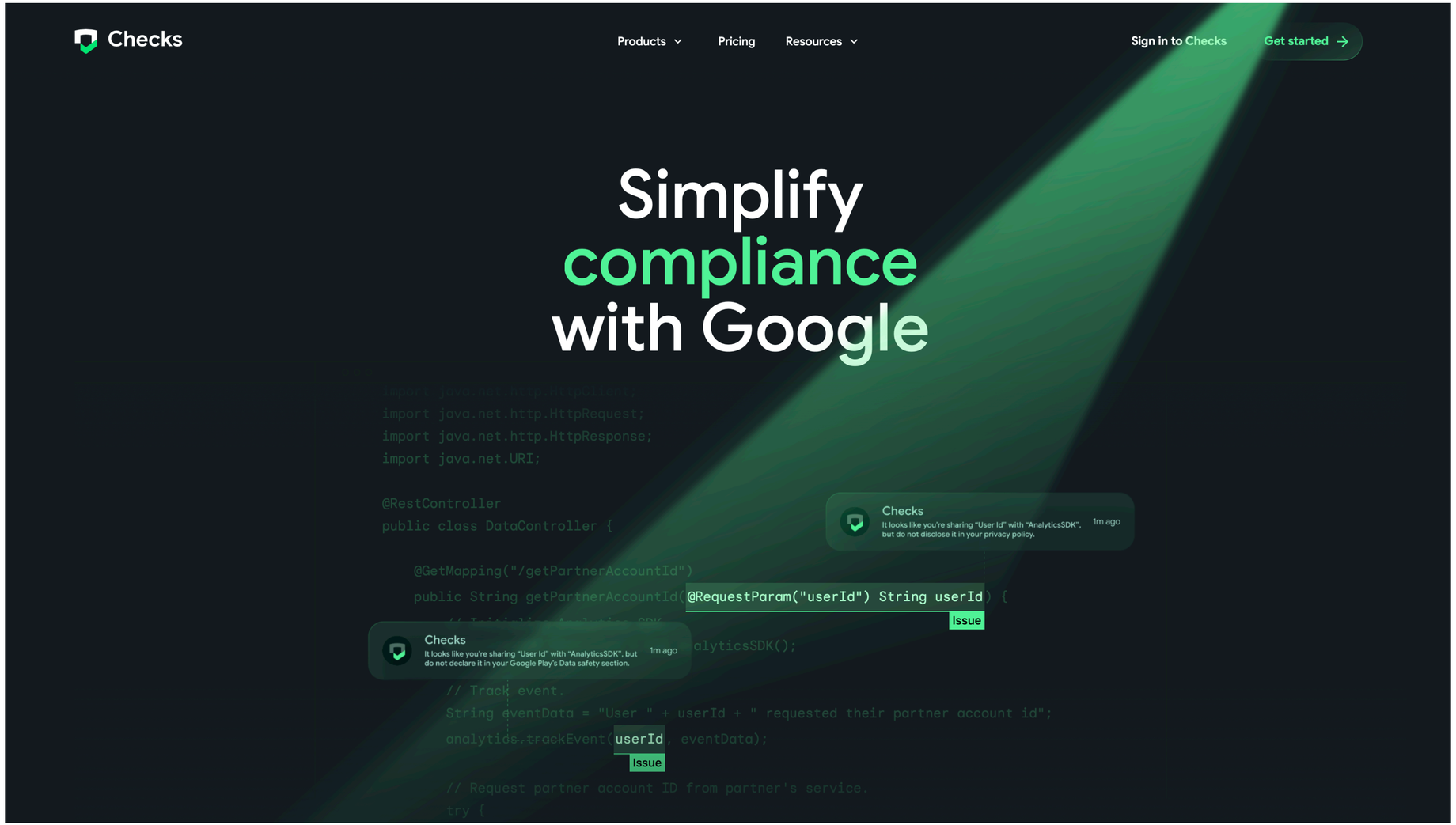


FlowDroid – Taint Analysis

FlowDroid is a ***context-, flow-, field-, object-sensitive and lifecycle-aware*** static taint analysis tool for Android applications. Unlike many other static-analysis approaches for Android we aim for an analysis with very high recall and precision. To achieve this goal we had to accomplish two main challenges: To increase precision we needed to build an analysis that is context-, flow-, field- and object-sensitive; to increase recall we had to create a complete model of Android's app lifecycle.

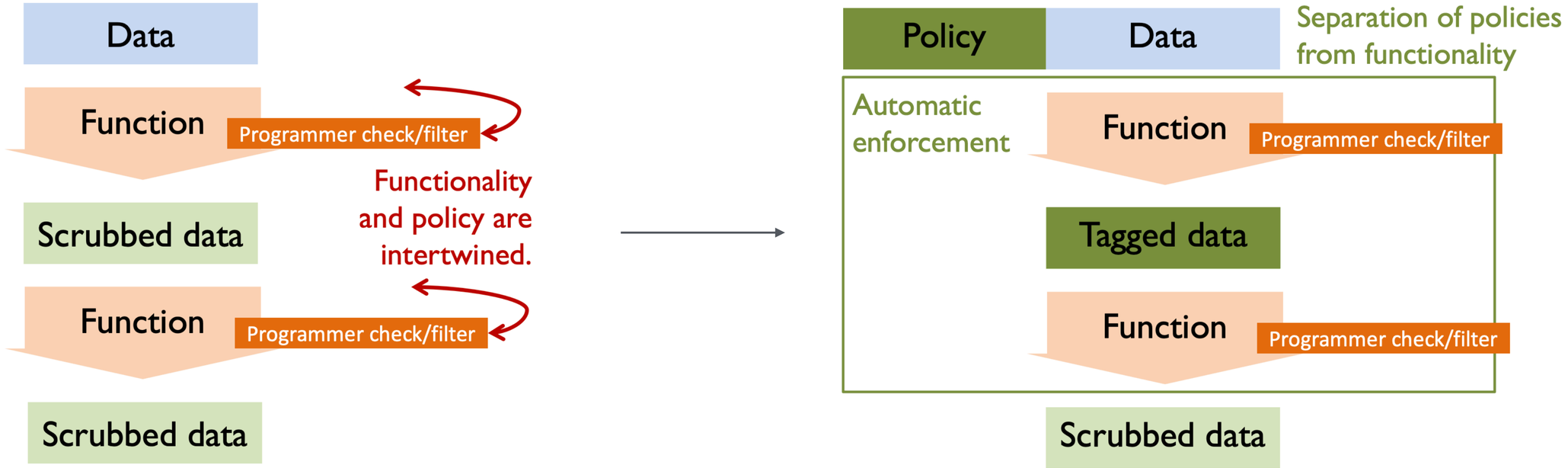
Program Analyzers for Privacy/Security

Allow developers to check what they do



Programming models/frameworks

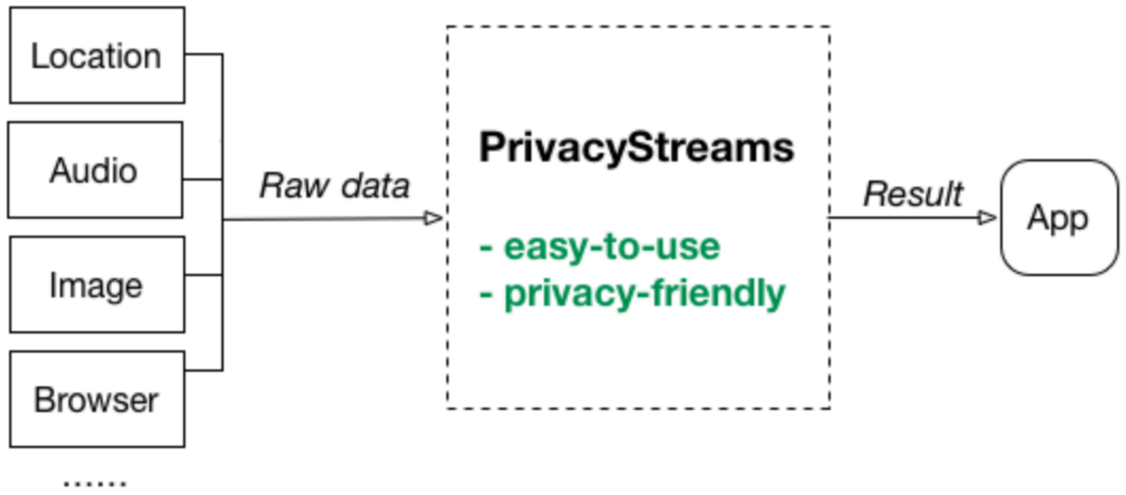
A programming language that automatically enforce privacy policies



Yang, Jean, Kuat Yessenov, and Armando Solar-Lezama. "A language for automatically enforcing privacy policies." ACM SIGPLAN Notices 47.1 (2012): 85-96.

Programming models/frameworks

Privacy-preserving APIs: Balancing Privacy and Utility




```
 uqi.getData(Audio.recordPeriodic(10*1000,
2*60*1000), Purpose.HEALTH("monitoring sleep")) //
Record a 10-second audio periodically with a 2-
minute interval between each two records.
    .setField("loudness",
AudioOperators.calcLoudness("audio_data")) // Set a
customized field "loudness" for each record as the
audio loudness
    .onChange("loudness", callback) // Callback with
loudness value when "loudness" changes
```

Li, Yuanchun, et al. "Privacystreams: Enabling transparency in personal data processing for mobile apps." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1.3 (2017): 1-26.

Programming models/frameworks

Privacy-preserving APIs: Balancing Privacy and Utility

Privacy-preserving APIs

In order to support core advertising use cases without reliance on cross-app identifiers, the Privacy Sandbox on Android proposes a set of APIs that enable ads personalization and measurement in a more private way.

These APIs protect user privacy through a combination of techniques such as retaining selected private data and processing on-device, aggregation and randomizing of data, and on-device ad selection. These API designs align closely with the corresponding efforts by the Privacy Sandbox for the Web to ensure consistency in the approach and the desired outcome, while taking into account the differences in browser and app technologies.

The initial design proposals include 3 core use cases:

- Topics infers coarse-grained interest signals, called *topics*, based on the apps on a user's device. Advertising SDKs may use these topics as an input to serve ads to relevant users.
- Protected Audience introduces a new way to show ads based on "custom audiences" defined by app developers and the interactions within their app. The solution stores this information and associated ads locally, and provides a framework to orchestrate ad selection workflows.
- Attribution Reporting supports the measurement of conversions, machine learning optimization use cases like predicted conversion-rate model building, and invalid activity detection.

Source: https://developers.google.com/privacy-sandbox/overview/android

Programming models/frameworks

Privacy annotations

@DataAccess

specifies accessed data types

```
@DataAccess(
  id = photo_attachment,
  dataType = {
    PhotosAndVideos_Photos})
Intent intent;
```

@DataTransmission

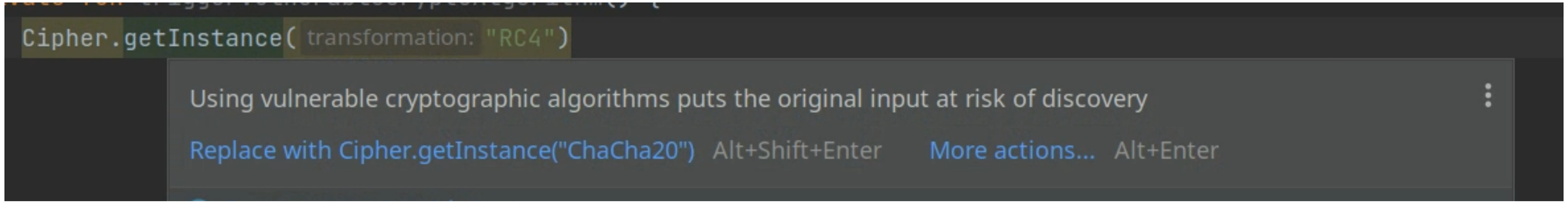
specifies how and why data is transmitted out of the app

```
@DataTransmission(
  accessId = {photo_attachment},
  collectionAttribute = {
    TransmittedOffDevice.True,
    OptionalCollection.False…},
  sharingAttribute = {
    SharedWithThirdParty.False})
NewUserInDbModel newUser;
```

By asking developers to provide one set of privacy annotations, multiple privacy-related tasks can be handled automatically or semi-automatically

Developer Tools — IDE Plugins

Built-in security lint checks in Android Studio



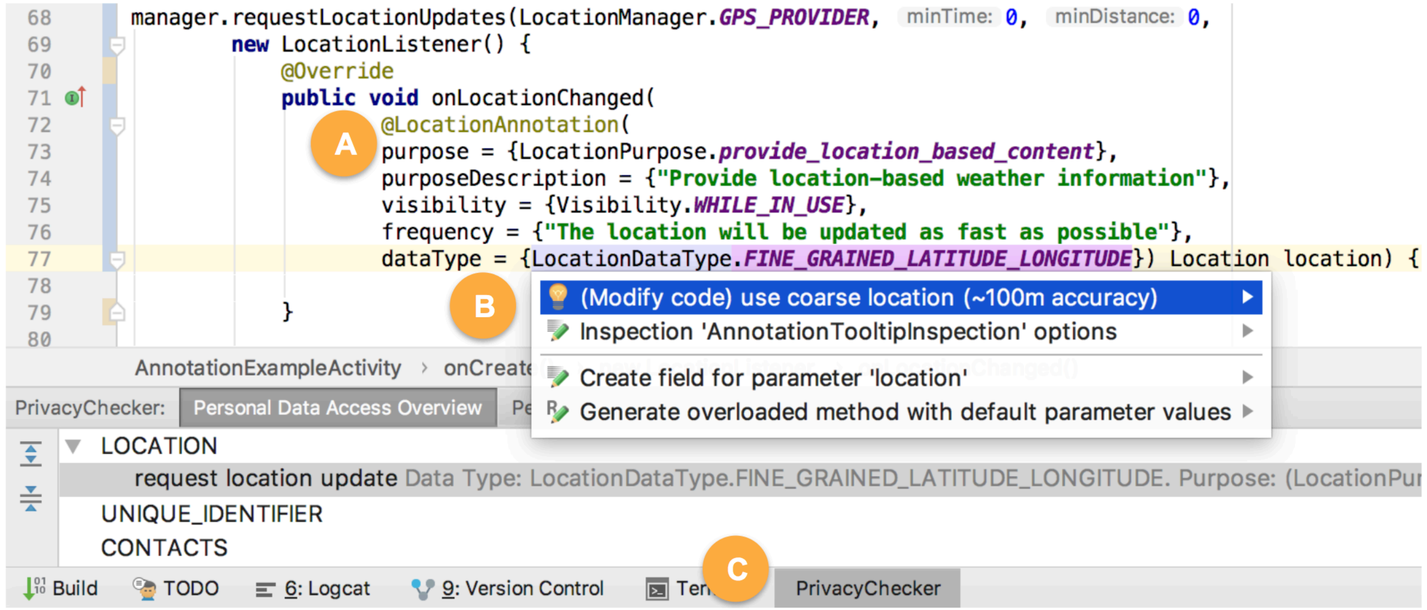
Source: https://android-developers.googleblog.com/2024/07/making-security-easy-vuln-remediation.html

Developer Tools — IDE Plugins

Built-in security lint checks in Android Studio

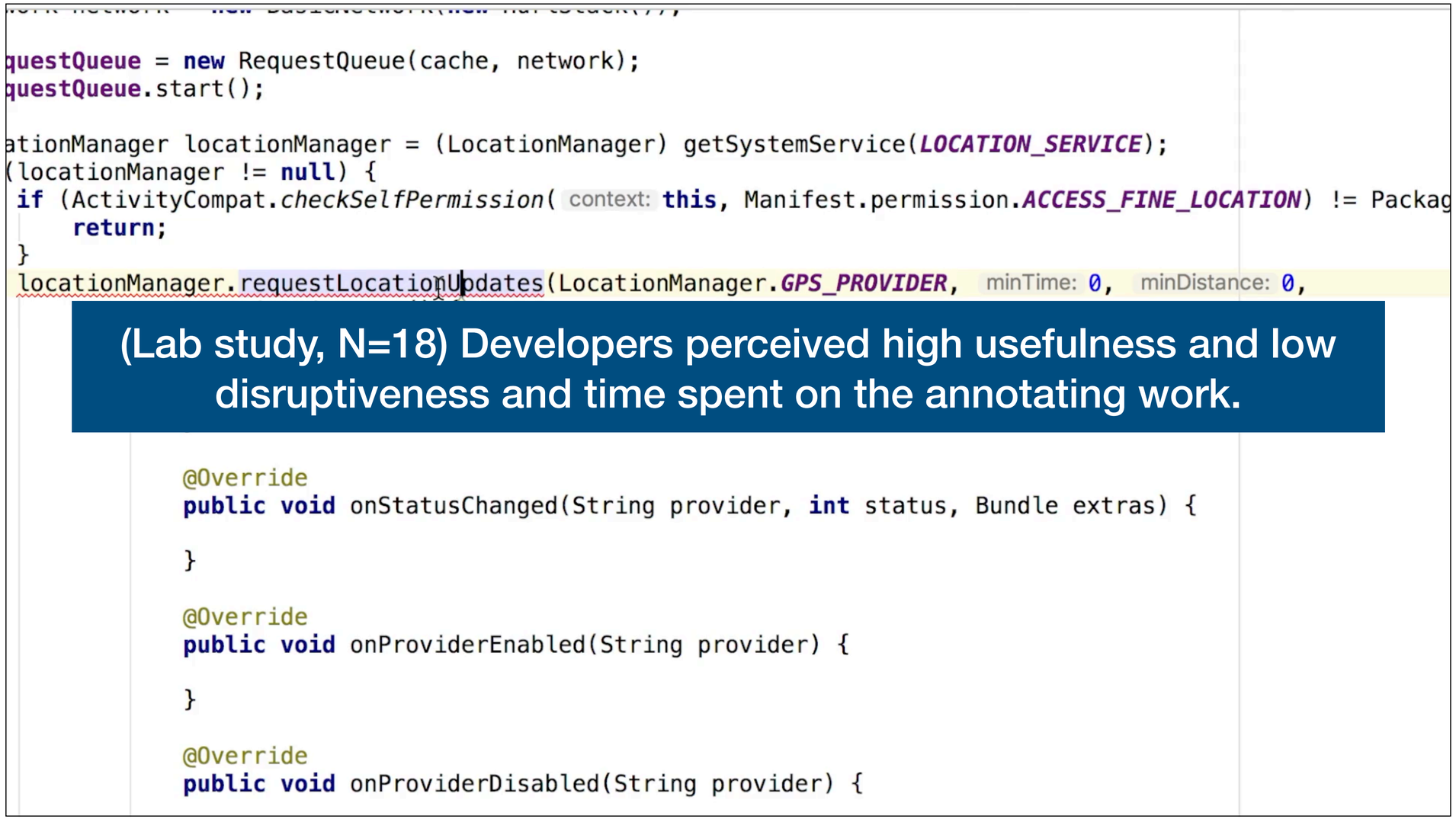
How can we design similar just-in-time checks for privacy?

Coconut: An IDE plugin for Privacy (IMWUT 2018)



Li, T., Agarwal, Y., & Hong, J. I. “Coconut: An IDE plugin for developing privacy-friendly apps.” IMWUT 2018

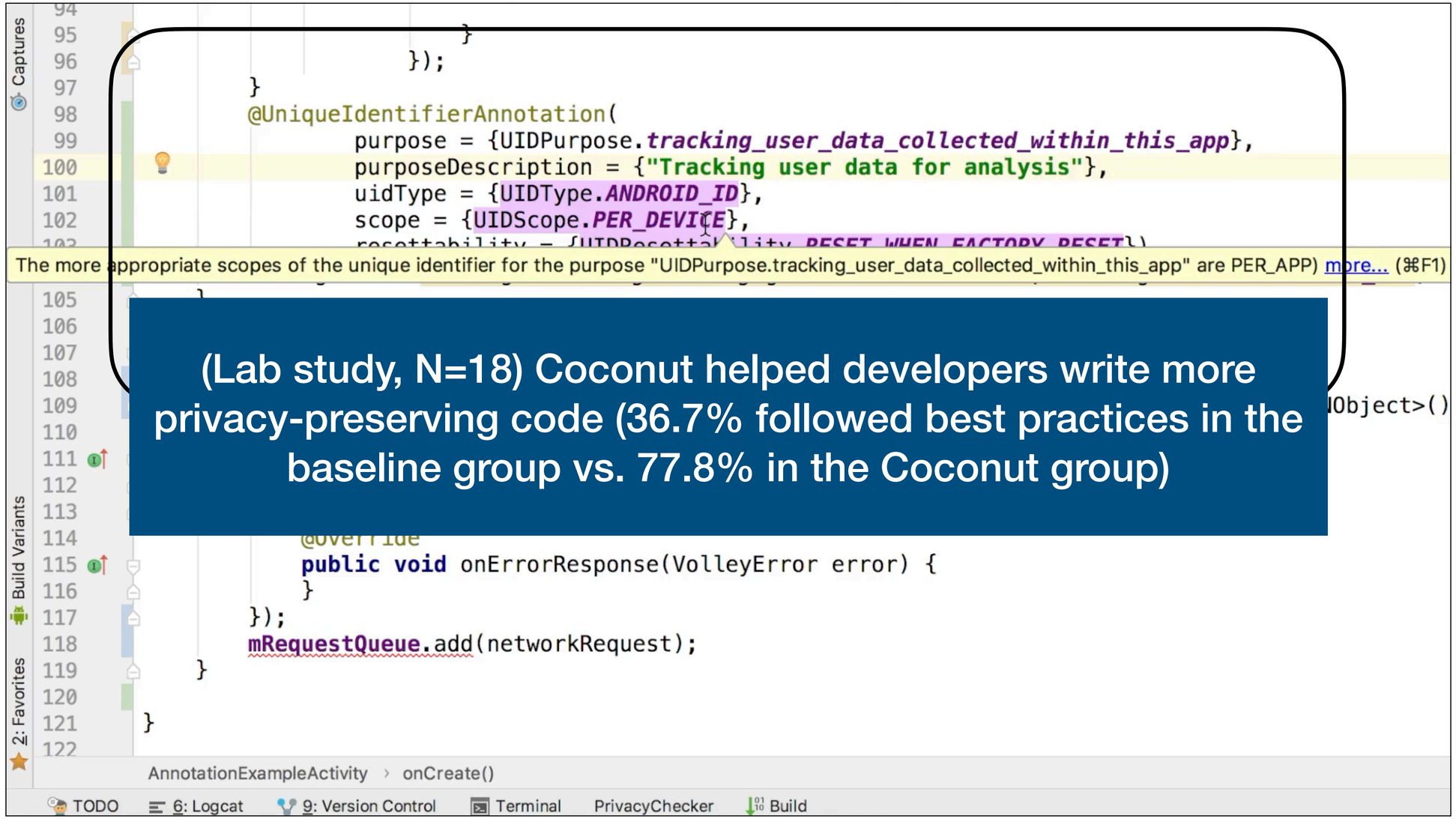
To tackle partial understanding of privacy: Coconut helps developers add privacy annotations




```
questQueue = new RequestQueue(cache, network);
questQueue.start();

ationManager locationManager = (LocationManager) getSystemService(LOCATION_SERVICE);
(locationManager != null) {
 if (ActivityCompat.checkSelfPermission( context: this, Manifest.permission.ACCESS_FINE_LOCATION) != Packag
     return;
 }
 locationManager.requestLocationUpdates(LocationManager.GPS_PROVIDER,  minTime: 0,  minDistance: 0,

    @Override
    public void onStatusChanged(String provider, int status, Bundle extras) {

    }

    @Override
    public void onProviderEnabled(String provider) {

    }

    @Override
    public void onProviderDisabled(String provider) {
```

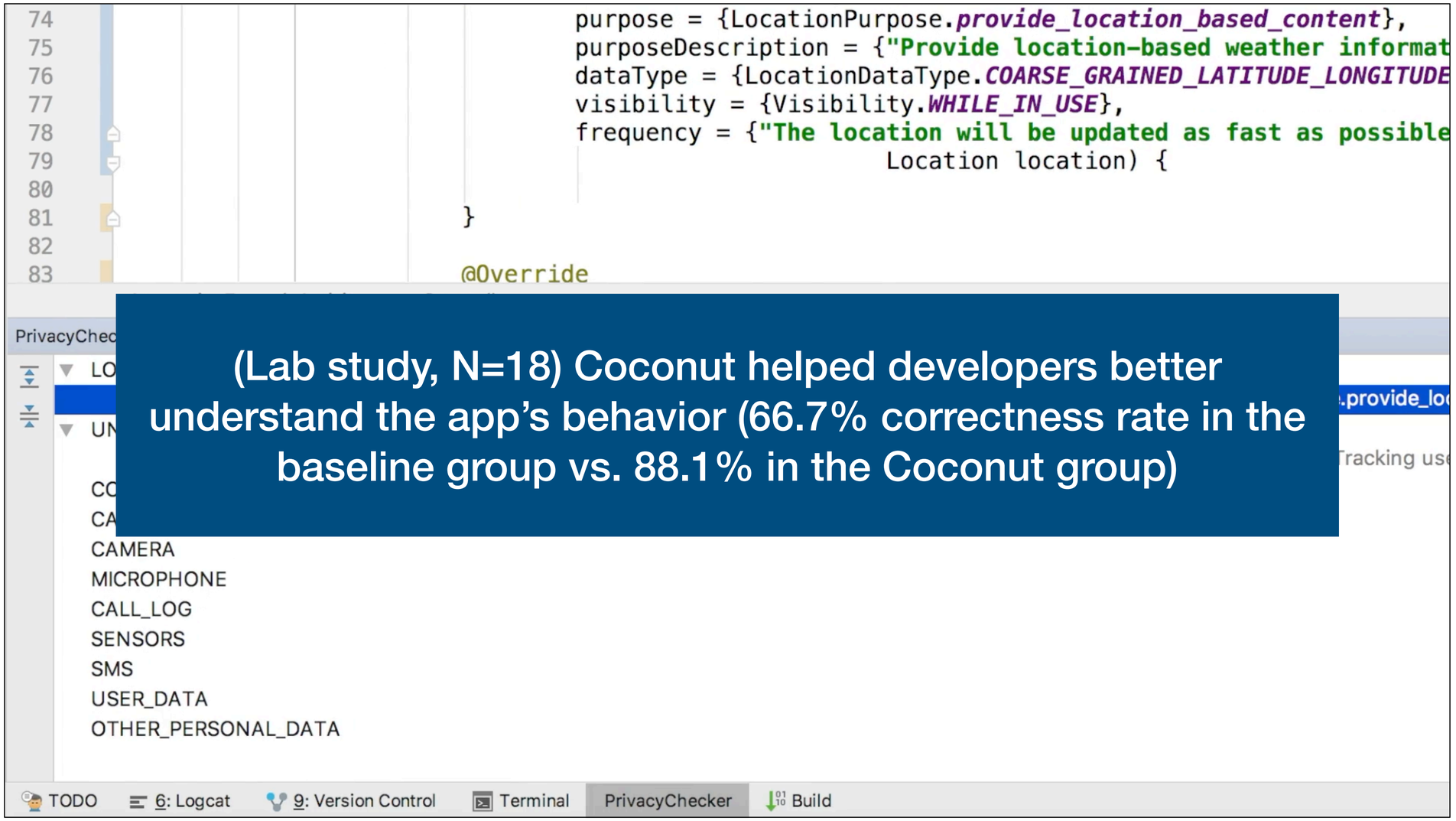
Li, T., Agarwal, Y., & Hong, J. I. “Coconut: An IDE plugin for developing privacy-friendly apps.” IMWUT 2018

To tackle lacking privacy knowledge: Coconut's *Privacy Lint* reminds developers about best practices



Li, T., Agarwal, Y., & Hong, J. I. "Coconut: An IDE plugin for developing privacy-friendly apps." IMWUT 2018

To tackle incorrect understanding of app behavior: Coconut's annotation-based privacy overview panel



Li, T., Agarwal, Y., & Hong, J. I. "Coconut: An IDE plugin for developing privacy-friendly apps." IMWUT 2018

@DataAccess

specifies accessed data types

```
@DataAccess(
   id = photo_attachment,
   dataType = {
     PhotosAndVideos_Photos})
Intent intent;
```

@DataTransmission

specifies how and why data is transmitted out of the app

```
@DataTransmission(
   accessId = {photo_attachment},
   collectionAttribute = {
     TransmittedOffDevice.True,
     OptionalCollection.False…},
   sharingAttribute = {
     SharedWithThirdParty.False})
NewUserInDbModel newUser;
```

Support

Coconut

Increase understanding of privacy

Honeysuckle

Streamline privacy feature implementation

Matcha

Increase accuracy of privacy notices

Nudges

What do my competitors do?

- Developers are incentivized by an automated alert, or “nudge”, shown in the Google Play Console when their apps ask for permissions that are requested by very few functionally-similar apps—in other words, by their competition.

Your app is requesting the permission, <permission _name>, which is used by less than X% of functionally similar apps. *<number> functionally similar apps which initially requested <permission_name> have stopped requesting it.* *Users prefer apps that request fewer permissions and requesting unnecessary permissions can affect your app’s visibility on Google Play. If these permissions aren’t necessary, you may be able to use alternative methods in your app and request fewer permissions. If they are, we recommend providing an explanation to users of why you need the permissions. Learn more.*

Table 1: Privacy warning shown to developers

Peddinti, Sai Teja, et al. "Reducing permission requests in mobile apps." Proceedings of the internet measurement conference. 2019.

Nudges

What do my competitors do?

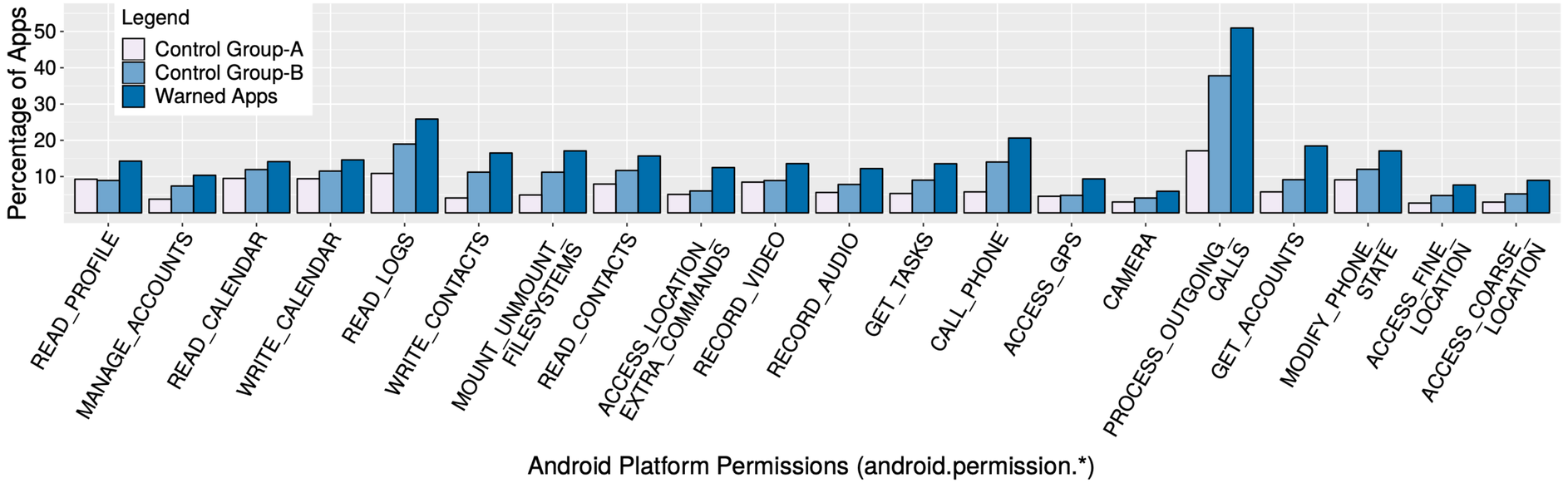


Figure 6: Percentage of apps that removed permissions (warned apps and control groups; 20 permissions shown)

Permissions have been redacted by 59% of apps that were warned, and this attenuation has occurred broadly across both app categories and app popularity levels.

Peddinti, Sai Teja, et al. "Reducing permission requests in mobile apps." Proceedings of the internet measurement conference. 2019.

Nudges

What do my competitors do?

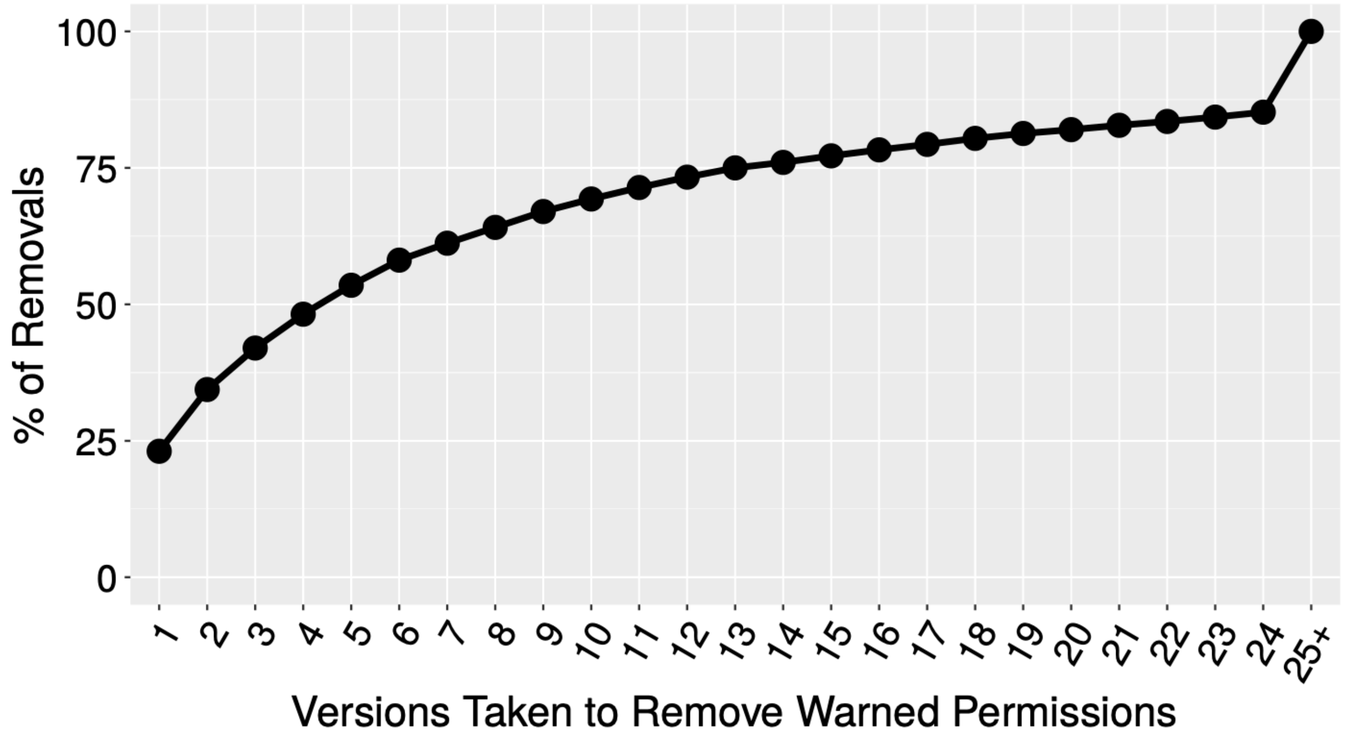


Figure 5: Versions taken to remove permissions

Peddinti, Sai Teja, et al. "Reducing permission requests in mobile apps." Proceedings of the internet measurement conference. 2019.

- How can we provide support for cross-platform development?
- How can we provide more proactive support?
- How can we incentivize the adoption of these developer tools for privacy?
  - Reliance on big tech adoption
- Are there any other ideas that you think are worth exploring?

Announcements

- Today is the final Lecture on a research topic
- The next two classes will be dedicated to paper discussions
- The project final presentation is on Dec 2
- The last lecture is on Dec 4
- The project final report is due on Dec 9