Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers
Sheffield, UK / 30 Oct - 2 Nov 2024

AUDITING THE CLOSED iOS ECOSYSTEM: IS THERE POTENTIAL FOR LARGE LANGUAGE MODEL APP INSPECTION

Jennifer Pybus
York University

Kristian Sick Svendsen
University of Copenhagen

Signe Sophus Lai
University of Copenhagen

Stine Lomborg
University of Copenhagen

Scholarly attention is increasingly paid to the dynamic and embedded ways in which third party tracking has become widespread in the mobile ecosystem. Much of this research has focused almost exclusively on Android and Google Play applications and their respective infrastructures of data capture, even though Apple and their App Store have a significant market share, with nearly 30 percent in Europe and 54 percent in the US (Sellers, 2024; Statista, 2023). One of the main reasons for this scholarly imbalance has to do with key differences between these two ecosystems, namely that Google's is open and accessible while Apple's is closed-off and encrypted, making it challenging, if not illegal to adequately inspect its apps. Moreover, few open access tools or databases like Exodus Privacy for Android apps, are available to make it easier for interested interlocutors or academics to inspect its applications. How then to ensure any kind of regulatory accountability if Apple's infrastructure lacks observability, outside the promises made about its new privacy preserving operating system?

We see an active need to better account for the ways in which this ecosystem accesses, shares and uses end-user information, increasing capacity for greater external oversight to ensure compliance with data protection legislation. As such, platform governance debates, particularly those that extend to the infrastructures of data capture within our applications should not only apply to Google. Our paper therefore seeks to address this gap through a methodological exploration that uses

Pybus, J.; Svendsen, K. S.; Lai, S.S.; Lomborg, S. (2024, October). *Auditing the Closed iOS Ecosystem: Is there Potential for Large Language Model App Inspections?*. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

ChatGPT-4's to decompile and open-up the IPA files of +1000 apps. On this basis, we assess its usability as an adequate approach to examine two key mobile applications infrastructures: i) the Plist file, which details how end-user data is accessed and used by the app and ii) the software developments kits (SDKs) or third-party platforms and companies that developers use to both build and develop apps.

Within app studies, the infrastructure that facilitates access to end-user data stems from the multi-situated (Dieter et al., 2019) and dynamic ways in which developers embed software from both platforms and third parties to build and monetize their applications (Blanke and Pybus, 2020). One of the primary ways in which these services are integrated is through SDKs. These third parties constitute practical technical objects that offer a range of complex services such as ready-made tools, code, libraries, and tutorials from either platforms and/or third parties. In addition, SDKs provide and encase APIs so that their services can be easily accessed and integrated into apps (Gontovinkas, 2020). For this reason, they take on a modular quality, wherein developers are free to pick and choose what they want to use (Pybus and Coté, 2024). However, given that these infrastructures belong almost uniquely to platforms, their opaque encased architectures raise a number of privacy and accountability questions, especially regarding what data gets accessed and shared, and with whom.

Scholarship on the relationship between apps and third parties tend to position the third parties as fundamental agents of surveillance (Zuboff, 2019), data (West, 2019), or platform (Srnicek, 2017) capitalism, playing a pivotal role in the monetization of end-user data. Datafication scholarship works alongside these insights and draws attention to the infrastructural agents that facilitate complex end-user data extraction for new kinds of value from a range of different actors (Flensburg and Lomborg, 2023; Blanke and Pybus, 2020; Flensburg and Lai, 2022). In this way, it draws our attention to the non-rivalrous quality of data, which enables its continuous (un)coupling, (re)use, and seamless integration into a constellation of different arrangements to maximise its value and exchange. We therefore see the infrastructures of apps, both the SDKs and the data governance policies - manifest files in Android or Plist files in Apple - as fundamental objects of study, given the way they amplify datafication, serving to (re)produce a data-for-service logic that underscores the app economy.

To examine Apple's ecosystem, we have first drawn on the open-source tools developed by Kollnig and his team (2021) to access IPA files and then to overcome the challenge of opening and inspecting, we have explored the role that a large language model (LLM) can play in enabling this process? We have opted to use Chat GPT-4 because of how it has been designed to assist humans in both interacting and understanding code. It has therefore been programmed to both simplify complexity and be able to provide documentation and commentary on existing codebases. We therefore ask: Can Chat-GPT assist scholars interested in Apple's mobile ecosystem through the auditing of IPA files?

To answer this question, we compare findings from the automated analysis with manually crafted dataset of SDKs derived from an ERC-funded project, in which we have mapped infrastructures of tracking and mobile datafication across Android and iOS platforms to gauge the infrastructural dependence on tech companies as peoples'

ordinary lives are increasingly orchestrated with the phone (Anonymized for review). As part of the project, we have collected a dataset of 1000+ apps, and inspected these for embedded permissions for accessing data via the app as well as identified the SDKs of each app.

The outcomes of the explorative study include 1) an evaluation of Chat-GPT4 as an assistant for reading code at scale as well as its potential for automating processes involved in future app monitoring and regulation; 2) a deep dive into the shortcomings of Chat-GPT4 when it comes to interventions into the mobile ecosystems; and 3) a discussion of ethical issues involved in harnessing LLM for forwarding a research agenda around Apple's otherwise understudied ecosystem. Our exploration aims to evaluate a methodological intervention that could bring more observability into an otherwise closed-ecosystem that promises the preservation of end-user privacy, without any meaningful external oversight.

References

Blanke, Tobias, and Jennifer Pybus. 'The Material Conditions of Platforms - Monopolisation through Decentralisation'. *Social Media + Society*, 2020. https://doi.org/10.1177/2056305120971632.

Dieter, Michael, Carolin Gerlitz, Anne Helmond, Nathaniel Tkacz, Fernando N. van der Vlist, and Esther Weltevrede. 'Multi-Situated App Studies: Methods and Propositions'. *Social Media + Society* 5, no. 2 (1 April 2019): 2056305119846486. https://doi.org/10.1177/2056305119846486.

Flensburg, Sofie, and Signe S. Lai. 'Datafied Mobile Markets: Measuring Control over Apps, Data Accesses, and Third-Party Services'. *Mobile Media & Communication* 10, no. 1 (1 January 2022): 136–55. https://doi.org/10.1177/20501579211039066.

Flensburg, Sofie, and Signe Sophus Lai. 'Networks of Power. Analysing the Evolution of the Danish Internet Infrastructure'. *Internet Histories* 5, no. 2 (2020): 79–100. https://doi.org/10.1080/24701475.2020.1759010.

Gontovnikas, Martin. 'What Is an SDK?' Auth0, 23 December 2020. https://auth0.com/blog/what-is-an-sdk/.

Kollnig, Konrad, Anastasia Shuba, Reuben Binns, Max Van Kleek, Nigel Shadbolt, and : De-. 'Are iPhones Really Better for Privacy? Comparative Study of iOS and Android Apps'. *Proceedings on Privacy Enhancing Technologies ..; .. (*, September 2021, 1–19. https://arxiv.org/abs/2109.13722v4.

Pybus, Jennifer, and Mark Coté. 'Super SDKs: Tracking Personal Data and Platform Monopolies in the Mobile'. *Big Data & Society* 11, no. 1 (1 March 2024): 20539517241231270. https://doi.org/10.1177/20539517241231270.

Sellers, Dennis. 'Apple Registered a Record-High Q4 Market Share in Europe's Smartphone Market'. *Apple World Today*, 26 February 2024.

https://appleworld.today/apple-registered-a-record-high-q4-market-share-in-europes-smartphone-market/.

Srnicek, Nick. *Platform Capitalism*. John Wiley & Sons, 2017. https://www.wiley.com/en-gb/Platform+Capitalism-p-9781509504862.

Statista. 'US Mobile Smartphone OS Market Share 2012-2023', 2023. https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/.

West, Sarah Myers. 'Data Capitalism: Redefining the Logics of Surveillance and Privacy'. *Business & Society* 58, no. 1 (2019): 20–41.
Zuboff, Shoshana. *The Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power*. London: Profile Books Ltd, 2019.