International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 12-No.1 (2026) pp. 840-846
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Optimizing Time-to-First-Ad in Mobile Advertising Systems

Rishi Kanth Alapati*

University of Southern California, Los Angeles, USA
* **Corresponding Author Email:** alapatirishikanth@gmail.com - **ORCID:** 0000-0002-0047-7660

Article Info:

DOI: 10.22399/ijcesen.4924
Received : 15 December 2025
Revised : 25 January 2026
Accepted : 02 February 2026

Keywords

Ad Latency Optimization,
Progressive Delivery,
Edge Computing,
Predictive Prefetching,
Performance Observability

Abstract:

The article on Time-to-First-Ad (TTFA) optimization in mobile advertising platforms emphasizes how the speed of delivery of an advertisement is vital to the success of a business in the competitive online market. The article discusses the technical strategies that are multifaceted and are used by the current platforms to ensure that latency is minimized when a cold start situation is experienced, where system resources are the most limited. The article provides a comprehensive discussion of the progressive delivery mechanism, predictive prefetching techniques, edge computing systems, and systematic latency budgeting habits through the analysis of which the latest generations of leading platforms have remodeled their delivery architecture to attain significant performance gains. The discourse is not restricted to the technical implementations, but also includes the organizational aspects of the matter, such as dedicated performance engineering teams and advanced observability systems that allow ongoing optimization. The article shows a positive relationship between TTFA improvements in user engagement metrics, more impression volume, and higher revenue generation capabilities. With the advertising technology industry still moving forward, the article puts latency optimization as a strategic necessity as opposed to a technical one, and the effect of its application cuts across the entire advertising ecosystem, between publishers and final consumers.

1. Introduction

In the mobile ad competitive environment, the delivery and rendering speed of ads determines both revenue and user experience directly. One key metric by which advertising platforms measure is "time-to-first-ad" (TTFA), or the latency from app launch until the first ad is shown to users. This piece discusses technical problems and solutions for reducing TTFA, especially in cold app starts when system resources are tightest.

The ad loading speed vs. revenue generation is one of the most important mobile ad optimization challenges. Recent server-side ad insertion framework research shows that even slight gains in delivery time can have disproportionate effects on revenue at large-scale systems. The dynamic nature of mobile networks exacerbates this challenge, especially on first-time app launches, where several competing processes compete for scarce device and network resources. Researchers who have studied adaptive video streaming platforms have reported significant user engagement disparities between optimized and non-optimized advertisement delivery systems, with advertisements loading quicker having better completion rates and subsequent interaction metrics [1].

Cold start situations pose specific challenges to advertising systems as a result of no cached assets and warm connection pools. In these situations, ad delivery systems have to juggle conflicting requirements: delivering valuable ads quickly while remaining mindful of device resource limitations that may be busy processing application launch at the same time. The computer architecture of modern ad delivery systems is actually designed precisely to manage these difficult situations, with smart prefetching technologies and progressive delivery strategies becoming increasingly prevalent throughout the industry. Such strategies acknowledge that even minimal reductions in initial ad presentation times are associated with quantifiable improvements in campaign performance metrics, as evidenced in many server-side ad insertion studies [1].

Contemporary mobile ad platforms utilize intelligent multi-layered architectures that are specifically designed to reduce TTFA across various network conditions. Edge computing deployments have become even more at the center of latency-cutting strategies, with large advertising platforms now running distributed infrastructure that places key rendering assets near end users. Industry experts looking at publisher-side ad technology use have reported that edge-optimized delivery infrastructures normally outperform centralized systems by quite high margins, especially for mobile users on variable-quality networks [2]. These innovations in architecture, along with progress in predictive technologies and real-time decision systems, have revolutionized the mobile advertising business to the point where platforms can deliver continually lower TTFA metrics even in poor network environments.

2. The Business Impact of Ad Latency

As consumers launch mobile apps, any lag in content rendering can have adverse impacts on engagement metrics. In ad-supported apps, this also means directly affecting revenue. Industry research across the ad industry illustrates that ad rendering speed improvement is related to increased user engagement rates and more monetization opportunities. Platforms that have optimized their TTFA successfully have seen corresponding revenue boosts relative to the amount of latency savings realized.

The economic effects of ad latency go far deeper than mere user experience factors, with thorough industry studies showing complex economic effects throughout the ad world. Publishers deploying sophisticated latency optimization strategies have shown dramatic gains in critical monetization metrics such as fill rates, viewability scores, and effective CPM values. Industry studies by advertising technology companies show that viewability—a factor directly affected by load times—is of central importance to publisher economics. When ads are loaded fast enough to be viewed correctly, publishers see real revenue gains, with some estimating that jumping from average to high viewability rates can boost earnings by double-digit percentages [3]. These differences in performance are particularly important for mobile-first publishers, where network uncertainty and device resource limitations add complexity to the ad serving process.

In addition to direct revenue impacts, ad latency also impacts key long-term business metrics like user retention and frequency of sessions. Long-term studies looking at mobile app usage patterns have determined strong correlations between performance metrics and subsequent user engagement. Studies using survival analysis techniques applied to app usage data have shown that technical performance attributes have a strong impact on how long users remain active with applications over time [4]. Applications that provide smooth, responsive experiences—efficient ad loading included—illustrate measurably higher retention curves than those with poor performance. This retention benefit accumulates over time, generating significant lifetime value differences that have major overall business sustainability implications for ad-supported applications.

The competitive dynamics of contemporary advertising marketplaces also further magnify the business significance of latency optimization. As programmatic platforms increasingly use performance signals in their auction and allocation calculations, publishers possessing better technical infrastructure enjoy significant benefits in impression monetization. Supply-side platforms now make it an automatic practice to include historical performance metrics within quality score mechanisms, essentially generating price premiums for high-performing inventory sources. Viewability rates, highly correlated with latency performance, have emerged as important drivers of how advertisers value impression opportunities [3]. These structural benefits generate strong incentives for ongoing investment in latency-optimization technologies across the publishing supply chain.

For advertising-reliant business models, the alignment between technical efficiency and financial performance establishes a direct strategic imperative around latency management. The most advanced publishers now treat TTFA optimization as an overarching business function and not a technical issue per se, with specific resources devoted exclusively to tracking and optimizing this key metric. This organizational priority is a response to increased awareness that even slight gains in ad delivery speed can produce dramatic returns when applied repeatedly across large impression volumes. Studies of user retention habits indicate that technical performance problems, such as slow-loading content and advertising, are among the most critical abandonment decision factors [4]. As the competitive environment keeps changing, those publishers that do not put high emphasis on latency optimization will increasingly find themselves at more serious disadvantages both in terms of user experience quality and revenue generation ability.

3. Structure of Modern Ad Delivery Systems

To manage the TTFA challenge, modern ad systems utilize a multi-dimensional technical strategy:

3.1 Response Streaming and Progressive Delivery

Instead of waiting for final ad candidate selection prior to transmitting any response, highly evolved ad platforms use response streaming capabilities. This enables the client application to get high-priority ad candidates instantaneously, while more, possibly less relevant choices keep processing in the background.

The answer is usually organized in layers of priority: high-relevance-probability (HP1) initial candidates are served first, and secondary and tertiary ones based on more intricate targeting logic.

Ad delivery architectures in the contemporary era have come a long way in adapting to the inherent trade-off between quality and speed in ad selection. Top platforms now employ high-level progressive delivery patterns that inherently redefine the classical request-response loop. These systems utilize similar principles to progressive software deployment, where modifications are implemented incrementally with close monitoring of system health indicators. For advertising, this translates into presenting content incrementally instead of holding out for full processing. Progressive delivery in ad systems allows sites to reconcile the dual goals of speed and targeting precision such that initial high-quality candidates can be delivered to users rapidly while subsequent evaluation is still ongoing in background streams [5]. The technique allows initial ad candidates to be delivered to client apps in milliseconds of the initial request, generating the illusion of near-instant responsiveness even as more sophisticated candidate evaluation proceeds in concurrent processing streams.

The implementation of such streaming delivery architectures technically usually involves dedicated buffer management systems that provide ongoing connections between server infrastructure and client applications. These ongoing connections facilitate incremental transmission of ad candidates upon completion of the evaluation pipeline, thus eliminating the request-response bottleneck inherent in traditional ad delivery systems. Industry deployments utilizing these methods can realize drastic latency reductions, with staged delivery methods making it possible for end-users to start engaging with advertising content as optimization operations remain operating in background threads.

3.2 Predictive Prefetching

Intelligent ad platforms study user behavior patterns to forecast likely ad requests in advance of when they happen. Prefetching probable ad contenders and caching them at the edge minimizes delivery latency by orders of magnitude when actual requests do arrive. This method works best for return visitors whose app behavior has been learned.

Intelligent predictive prefetching systems run on multiple timescales to maximize ad delivery performance. Long-term pattern analysis looks at past user behavior over multiple sessions to determine frequent sequences of interaction that are associated with certain advertisement needs. Short-term prediction models update in real time based on activity in a session, further evolving prefetch priorities as the user moves through application content. The more advanced implementations also include contextual information like time of day, device attributes, and network quality to add yet more accuracy to predictions. Carefully crafted prediction systems can keep backend request latency low for repeat users, making the system feel faster while also offloading infrastructure load.

The technical complexity of the prediction systems is advancing at a quick pace, with the latest implementations using machine learning-based techniques that continually optimize prefetching decisions based on seen results. These systems balance the tradeoffs between prefetching accuracy and resource use dynamically, modulating their aggressiveness in response to prevailing network conditions and device capabilities. By keeping fine-grained histories of performance per user, these systems are able to vary their strategy intelligently as a function of the quality of prediction signals, being more conservative for users with highly dynamic behavior and more aggressive prefetching for users with stable behaviors.

3.3 Edge Caching and Distributed Architecture

By putting ad delivery infrastructure near end users with edge computing principles, platforms reduce network latency. With this distributed architecture, there are smaller round-trip times for ad requests, less bandwidth usage, and better tolerance of network volatility.

Advances in edge computing capabilities have revolutionized advertising delivery architectures to bring unprecedented latency improvements across many different network environments. Studies of edge computing deployments illustrate that local processing distribution to end users can significantly lower data-intensive application response times. Research into real-time data processing systems has indicated that architectures utilizing the edge can cut average latency by 30-40% against cloud-centric methods, with considerably greater enhancements in times of high traffic [6]. These performance benefits are specifically important in advertising systems, where tens of milliseconds determine between successfully serving and lost impression opportunities.

In addition to straightforward content caching, edge deployments today also support advanced request handling features that enable early ad selection decisions to be made directly at edge sites. These distributed decision systems run on locally cached targeting information and reduced evaluation heuristics, making it possible for them to respond to a large proportion of the incoming requests without communicating with centralized infrastructure. This method also decreases perceived latency by removing round-trip communication in the case of simple ad selection scenarios. Edge computing studies have shown that this architectural paradigm provides significant advantages to applications that need real-time processing of data, with edge-based deployments exhibiting better reliability during network congestion phenomena and more uniform performance among geographically dispersed populations of users [6].

4. Latency Budget Allocation

Rigorous latency budgets are imposed by engineering teams on the ad delivery pipeline. An average breakdown permits approximately 20ms to the ranking engine, 15ms to the retrieval of creative assets, 10ms to the validation and safety checks, and 5ms to fallbacks.

This budgeting methodology ensures that there is no undue influence of a single constituent of the system on the general performance. Teams apply continuous monitoring systems where alarms are set to notify whenever any of the parts exceeds its allocation, and, therefore, performance bottlenecks are easily spotted and repaired.

The adoption of strict latency budgeting is a fundamental shift in advertising technology management, echoing methods developed in high-performance computing applications. Current ad platforms now utilize advanced decomposition methods to define accurate performance targets for every system element, with budgets usually specified at the level of milliseconds. Such detailed provisioning mirrors the essential importance of managing time in ad delivery, with even slight latencies potentially drastically influencing monetization opportunities. A study of the economic effects of latency in digital systems shows that delays in performance are directly converted into business expenses, with each millisecond possibly affecting user engagement and conversion rates [7]. For ad platforms, this fact has fueled ever-more-sophisticated methods for managing performance, with current architectures including detailed budgeting models that set well-defined expectations for every system element.

The allocation of latency to pipeline elements is determined both by technical limitations and business priorities. Operations ranking generally receive the highest allocation (around 20ms) because of their computational overhead and immediate effect on ad relevance and value. Creative asset retrieval receives the second-largest allocation (around 15ms) because of the variable nature of network operations and the essential role visual assets play in the end-user experience. Validation and safety verification operations are given moderate budgets (around 10ms), trading off between needing good verification and performance limitations. Fallback routines and error handling paths are given the lowest budgets (around 5ms), since these parts of the system need to run quickly when primary systems are delayed or fail. This tiered budgeting strategy provides the most important functions with adequate resources while keeping tight reins on system performance as a whole.

Successful use of latency budgets calls for advanced instrumentation and monitoring features that give insight into component-level performance metrics. Successful ad platforms currently employ distributed tracing systems that monitor request processing through all stages of the pipeline, allowing accurate measurement of execution time for every system component. These monitoring tools often include advanced visualization software that emphasizes budget violations and trends in performance, allowing operations teams to detect issues early before they affect system performance as a whole. Observability practice today strongly asserts the need for end-to-end monitoring in distributed systems, with successful implementations blending metrics, traces, and logs to ensure total visibility into system behavior [8]. To advertising platforms running on a worldwide scale, this enhanced observability immediately maps to better revenue protection and stability.

Beyond reactive monitoring, newer latency management systems now also include predictive features that foresee imminent performance degradation before it ever gets to the point of actual budget breaches. Such systems cross-analyze historical performance trends and existing load profiles to predict future constraints, allowing ahead-of-time resource allocation or traffic diversion to avoid latency spikes. Advanced event processing engines constantly compare real-time statistics to predefined baselines, initiating automated remediation processes when statistical outliers indicate future performance problems. Good observability systems enable teams to transcend mere monitoring of discrete parts to comprehension of sophisticated interactions among distributed services, allowing them to construct mental models of system behavior that underpin both debugging and performance improvement [8]. For ad platforms, where performance directly translates to revenue, these end-to-end observability features bring important operational and financial benefits.

5. Observability and Performance Management

Having in place an optimal TTFA necessitates strong monitoring capabilities, such as per-request telemetry taking timing at every pipeline stage, canary deployment to catch performance regressions before full deployment, A/B testing infrastructure to check optimization improvement, and anomaly detection systems to pick up unusual latency trends.

The sophistication of contemporary ad delivery infrastructure requires high-end observability architectures that go far beyond rudimentary monitoring strategies. Top platforms establish end-to-end telemetry systems that monitor hundreds of unique performance indicators across all key request flows, generating rich data sets that both inform operational monitoring and enable ongoing optimization activities. These observability implementations record detailed timing information for every element of each pipeline, with high-resolution timestamping enabling microsecond-level accuracy for key processing phases. The resulting telemetry streams support real-time performance dashboards that graph system behavior in multiple dimensions, such as geographic distribution, device attributes, network characteristics, and campaign attributes. Distributed tracing deployments in contemporary systems support rich capabilities for insight into request flow in complex service architectures so that teams can identify the actual components of latency issues instead of just knowing there are issues [9]. For global-scale advertising platforms, this enhanced diagnostic functionality directly translates into increased reliability and revenue maintenance.

Sophisticated deployment methods are critical to maintaining TTFA performance as the system evolves. Canary deployment regimes support the gradual rollout of new code through routing a small percentage of production traffic to refreshed instances and monitoring the critical performance indicators. This methodology enables engineering teams to identify potential regressions with minimal user effect, stopping deployments automatically when measured performance strays outside predefined baselines. Advanced canary implementations feature automated statistical analysis that considers multiple dimensions of performance at once, spotting nuanced degradations that might elude human inspection. These deployment protections work alongside thorough A/B testing systems that rigorously test suggested optimizations against control versions, making sure that modifications yield quantifiable performance improvements prior to widespread deployment. Controlled experimentation studies prove that well-implemented A/B testing systems offer essential functionality for confirming optimization work, allowing teams to make decisions based on facts instead of intuition or mathematical models of performance alone [10].

In addition to traditional monitoring methods, top ad platforms now employ dedicated anomaly detection software written specifically to detect emerging performance problems. The software uses sophisticated machine learning methods to create dynamic baselines of acceptable performance that consider typical traffic fluctuations, seasonal trends, and known system actions. Dedicated algorithms constantly compare incoming telemetry against changing baselines, detecting statistical outliers that should be investigated. The most advanced deployments include multi-dimensional analysis features that cross-correlate anomalies in correlated metrics so that sophisticated degradations in performance may be detected by considering signals of multiple sets in combination. Distributed tracing tools aid teams in transcending symptom treatment to insight into causes, with in-depth request visualization features allowing engineers to rapidly determine ailing service interactions and performance bottlenecks [9].

The technical implementations of performance management are of equal significance to organizational considerations. Top-tier ad platforms have specialized performance engineering teams that manage TTFA metrics throughout the technology stack. These groups combine skills in distributed systems architecture, performance analysis, and knowledge of the advertising domain so they can understand the intricate relationships between technical limitations and business needs. Performance engineers collaborate with product development teams to define proper latency budgets for new features, perform pre-release performance testing, and ensure observability instrumentation. They also enable regular performance reviews, analyzing trend data across critical metrics such that gradual degradations get the proper focus, even when individual changes do not dip below alert thresholds. Evidence on controlled experimentation culture through research suggests that companies rolling out strict testing programs make much better product decisions, with well-planned experiments yielding hard evidence about the effect of suggested changes [10].

6. Results and Industry Benchmarks

Deployments of these methods have produced stunning returns throughout the advertising tech ecosystem. Top platforms reported TTFA decreases of as much as 40% by methodically adopting the strategies detailed above. The enhancements translate directly into quantifiable improvements in chief business metrics, such as higher impression volume, more elevated session depth metrics, better revenue per session, and better user retention.The measurable effect of TTFA optimization efforts shows the essential role of latency management in advertising ecosystems. In-depth industry studies, conducting multi-advertising platform performance data analysis consistently show correlations between speed of delivery enhancements and material business results. Platforms that apply the full set of latency optimization methods often obtain TTFA reductions of 30-40% from their former baseline levels, with some implementations reporting even more extreme reductions under certain network conditions. These engineering gains directly translate to concrete business benefits, with optimized platforms realizing quantifiable advantages across several dimensions of performance. Studies that analyze performance optimization in software systems illustrate how latency gains can have a noteworthy effect on user experience metrics, with mere reductions in response time causing quantifiable increments in engagement behaviors [11]. This correlation is especially marked in ad technology, where milliseconds can distinguish between whether or not an impression opportunity is properly monetized.In addition to basic impression delivery metrics, latency optimization also has a beneficial effect on more granular engagement indicators that signal general application health. User session depth analyses show that apps using innovative TTFA optimization methods usually see an enhancement in average screens shown per session, which means accelerated ad experiences are a contributor to user satisfaction as well as continued interaction. These improvements in engagement over time drive larger cumulative effects on lifetime value metrics. Guidelines from the mobile advertising industry highlight the paramount role technical performance plays in ensuring successful campaigns, as the speed of loading ads has a direct influence on user experience and conversion rates [12]. The advertising tech industry increasingly acknowledges that optimizing technical performance is a basic element of business strategy rather than an engineering imperative.

The competitive forces of the ad marketplace itself only serve to heighten the significance of such performance benefits. As the leading advertising platforms have adopted ever-more advanced latency optimization strategies, performance levels throughout the ecosystem have grown correspondingly ambitious. Platforms that are not competitive on TTFA metrics are progressively placed at structural handicaps in both user experience quality and marketplace competitiveness. This creates strong incentives for ongoing innovation in latency management strategies, with platforms constantly seeking out new optimization methods to preserve or grow their performance leads. Literature on software performance optimization and quality suggests that companies investing in performance engineering over the long term generally gain a substantial competitive edge over time, especially in markets where the quality of the user experience has direct implications for business results [11].Gazing into the future, frontier technologies are set to continue revolutionizing TTFA optimization abilities. More sophisticated machine learning methods are being used in predictive prefetching infrastructure to better anticipate user behavior patterns and related ad demand. Edge computing infrastructure keeps growing into new geographic areas, lowering network transit times for end-users in previously unserved locations. Hardware acceleration for typical advertising workloads is starting to show up in production environments, allowing for more efficient processing of computationally expensive ranking and targeting operations. Industry counsel on mobile ad strategy increasingly underscores the need for ongoing technical refinement, knowing that performance demands will persist to adapt as user expectations and competitive forces increase [12]. As these advancements become a reality, business benefits tied to leadership in performance are likely to be even greater, further solidifying latency optimization as a key success factor for ad tech platforms.

7. Conclusions

The history of mobile advertising technology development has created time-to-first-ad optimization as an essential point of difference in the competitive digital ecosystem, with ramifications that extend to the user experience and revenue production. With advanced architectural designs such as response streaming, edge computing, predictive prefetching, and strict latency budgeting, the best platforms have shown that technical performance is directly correlated with business value. The compounding effect of these optimizations is not just on immediate measures of impression delivery, but also on more profound measures of engagement and long-term user retention trends, and these effects can be experienced at the level of publisher sustainability. With the performance metrics becoming part of the advertising marketplace auction facade and allocation mechanisms, there will only be more demand for the strategic significance of latency management, necessitating further innovation and investment. In the future, optimization potential is set to be further enhanced with emerging technologies such as more advanced machine learning and prediction, larger edge infrastructure, and purpose-built hardware acceleration; latency management will continue to be a key interest of platforms aiming at gaining a competitive edge in mobile advertising ecosystems.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] Kiran Kumar, "Optimizing Server-Side Ad Insertion to Improve User Experience in Adaptive Video Streaming," ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/publication/385885623_Optimizing_Server-Side_Ad_Insertion_to_Improve_User_Experience_in_Adaptive_Video_Streaming

[2] Gareth Glaser, "A Publisher's Guide to Latency," Gareth Hates Ad Tech, 2023. [Online]. Available: https://www.garethhatesadtech.com/p/a-publishers-guide-to-latency

[3] Rayomand Engineer, "Ad Viewability and its Impact on Publishers' Revenue," Publir, 2022. [Online]. Available: https://publir.com/blog/2022/03/ad-viewability-and-its-impact-on-publishers-revenue/

[4] Yu-Hsuan Lin et al., "Assessing User Retention of a Mobile App: Survival Analysis," ResearchGate, 2020. [Online]. Available: https://www.researchgate.net/publication/344980859_Assessing_User_Retention_of_a_Mobile_App_Survival_Analysis

[5] Octopus Deploy, "What is progressive delivery?" Octopus.com, 2023. [Online]. Available: https://octopus.com/devops/software-deployments/progressive-delivery/

[6] Brian Kelly, "The Impact of Edge Computing on Real-Time Data Processing," International Journal of Computing and Engineering 5(5):44-58, 2024. [Online]. Available: https://www.researchgate.net/publication/382156395_The_Impact_of_Edge_Computing_on_Real-Time_Data_Processing

[7] Farhan Khan, "The Cost of Latency," Digital Realty, 2024. [Online]. Available: https://www.digitalrealty.co.uk/resources/articles/the-cost-of-latency

[8] Cindy Sridharan, "Distributed Systems Observability," O'Reilly Media, 2018. [Online]. Available: https://www.oreilly.com/library/view/distributed-systems-observability/9781492033431/

[9] Austin Parker, Daniel Spoonhower, Jonathan Mace, & Rebecca Isaac, "Distributed Tracing in Practice: Instrumenting, Analyzing, and Debugging Microservices," O'Reilly Media, 2020. [Online]. Available: http://103.203.175.90:81/fdScript/RootOfEBooks/E%20Book%20collection%20-%202023%20-%20G/CSE%20%20IT%20AIDS%20ML/Distributed_Tracing_in_Practice_Instrumenting,_Analyzing,_and_Debugging.pdf

[10] Ron Kohavi et al., "Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing," ResearchGate, 2020. [Online]. Available: https://www.researchgate.net/publication/339914315_Trustworthy_Online_Controlled_Experiments_A_Practical_Guide_to_AB_Testing

[11] Christos Petalotis et al., "An empirical study on the performance and energy costs of ads and analytics in mobile web apps," Information and Software Technology, Volume 166, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584923002252

[12] Adjust, "The ultimate guide to mobile advertising,". [Online]. Available: https://www.adjust.com/resources/guides/mobile-advertising/