Webmunk 2: A Tool for Studying Online Behavior and Digital Platforms

Chiara Farronato, Andrey Fradkin*

January 15, 2026

Abstract

Understanding the behavior of users online is important for researchers, policymakers, and companies. But measuring behavior online and conducting experiments is difficult for independent researchers, who do not have access to the user bases or software of technology companies. We introduce Webmunk 2, an improved open-source tool designed to make conducting online studies much easier. Webmunk 2 is a substantial improvement over the original Webmunk, released in 2024. It is easier to use, cheaper to deploy, and more powerful than the original. The user-facing side of Webmunk 2 is a browser extension that can track consumer browsing behavior, capture screenshots for analysis, and experimentally modify consumers' experiences as they browse the Internet. A key innovation is that Webmunk 2 uses a standard back-end stack consisting of Jitsu for data ingestion, Firebase for user management and configuration, and BigQuery for data storage. This makes it more accessible to researchers and software developers who are familiar with these widely-used tools. Through this extension, researchers can collect a host of consumer data, from URLs to ad impressions, clicks, and user responses to in-browser surveys. The extension can also modify information and change the look of a web page, allowing researchers to implement interventions that vary across study participants. We demonstrate the power of Webmunk 2 by discussing ongoing research applications.

*Farronato: Harvard University, CEPR, NBER, cfarronato@hbs.edu. Fradkin: Boston University, fradkin@bu.edu. Chris Karr (Audacious Software) developed the first version of Webmunk and Unibrix developed the second. We thank Tesary Lin, Alex MacKay, and Chuan Yu for productive collaborations using Webmunk. We thank Jean-Michael Lambert and Wenhan Wang for additional software development. We thank James Dana, Yutao Chen, Yunjie Song, Richard Xu, and Hannah Zhang for outstanding research assistance. We thank Stefan Bucher for comments on the draft. We thank the Digital Business Institute at Boston University and the Internet Society Foundation for helping to fund the project.

The internet has become central to modern economic and social activities. As such, understanding online behavior is critical for policymakers and businesses alike. Researchers have been increasingly interested in exploring consumer behavior online, yet, the tools available to them for conducting experimental studies remain limited. In this paper, we introduce Webmunk 2, an improved open-source tool that allows researchers to manipulate a consumer's online experience and track its effects on the consumer's behavior. Webmunk 2 replaces the original Webmunk (Farronato, Fradkin and Karr (2024)) released in 2024 with a more accessible, cost-effective, and powerful platform. We describe the tool's capabilities and illustrate them with research applications.

Researchers of digital behavior have started to use custom software to conduct studies. Under this paradigm, recruited participants install software on a device such as a laptop or mobile phone, and this software tracks behavior and implements interventions in the course of typical online behavior. This approach has recently been used to study social media (Allcott, Gentzkow and Song (2022), Aridor (2022), Levy (2021), Beknazar-Yuzbashev et al. (2022)). Although these studies require a similar technology, there is no software infrastructure that allows researchers to easily run such studies. Developing this type of technology is costly–both financially and time-wise–, which inevitably creates entry barriers for academics to develop online research studies with credible random variation.

To overcome the challenges imposed by data availability and software development, we developed and open-sourced Webmunk in 2024. Building on lessons learned from deploying the original version, we have now released Webmunk 2, which offers substantial improvements in three key dimensions: ease of use, cost, and capabilities.

Ease of use. The original Webmunk required researchers to deploy and maintain custom Django servers with PostgreSQL databases. Webmunk 2 replaces this custom infrastructure with a standard stack of widely-used cloud services: Jitsu for data ingestion, Firebase for user management and configuration, and BigQuery for data storage and analysis. These are services that many software developers and data scientists are already familiar with, dramatically lowering the barrier to entry. Comprehensive documentation guides researchers and AI coding tools through the setup process, and the modular architecture allows for easy customization.

Cost. The original Webmunk required substantial server infrastructure that scaled with study size. In our original Amazon study, we deployed five large Amazon Web Services servers with several terabytes of storage. Webmunk 2 leverages managed cloud services that scale automatically and offer pay-as-you-go pricing. For most research studies, the infrastructure costs are minimal, as Jitsu, Firebase, and BigQuery all offer generous free tiers that can accommodate pilot studies and small-to-medium scale research projects.

Capabilities. Webmunk 2 includes several new features not available in the original version. These include automated screenshot capture with cloud storage, AI-powered screenshot analysis for detecting ads and other content, an in-browser ad rating system that collects user feedback on ad relevance, and a sophisticated ad personalization module that can interact with ad preference settings on major platforms including Amazon, Facebook, and Google.

Webmunk 2 consists of two main components: a web browser extension and cloud infrastructure. The browser extension component is designed for Chrome and Edge browsers using the modern Manifest V3 specification. The extension can modify the content of web pages in real time and collect data as users browse the Internet. By keeping track of user identities through IDs (e.g., Prolific IDs for studies with subjects recruited on Prolific), the extension can implement different web page manipulations across users or over time, enabling researchers to assign different experimental interventions. The extension uses various techniques to engage study participants in specific tasks beyond their regular browsing activity, such as completing surveys or rating advertisements.

The cloud infrastructure of Webmunk 2 consists of three main services. First, Jitsu handles real-time data ingestion, streaming events from the browser extension directly to BigQuery. Second, Firebase provides user authentication, stores user data in Firestore, manages study configuration through Remote Config, and hosts serverless functions for custom logic. Third, BigQuery serves as the data warehouse where all collected data is stored and can be queried using standard SQL.

All components of Webmunk 2 are open-source and designed to be easily extended. Additional functionality can be added through modular npm packages. We have developed modules for tracking ads across websites, scraping cookies, and managing ad personalization settings on major platforms. Modules are particularly useful for implementing manipulation and tracking that is specific to a web domain or that broadly applies to many domains.

This paper, along with relevant companion documentation available at `webmunk2.org` and `https://github.com/webmunk2/webmunk-extension`, aims to provide researchers and software developers with the necessary information to use Webmunk 2 for conducting new research studies. The use of standard cloud services means that researchers using AI tools can work with Webmunk 2 without specialized training, significantly reducing the cost and time required to deploy studies.

We showcase the functionalities of Webmunk 2 through our ongoing research on online advertising and privacy. Webmunk 2's capabilities make it particularly well-suited for studying how platforms display advertising, how users interact with ads, and how privacy choices affect the advertising ecosystem.

As an illustration of how the software can be used, we describe the software setup used in studies of ad personalization and user privacy preferences. The version of Webmunk 2 customized for these studies tracks user browsing across websites, captures screenshots for ad detection and analysis, and guides users through changing their ad personalization settings on major platforms.

Our work contributes to the broader Open Science movement (Nosek, Spies and Motyl (2012), Spellman, Gilbert and Corker (2017)) which calls for researchers to make research data and software widely available. In economics, development of open-source software is becoming part of a researcher's professional activity. For example, Conlon and Gortmaker (2020) develop PyBLP, an open-source package to estimate demand models, and Shen et al. (2021) develop a novel framework to improve the efficiency of layout detection and extract text from historical documents.

In addition to software-based studies, there are other paradigms for studying digital behavior. One paradigm is to use data collected by third-parties or scraped by researchers.[1] For example, Santos, Hortaçsu and Wildenbeest (2012) use data from ComScore[2] to study search behavior online, Calder-Wang (2021) uses Airbnb data scraped by AirDNA[3] to study the effects of Airbnb entry on the rental market, and Lewis (2011) uses scraped data from eBay motors to study asymmetric information and disclosure. The limit of this approach is that third-party or scraped data may measure a limited set of variables and may not contain experimental variation of research relevance.

Another approach to studying digital behavior is to use confidential data directly from the company that generates it. This approach has two main advantages. First, it is less susceptible to measurement error and selection issues (Farronato and Fradkin 2022). Second, it allows researchers to use randomized controlled trials conducted by the company in order to identify causal effects (Blake, Nosko and Tadelis 2015). Despite the advantages, most researchers do not have the option of using confidential data. Companies are hesitant to share these data for a variety of reasons, including the perception of regulatory, legal, or competitive risk. When companies do collaborate with researchers, researchers may be limited to studying topics that may benefit the company, or, at the very least, will not hurt it. This creates a selection problem for the type of questions that researchers can hope to answer through company collaborations. A key advantage of our software based paradigm is that it only requires permission from the researchers' Internal Review Boards (IRB) and from study participants.

[1] Although the legality of scraping has faced some recent challenges (`https://techcrunch.com/2022/04/18/web-scraping-legal-court/`).

[2] `https://www.comscore.com/`.

[3] `https://www.airdna.co/`

The rest of the paper proceeds as follows. Section 1 illustrates Webmunk 2's functionalities, and Section 2 describes research applications. Section 3 discusses the technical skills required to use Webmunk 2 for new studies. Section 4 presents Webmunk 2's technical aspects and its modular structure. Section 5 concludes.

1 Webmunk 2's Functionalities

In this section, we provide a comprehensive overview of Webmunk 2's functionalities. Our aim is to offer researchers who are considering deploying Webmunk 2 a clear understanding of the possibilities it enables. By delving into the details, we aim to showcase the full range of capabilities that Webmunk 2 offers for conducting research studies. We hope that by leveraging this technology, other researchers can add to Webmunk 2's existing functionalities.

It is useful to compare Webmunk 2 to an ad blocker, which many readers may be familiar with. From the user's perspective, Webmunk 2 can be found on the Chrome store and can be installed like any other Chrome extension. After installation, the user is prompted to enter their Prolific ID, which serves as their unique identifier.[4] A Firebase function validates the Prolific ID and creates a user record in Firestore. After enrollment, Webmunk 2 operates in the background while the browser is open, and can be found in the list of installed Chrome extensions.[5]

The capabilities of Webmunk 2 can be broadly classified into three main groups: tracking capabilities, manipulation capabilities, and new capabilities for studying advertising. We devote a subsection to each.

1.1 What Can Webmunk 2 Track?

Webmunk 2 can track a variety of digital data that the consumer generates while browsing the Internet. We have included a number of tracking modules that we developed for our ongoing studies. Researchers can easily include a subset of the existing data tracking capabilities or develop other modules to track additional behavior and data.

When it comes to tracking user behavior online, researchers need to strike the right balance between tracking everything, which may put private data at risk or dissuade users from participating in a research study, and not tracking enough, which prevents the researchers from conducting a thorough study of user behavior. The approach we take with Webmunk 2 is a

[4] Webmunk 2 uses Prolific IDs as the primary identification method, which integrates seamlessly with Prolific's participant recruitment platform.

[5] Extensions on the Chrome browser can be accessed by navigating to `chrome://extensions/`.

conservative one, where the default is not to track, unless explicitly stated otherwise. This is implemented through configurable domain exclusion lists that prevent tracking on sensitive websites.

In addition to ex-ante conditions for tracking, researchers can ensure the privacy of certain confidential information ex-post. For example, researchers can implement models to identify searches that likely contain personally identifiable information (such as addresses or first and last names) and remove them from the data ex-post. Researchers should also consult their respective Institutional Review Boards and information technology support resources to ensure data are stored in accordance with best practices.

We classify the types of data that Webmunk 2 can track into six categories: URL tracking, ad detection, cookies, screenshots, user activity, and installed extensions. We describe each of them below.

URL tracking. Webmunk 2 can track URLs that users visit. This is already a very useful set of data because it contains information about which websites users visit and when. Athey, Mobius and Pal (2021) use this type of data to study the effect of news aggregator on the consumption of news. The URL tracking module allows researchers to configure which domains to exclude from tracking, ensuring that sensitive websites (such as banking or healthcare portals) are not monitored.

Ad detection. The ad detection module (`@webmunk/extension-ads`) identifies and tracks advertisements displayed on web pages. When an ad is detected, the module captures metadata including the ad's title, company, text content, coordinates on the page, and URLs. The module also tracks when users click on ads, capturing the click event along with the initial and redirected URLs. This comprehensive ad tracking enables researchers to study how users interact with advertising across the web.

Cookies. The cookies module (`@webmunk/cookies-scraper`) tracks cookies stored by the user's browser. This allows researchers to understand how websites use cookies for advertising and tracking purposes, and how cookies change as a function of users' privacy preferences. The module can be triggered to record all cookies at specific points during a study, such as before and after a user changes their privacy settings.

Screenshots. A major new capability in Webmunk 2 is automated screenshot capture. The extension can capture screenshots of the visible browser tab at configurable intervals (by default, once per minute when the user is actively browsing). Screenshots are compressed and uploaded to Firebase Storage, where they can be analyzed using AI-powered services. This capability is particularly useful for detecting and analyzing visual content such as advertisements that may be difficult to capture through DOM inspection alone. The screenshot service respects the domain exclusion list, ensuring that screenshots are not captured on sensitive websites.

User activity. Webmunk 2 tracks user activity including scrolling and clicking. The extension monitors scroll events (throttled to avoid excessive data collection) and click events, which are used to trigger screenshot capture when users are actively engaging with a page. This activity tracking helps researchers understand how users interact with web content.

Installed extensions. Upon enrollment, Webmunk 2 records the list of browser extensions installed by the user. This is particularly important for studies of advertising, as ad blockers can significantly affect the ads that users see. By knowing which extensions are installed, researchers can account for these factors in their analysis.

1.2 What Can Webmunk 2 Manipulate?

Browser extensions have the capability of manipulating many aspects of a user's experience while using the Internet. In this section, we describe the capabilities we have developed for Webmunk 2.

Inserting notifications and surveys. Webmunk 2 has the ability to inject notifications and survey prompts into web pages. Notifications appear as overlays on the current page and can be used to remind users to complete tasks, inform them about study milestones, or prompt them to uninstall the extension at the end of a study. Surveys are managed through Firebase Remote Config, which allows researchers to configure survey URLs and timing without updating the extension itself.

Ad rating prompts. A key new feature in Webmunk 2 is the ability to prompt users to rate advertisements. When ads are detected on a page, the extension can display a non-intrusive popup asking users questions such as "Did you find this ad relevant?" with Yes/No response options. User responses are recorded along with the ad metadata, enabling researchers to study the relationship between ad characteristics and user perceptions. The rating prompts include a configurable cooldown period (by default, 10 minutes) to avoid survey fatigue.

Ad personalization automation. The ad personalization module (`@webmunk/ad-personalization`) can automatically navigate to and interact with ad preference settings on major platforms including Amazon, Facebook, Instagram, YouTube, and Google. This capability allows researchers to experimentally manipulate users' ad personalization settings as part of a study design. The module uses platform-specific strategies to locate and interact with the relevant settings pages, and records whether the manipulation was successful.

Acting as a user proxy. Webmunk 2 can act on behalf of the user in filling out forms and selecting options on a webpage. The ad personalization module uses this capability to change ad preference settings. Researchers can extend this functionality for other purposes, such as automatically dismissing cookie consent forms or promotional banners.

1.3 New Capabilities for Studying Advertising

Webmunk 2 includes several new capabilities specifically designed for studying online advertising:

AI-powered screenshot analysis. Screenshots captured by the extension can be analyzed using a cloud-based AI service hosted on Google Cloud Run. The analysis can detect the presence of advertisements in screenshots and extract relevant information. This enables researchers to identify ads that may not be detectable through DOM inspection, such as video ads or ads loaded in complex iframe structures.

Cross-platform ad personalization. The ad personalization module supports automated manipulation of ad settings across multiple major platforms. Researchers can assign participants to different conditions that affect their ad personalization settings on Amazon, Facebook, Google, and YouTube. The module tracks whether each manipulation was successful and records any errors encountered.

Comprehensive ad event tracking. The extension tracks multiple ad-related events including ad detection (`ad_detected`), ad clicks (`ad_clicked`), and ad ratings (`ads_rated`). Each event includes rich metadata about the ad and the context in which it was displayed, enabling detailed analysis of the advertising ecosystem.

2 Research Applications

In this section, we describe research questions for which Webmunk 2 has proved useful in collecting user behavior and implementing experimental designs.

Studying Online Advertising. Webmunk 2 is particularly well-suited for studying online advertising. The extension can track ads across websites, capture screenshots for visual analysis, and collect user feedback on ad relevance. Researchers can study questions such as: How do ad personalization settings affect the types of ads users see? Do users find personalized ads more or less relevant? How does ad blocking affect user browsing behavior?

To study these questions, researchers can recruit participants through Prolific and have them install the Webmunk 2 extension. Participants enter their Prolific ID, which creates their user record in Firebase. The extension then tracks their browsing activity, captures screenshots, detects ads, and periodically prompts users to rate ad relevance.

The experimental component can involve randomly assigning participants to different ad personalization conditions. For some participants, the extension can use the ad personalization module to enable ad personalization on major platforms, while for others it can disable personalization or leave settings unchanged. By comparing outcomes across these conditions, researchers can identify the causal effects of ad personalization on user experience and advertising effectiveness.

Survey Integration. Webmunk 2 integrates with external survey platforms such as Qualtrics. Surveys are configured through Firebase Remote Config and appear as tasks in the extension popup. The extension tracks when users start and complete surveys, and can pass participant identifiers and experimental condition information to the survey URL as query parameters. This enables researchers to collect survey responses that are linked to the behavioral data collected by the extension.

Survey timing is configurable, allowing researchers to schedule surveys at specific intervals (e.g., one week after installation, two weeks after installation). The extension shows notifications reminding users to complete pending surveys, with configurable delays between notifications to avoid fatigue.

Study Lifecycle Management. Webmunk 2 includes features to manage the full lifecycle of a research study. The extension can detect when all study tasks are complete (e.g., all surveys finished, all ad personalization settings checked) and prompt users to uninstall the extension. Researchers can also configure an uninstall URL that redirects users to a completion page when they remove the extension, enabling proper study closeout and compensation.

3 Webmunk 2 Project Requirements

We anticipate that most social science researchers can implement modifications to Webmunk 2 for their research purposes by collaborating with AI-assisted coding tools. This section provides guidance regarding the services that are necessary to work with Webmunk 2.

Cloud service setup. Setting up the cloud infrastructure for Webmunk 2 requires familiarity with three main services:

1. **Jitsu** handles data ingestion. Jitsu can be deployed through Elestio or self-hosted. Configuration involves setting up a BigQuery destination and generating API keys for the extension.

2. **Firebase** handles user management and configuration. Setup involves creating a Firebase project, configuring Firestore security rules, setting up Remote Config parameters for

surveys and other settings, and deploying Firebase Functions for user authentication.

3. **BigQuery** stores the collected data. Setup involves creating a dataset and ensuring proper permissions for the Jitsu service account.

Comprehensive documentation is provided in the `jitsu-bigquery-firebase-setup.md` file in the repository, which walks through each step of the setup process.

Extension customization. The Webmunk 2 extension is built with TypeScript and uses Webpack for bundling. Customizing the extension requires:

1. TypeScript and modern JavaScript development practices
2. Chrome extension development using Manifest V3
3. Configuration of environment variables for the different cloud services
4. Npm (Node Package Manager) for managing dependencies and building the extension

The extension uses a modular architecture where tracking and manipulation capabilities are implemented as separate npm packages. Researchers can include or exclude modules based on their study needs, and can develop custom modules following the established patterns.

Chrome Web Store submission. Deploying the extension to participants requires submitting it to the Chrome Web Store. This involves:

1. Creating a Chrome Web Store developer account
2. Building the production version of the extension
3. Submitting the extension for review
4. Responding to any issues raised by Google's reviewers

The build process is automated through npm scripts, and the repository includes configuration for creating a zip file suitable for submission.

Comparison to original Webmunk. The original Webmunk required expertise in Unix server administration, Django development, and PostgreSQL database management. Webmunk 2's use of managed cloud services eliminates the need for server administration.

4 Technical and Implementation Details

The Webmunk 2 platform has been redesigned from the ground up to leverage modern cloud services while maintaining the flexibility and extensibility that made the original Webmunk valuable.

4.1 Technical Details About Webmunk 2 Browser Extensions

Webmunk 2 extensions are built using Chrome's Manifest V3 specification, the latest standard for browser extensions. The extension is written in TypeScript, which provides type safety and improved developer experience compared to plain JavaScript.

Every browser extension features two main functional components: content scripts, which interact with web pages and the users directly, and service workers, which provide functionality and services in the background. For a variety of security and functional considerations, the browser permits each to accomplish specific tasks. A content script may directly observe and manipulate a page element in the page it is loaded, while a service worker can transmit data to a server, even if no pages are open.

The main service worker (`Worker.ts`) coordinates the extension's functionality by:

1. Initializing Firebase authentication
2. Loading survey configurations from Firebase Remote Config
3. Setting up event listeners for messages from content scripts
4. Managing the ad personalization workflow
5. Coordinating screenshot capture and analysis

Content scripts are injected into web pages to track user activity and display notifications. The content script monitors scroll and click events, which trigger the screenshot service to capture the visible tab.

4.2 Event Tracking with Jitsu

Data collection in Webmunk 2 uses Jitsu's JavaScript SDK for real-time event streaming. The `EventService` class wraps the Jitsu client and handles all event tracking:

```
import { jitsuAnalytics } from "@jitsu/js";

export class EventService {
  private client: any;

  constructor() {
    this.client = jitsuAnalytics({
      writeKey: JITSU_WRITE_KEY,
      host: JITSU_INGEST_URL
    });
  }

  async track(event: string, properties: any): Promise<void> {
    const user = await this.firebaseAppService.getUser();
    this.client.identify(user.uid, { $doNotSend: true });
    await this.client.track({ event, ...properties });
  }
}
```

Events are streamed in real-time to Jitsu, which forwards them to BigQuery. This eliminates the store-and-forward complexity of the original Webmunk while providing automatic retry and buffering through Jitsu's infrastructure.

4.3 User Management with Firebase

User authentication and data storage use Firebase services. The `FirebaseAppService` class manages:

1. **Authentication**: Users sign in anonymously through Firebase Auth, with their Prolific ID stored in Firestore.
2. **User data**: User records are stored in Firestore and include the Prolific ID, Firebase UID, and active status.
3. **Screenshot storage**: Screenshots are uploaded to Firebase Storage and URLs are recorded in event data.

A Firebase Function (`signIn`) handles user registration:

```
export const signIn = functions.https.onCall(async (data) => {
  const { prolificId } = data;
  // Validate Prolific ID format
  // Create or retrieve user record in Firestore
  // Return user data to extension
});
```

4.4 Configuration with Firebase Remote Config

Study configuration is managed through Firebase Remote Config, which allows researchers to update settings without deploying a new version of the extension. Configuration parameters include:

- Survey URLs and names
- Timing intervals for surveys, notifications, and ad personalization
- Domain exclusion lists for URL tracking and screenshots
- Flags to stop data collection globally or for specific users

The `ConfigService` class fetches and caches configuration values:

```
async getConfigByKey(key: string): Promise<string> {
  await fetchAndActivate(this.remoteConfig);
  const value = getValue(this.remoteConfig, key);
  return value.asString();
}
```

4.5 Modular Architecture

Webmunk 2 uses a modular architecture where tracking and manipulation capabilities are implemented as separate npm packages:

- `@webmunk/extension-ads`: Ad detection and tracking
- `@webmunk/cookies-scraper`: Cookie tracking
- `@webmunk/ad-personalization`: Ad preference automation

- `@webmunk/utils`: Shared utilities

Modules communicate through a message-driven architecture. The main worker registers listeners for module events:

```
messenger.addModuleListener('ads-scraper',
  this.onModuleEvent.bind(this));
messenger.addModuleListener('cookies-scraper',
  this.onModuleEvent.bind(this));
messenger.addModuleListener('ad-personalization',
  this.onAdPersonalizationModuleEvent.bind(this));
```

When modules detect events (e.g., ad detected, cookie recorded), they emit events that the main worker forwards to the EventService for recording in BigQuery.

4.6 Data Flow

The complete data flow in Webmunk 2 is:

1. User browses the web with the extension installed
2. Content scripts detect events (page loads, ads, clicks)
3. Events are sent to the service worker via Chrome messaging
4. The service worker enriches events with user context
5. The EventService sends events to Jitsu
6. Jitsu streams events to BigQuery in real-time
7. Screenshots are uploaded to Firebase Storage with URLs included in event data

This architecture provides reliability (Jitsu handles retries), scalability (managed cloud services), and queryability (standard SQL in BigQuery).

5 Conclusion

This paper has presented Webmunk 2, an improved open-source framework for conducting digital experiments. Webmunk 2 represents a substantial upgrade over the original Webmunk released in 2024, offering three key improvements: easier setup through standard cloud services, lower costs, and new capabilities for studying online advertising.

The use of Jitsu, Firebase, and BigQuery as the backend infrastructure makes Webmunk 2 more accessible to a wider range of developers and researchers. These are widely-used services with extensive documentation and community support.

New capabilities in Webmunk 2, including automated screenshot capture, AI-powered analysis, ad rating prompts, and cross-platform ad personalization automation, make it particularly well-suited for studying the online advertising ecosystem. These features enable research questions that were difficult or impossible to address with the original Webmunk.

The paper has explained how to use Webmunk 2 for research studies and described the technical skills required for customization. The modular TypeScript architecture and standard npm tooling mean that developers familiar with modern JavaScript can quickly become productive with the codebase.

Several directions for future work remain. Privacy-preserving data sharing through differential privacy techniques could enable researchers to share datasets more broadly. Integration with mobile platforms remains a challenge, as the browser extension approach does not work on phones where most activity occurs through apps. Improved recruitment strategies could reduce the cost of finding study participants.

We hope that Webmunk 2 will enable researchers to study important questions about online behavior, digital platforms, and the advertising ecosystem. The open-source nature of the project means that improvements and extensions developed by the research community can benefit all users.

References

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2022. "Digital addiction." *American Economic Review*, 112(7): 2424–63.

Aridor, Guy. 2022. "Drivers of Digital Attention: Evidence from a Social Media Experiment." *Available at SSRN 4069567*.

Athey, Susan, Markus Mobius, and Jeno Pal. 2021. "The impact of aggregators on internet news consumption."

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski. 2022. "Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment." *Working Paper*.

Blake, Thomas, Chris Nosko, and Steven Tadelis. 2015. "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment." *Econometrica*, 83(1): 155–174.

Calder-Wang, Sophie. 2021. "The distributional impact of the sharing economy on the housing market." *Available at SSRN 3908062*.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best practices for differentiated products demand estimation with pyblp." *The RAND Journal of Economics*, 51(4): 1108–1161.

Farronato, Chiara, and Andrey Fradkin. 2022. "The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry." *American Economic Review*, 112(6): 1782–1817.

Farronato, Chiara, Andrey Fradkin, and Chris Karr. 2024. "Webmunk: A new tool for studying online behavior and digital platforms." National Bureau of Economic Research.

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American economic review*, 111(3): 831–870.

Lewis, Gregory. 2011. "Asymmetric information, adverse selection and online disclosure: The case of eBay motors." *American Economic Review*, 101(4): 1535–1546.

Nosek, Brian A, Jeffrey R Spies, and Matt Motyl. 2012. "Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability." *Perspectives on Psychological Science*, 7(6): 615–631.

Santos, Babur De los, Ali Hortaçsu, and Matthijs R Wildenbeest. 2012. "Testing models of consumer search using data on web browsing and purchasing behavior." *American economic review*, 102(6): 2955–2980.

Shen, Zejiang, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, and Weining Li. 2021. "LayoutParser: A unified toolkit for deep learning based document image analysis." 131–146, Springer.

Spellman, Bobbie, Elizabeth Gilbert, and Katherine S Corker. 2017. "Open science: What, why, and how."