

Infraestrutura de Rastreamento Comportamental na Web: Implicações Sociais no Contexto Brasileiro

Antony Seabra, Claudio Cavalcante, Sérgio Lifschitz

¹PUC-Rio - Departamento de Informática, Rio de Janeiro, RJ

{amedeiros, ccavalcante, sergio}@inf.puc-rio.br

Abstract. *This paper empirically investigates the infrastructure of behavioral data collection that underpins the digital advertising ecosystem. Through network traffic analysis, we examine the data flows generated during interactions of Brazilian users with websites and search engines. The results reveal the systematic presence of multiple third-party entities that collect and correlate information in real time, forming a distributed tracking architecture. This infrastructure extends beyond commercial personalization, constituting a sociotechnical system with the potential to influence individual preferences and decisions. The study contributes to ongoing discussions on algorithmic transparency, data governance, and the societal risks associated with surveillance-based economies.*

Resumo. *Este artigo investiga empiricamente a infraestrutura de coleta de dados comportamentais que sustenta o ecossistema de publicidade digital. Por meio de análise de tráfego de rede, examinamos os fluxos de dados gerados durante interações de usuários brasileiros com websites e mecanismos de busca. Os resultados evidenciam a presença sistemática de múltiplas entidades terceiras que coletam e correlacionam informações em tempo real, configurando uma arquitetura distribuída de rastreamento, que vai além da personalização comercial, constituindo um sistema sociotécnico com potencial de influenciar preferências e decisões. O estudo contribui para o debate sobre transparência algorítmica, governança de dados e riscos sociais associados à economia baseada em monitoramento.*

1. Introdução

A crescente digitalização das atividades cotidianas transformou a web em uma infraestrutura central para comunicação, consumo, trabalho e participação social. Nesse cenário, a coleta e o processamento massivo de dados comportamentais tornaram-se elementos estruturais dos modelos econômicos que sustentam serviços digitais contemporâneos. Plataformas online, mecanismos de busca e redes de publicidade operam por meio de sistemas automatizados capazes de capturar, correlacionar e analisar sinais detalhados de interação dos usuários, incluindo histórico de navegação, padrões de consumo, localização e preferências inferidas. Esse processo ocorre de forma contínua e distribuída, frequentemente envolvendo múltiplas entidades terceiras que participam de cadeias complexas de transmissão e enriquecimento de dados.

Embora práticas de monitoramento digital sejam frequentemente justificadas como necessárias para personalização de serviços e otimização de experiências, estudos recentes indicam que tais mecanismos constituem uma infraestrutura técnica ampla e persistente de rastreamento comportamental. Essa infraestrutura não apenas viabiliza modelos

comerciais baseados em segmentação algorítmica, mas também introduz novas dinâmicas de poder informacional, nas quais organizações privadas passam a deter capacidades extensivas de observação, previsão e potencial influência sobre indivíduos e grupos sociais. A opacidade dessas operações, combinada com a dificuldade de auditoria por parte dos usuários, levanta questionamentos relevantes acerca de transparência, autonomia informacional e responsabilidade institucional.

Apesar do crescimento do debate acadêmico e regulatório sobre privacidade e proteção de dados, ainda existe uma lacuna significativa de evidências empíricas diretas sobre o funcionamento concreto dessas infraestruturas em contextos reais de uso da web, especialmente em ambientes nacionais específicos. Grande parte da literatura concentra-se em análises conceituais, avaliações normativas ou estudos das políticas declaradas por plataformas, enquanto investigações que observam diretamente os fluxos de dados transmitidos durante interações reais permanecem relativamente escassas. Essa limitação dificulta a compreensão precisa da extensão, frequência e natureza dos mecanismos de rastreamento efetivamente operantes.

Este trabalho busca contribuir para o preenchimento dessa lacuna por meio de uma análise empírica do tráfego de rede gerado durante sessões reais de navegação e uso de mecanismos de busca em ambiente brasileiro. Utilizando uma infraestrutura controlada de interceptação e inspeção de pacotes, capturamos e examinamos comunicações entre dispositivos de usuário e servidores externos, permitindo identificar domínios de terceiros acionados, tipos de dados transmitidos e padrões de circulação de informações comportamentais. Os resultados evidenciam que mesmo interações simples com websites podem desencadear cadeias extensas de compartilhamento de dados com múltiplos serviços externos, caracterizando uma arquitetura distribuída de coleta e redistribuição de informações.

As evidências apresentadas indicam que a infraestrutura de rastreamento comportamental deve ser compreendida como um sistema sociotécnico com implicações que ultrapassam o domínio comercial. A capacidade de construir perfis comportamentais detalhados e continuamente atualizados cria condições para mecanismos de segmentação altamente precisos, os quais podem ser utilizados não apenas para publicidade personalizada, mas também para direcionamento informacional, modulação de conteúdo e potencial influência sobre percepções e decisões individuais. Esse cenário introduz desafios relevantes para o equilíbrio informacional em ambientes digitais, sobretudo quando tais processos ocorrem de forma invisível aos usuários.

Além das dimensões sociais, essa infraestrutura levanta questões jurídicas e regulatórias relevantes. No contexto brasileiro, a Lei Geral de Proteção de Dados estabelece princípios como finalidade, necessidade, transparência e responsabilização no tratamento de dados pessoais. Entretanto, a natureza distribuída e transnacional dos ecossistemas de rastreamento, não necessariamente em território brasileiro, dificulta a aplicação efetiva desses princípios, especialmente quando dados são compartilhados entre múltiplos agentes com diferentes jurisdições e níveis de responsabilização. Compreender empiricamente como esses fluxos ocorrem constitui, portanto, um passo fundamental para subsidiar discussões sobre governança de dados, fiscalização regulatória e formulação de políticas públicas.

Este artigo apresenta três contribuições principais. Primeiro, fornece evidência empírica direta da operação de infraestruturas de rastreamento comportamental durante interações reais na web, obtida por meio de análise de tráfego de rede em nível de pacote. Segundo, caracteriza o ecossistema técnico envolvido nesses fluxos, identificando categorias de entidades terceiras e mecanismos responsáveis pela coleta e correlação de dados. Terceiro, discute as implicações sociais e legais associadas a essa infraestrutura, argumentando que ela deve ser interpretada como um componente estrutural do ambiente digital contemporâneo, com potenciais efeitos sobre autonomia informacional e processos sociais mediados por tecnologia.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta o referencial teórico sobre monitoramento digital e capitalismo de vigilância. A Seção 3 descreve a metodologia experimental utilizada para captura e análise de tráfego de rede. A Seção 4 apresenta os resultados empíricos e estudos de caso observados. A Seção 5 discute as implicações sociotécnicas e regulatórias dos achados. Por fim, a Seção 6 apresenta as conclusões e direções para trabalhos futuros.

2. Referencial Teórico e Trabalhos Relacionados

2.1. Capitalismo de Vigilância e Infraestruturas de Dados

O conceito de capitalismo de vigilância descreve um modelo econômico no qual dados comportamentais extraídos das interações digitais são convertidos em ativos estratégicos para previsão e modulação de comportamento [Zuboff 2019, Zuboff 2022]. Experiências humanas são traduzidas em sinais computacionais que alimentam sistemas preditivos e mercados de segmentação. A captura contínua de dados deixa de ser um efeito colateral da prestação de serviços e passa a constituir o núcleo do modelo de geração de valor.

Autores recentes ampliam essa perspectiva ao interpretar tais práticas como parte de infraestruturas sociotécnicas persistentes.[Couldry and Mejias 2019] argumentam que a extração sistemática de dados configura uma forma de colonialismo de dados, na qual a vida social é apropriada como matéria-prima para sistemas computacionais. De forma complementar, [Pasquale 2015] discute a opacidade algorítmica e como ela dificulta a responsabilização e amplia assimetrias informacionais entre plataformas e usuários. Essas abordagens convergem ao indicar que a coleta de dados deve ser analisada como estrutura de poder e não apenas como prática técnica.

No contexto da publicidade comportamental, a literatura demonstra que mecanismos de rastreamento online operam por meio de redes complexas de intermediários, incluindo plataformas de leilão em tempo real e serviços de sincronização de identificadores [Marotta et al. 2022, Englehardt and Narayanan 2016]. Esses sistemas permitem a consolidação de perfis detalhados a partir de múltiplas fontes, viabilizando segmentação automatizada em larga escala.

2.2. Rastreamento Web e Evidências Empíricas

Estudos empíricos sobre rastreamento web têm evidenciado a ubiquidade de mecanismos de coleta distribuída. [Englehardt and Narayanan 2016] demonstraram a presença massiva de rastreadores de terceiros em websites populares, incluindo técnicas avançadas de fingerprinting. Trabalhos mais recentes mostram que tais práticas continuam evoluindo,

incorporando métodos resilientes a bloqueadores de anúncios, algoritmos de ofuscamento e mecanismos de anonimização [Acar et al. 2014, Iqbal et al. 2021].

[Marotta et al. 2022] analisaram o ecossistema de publicidade programática, evidenciando a circulação de dados em processos automatizados de leilão em tempo real. Esses fluxos envolvem múltiplos atores e frequentemente ocorrem em frações de segundo, dificultando a compreensão do percurso completo das informações transmitidas. Tais achados reforçam a caracterização da infraestrutura de rastreamento como distribuída e altamente interconectada.

Apesar da consolidação desses estudos no cenário internacional, investigações empíricas específicas no contexto brasileiro ainda são limitadas. Considerando particularidades regulatórias, culturais e de mercado, análises locais tornam-se relevantes para compreender como infraestruturas globais se manifestam em ambientes nacionais.

2.3. Implicações Sociais e Regulatórias

A literatura recente também discute os impactos sociais da publicidade comportamental e da coleta extensiva de dados. [Wu et al. 2023] argumentam que a publicidade baseada em comportamento pode produzir efeitos cumulativos de dano social, incluindo manipulação informacional e reforço de desigualdades. Essas preocupações ganham relevância quando mecanismos de segmentação são aplicados a conteúdos políticos ou narrativas sensíveis.

Do ponto de vista regulatório, a implementação de marcos legais como o Regulamento Geral de Proteção de Dados na União Europeia e a Lei Geral de Proteção de Dados no Brasil busca estabelecer limites e princípios para o tratamento de dados pessoais [Andrew and Baker 2021]. Entretanto, a literatura aponta que a complexidade técnica das cadeias de compartilhamento e a fragmentação transnacional dos serviços dificultam a fiscalização efetiva e a aplicação prática de princípios como transparência e finalidade.

Nesse cenário, a produção de evidências empíricas observacionais sobre fluxos reais de dados torna-se fundamental para subsidiar debates normativos e políticas públicas. Ao articular análise técnica e discussão sociotécnica, o presente trabalho dialoga com essa linha de pesquisa e contribui para a compreensão da infraestrutura de rastreamento comportamental no contexto brasileiro.

3. Metodologia

Utilizamos uma abordagem empírica baseada na análise observacional de tráfego de rede, com o objetivo de identificar e caracterizar os fluxos de dados gerados durante sessões reais de navegação na web. O foco metodológico consistiu na captura e inspeção de comunicações entre dispositivos de usuário e servidores externos, permitindo observar interações que não são diretamente visíveis na interface do navegador.

3.1. Ambiente Experimental

O experimento foi conduzido em ambiente controlado, utilizando um servidor proxy configurado no modelo man-in-the-middle para interceptação de tráfego HTTP e HTTPS. O proxy foi instalado entre o dispositivo cliente e a rede externa, possibilitando o registro das requisições realizadas durante sessões de navegação. Para permitir a inspeção de conexões criptografadas, foi instalado no dispositivo um certificado digital associado ao

proxy experimental, viabilizando a descriptografia local dos pacotes para fins exclusivamente analíticos.

As sessões de navegação foram realizadas em dispositivos configurados especificamente para o experimento, sem uso de extensões bloqueadoras de anúncios ou ferramentas de anonimização, a fim de observar o comportamento padrão dos websites em condições ordinárias de acesso.

3.2. Procedimento de Coleta

Foram conduzidas sessões controladas de navegação em websites populares no contexto brasileiro, bem como consultas em mecanismos de busca amplamente utilizados. Durante cada sessão, o proxy registrou:

- Domínios acionados
- Requisições HTTP e HTTPS
- Parâmetros transmitidos nos pacotes
- Identificadores persistentes observados
- Encadeamento temporal de chamadas subsequentes

Os registros foram armazenados para posterior análise, preservando-se apenas informações relevantes à caracterização dos fluxos técnicos, sem retenção de dados pessoais identificáveis além daqueles estritamente necessários à análise estrutural do tráfego.

3.3. Critérios de Análise

A análise concentrou-se em quatro dimensões principais:

1. Volume de domínios terceiros acionados por sessão
2. Classificação funcional dos serviços identificados
3. Persistência de identificadores entre sessões
4. Encadeamento e propagação de fluxos de dados

Os domínios observados foram categorizados manualmente com base em sua função declarada e documentação pública disponível, distinguindo redes de publicidade, serviços de analytics, sistemas de sincronização de identidade e infraestruturas de publicidade programática.

3.4. Limitações

A metodologia adotada apresenta limitações inerentes. A amostra de navegação não representa a totalidade dos padrões de uso da web no Brasil, e os resultados refletem condições específicas do ambiente experimental. Além disso, parte do conteúdo transmitido pode permanecer opaco devido a técnicas adicionais de ofuscação ou criptografia ponta a ponta. Ainda assim, a abordagem observacional adotada permite caracterizar padrões estruturais da infraestrutura de rastreamento operante durante interações reais.

3.5. Considerações Éticas

O experimento foi conduzido exclusivamente em ambiente controlado, com dispositivos configurados para fins de pesquisa e sem interceptação de tráfego de terceiros não consentidos. A análise concentrou-se na estrutura técnica dos fluxos de dados, sem divulgação de informações sensíveis associadas a indivíduos específicos.

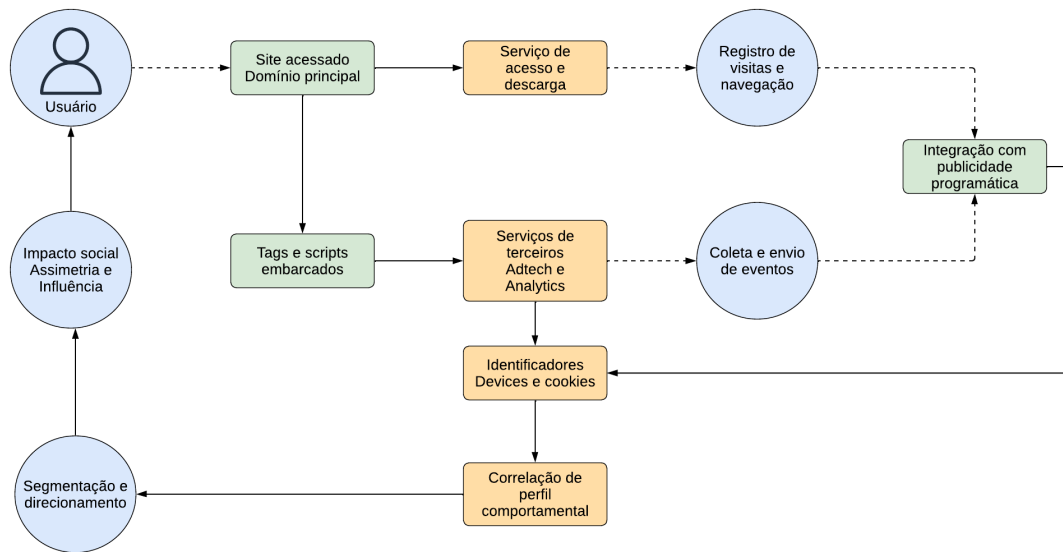


Figure 1. Visão conceitual da infraestrutura de rastreamento comportamental: a interação do usuário aciona scripts embarcados que comunicam com serviços terceiros, viabilizando coleta de eventos, identificação, correlação de perfil e integração com publicidade programática, com efeitos de segmentação e impactos sociais associados.

4. Resultados Empíricos

Esta seção apresenta os resultados obtidos a partir da captura e análise do tráfego de rede gerado durante sessões controladas de navegação na web e utilização de mecanismos de busca. A análise concentrou-se na identificação de domínios terceiros acionados, tipos de serviços envolvidos, padrões de transmissão de dados e recorrência de mecanismos de rastreamento.

A Figura 1 sintetiza o fluxo conceitual que orienta a leitura dos achados empíricos apresentados nesta seção. O processo inicia-se quando o usuário acessa um website no navegador ou em um aplicativo. Essa ação estabelece a comunicação com o domínio principal do site, responsável por entregar o conteúdo solicitado. A partir desse primeiro contato, o site frequentemente carrega recursos adicionais que não fazem parte do conteúdo em si, como tags, bibliotecas e scripts embarcados, os quais são executados no ambiente do cliente e passam a coordenar chamadas subsequentes.

Na sequência, os scripts embarcados acionam serviços de terceiros, tipicamente associados a ecossistemas de adtech e analytics. Esses serviços podem incluir mensuração de audiência, monitoramento de desempenho, recomendação de conteúdo, entrega de anúncios e sincronização de identificadores. Do ponto de vista observacional, esse estágio se manifesta como a abertura de conexões para domínios externos que não pertencem ao site originalmente acessado, muitas vezes com requisições realizadas em paralelo e em alta frequência.

Uma vez ativados, os serviços de terceiros passam a coletar e receber eventos de navegação. Esses eventos podem representar ações do usuário, páginas visitadas, parâmetros de sessão e sinais técnicos do ambiente do dispositivo. Em paralelo, observa-se a utilização de identificadores persistentes, obtidos por meio de cookies, identificadores

de dispositivo ou outros mecanismos equivalentes. O papel desses identificadores é permitir a vinculação das interações atuais a registros anteriores, estabelecendo continuidade entre sessões e entre diferentes domínios.

Com base nessa camada de identificação, ocorre a correlação de perfil comportamental. Nesse estágio, informações provenientes de múltiplas interações são agregadas e associadas a um mesmo identificador, viabilizando inferências sobre interesses, preferências e padrões de comportamento. Essa etapa é central para a transição entre coleta de eventos e capacidade de segmentação, pois converte dados observacionais em perfis utilizáveis por sistemas de decisão automatizada.

Em seguida, a infraestrutura é integrada a mecanismos de publicidade programática. Essa integração conecta a informação comportamental e os identificadores a processos automatizados de compra e venda de inventário publicitário. O resultado prático é a ativação de mecanismos de segmentação e direcionamento, nos quais conteúdos e anúncios são escolhidos e exibidos de maneira diferenciada conforme o perfil associado ao usuário.

Por fim, os efeitos dessa segmentação produzem impactos sociais associados a assimetria informacional e potencial influência. A personalização resultante tende a operar como uma camada de mediação invisível, na qual diferentes usuários recebem estímulos distintos com base em perfis não transparentes. Essa dinâmica retroalimenta o ciclo, pois as respostas do usuário a conteúdos direcionados geram novos eventos de navegação, reforçando registros, atualizando perfis e ampliando a capacidade preditiva do ecossistema de rastreamento ao longo do tempo.

4.1. Volume e Diversidade de Domínios Terceiros

A navegação em websites populares no contexto brasileiro revelou a presença sistemática de múltiplos domínios externos acionados durante uma única sessão de acesso. Mesmo em interações simples, como a visita a uma página inicial ou a realização de uma busca, observou-se a ativação de diversas conexões com serviços não diretamente relacionados ao conteúdo principal acessado pelo usuário.

Em média, cada sessão analisada resultou no contato com múltiplas entidades terceiras, incluindo redes de publicidade, plataformas de analytics, serviços de resolução de identidade e sistemas de recomendação de conteúdo. Esses domínios atuam como intermediários na coleta, processamento e redistribuição de dados comportamentais, configurando um ecossistema técnico distribuído. A Tabela 1 resume as principais métricas observadas durante os experimentos.

Table 1. Resumo das métricas observadas nas sessões de navegação

Métrica	Valor Observado
Domínios terceiros acionados por sessão	10-20
Categorias distintas de serviços identificadas	4-7
Transmissões médias por sessão	400-1.200
Requisições associadas a rastreadores	50%-70%

4.2. Análise dos Indicadores Observados

Os valores apresentados na Tabela 1 permitem extrair alguns insights relevantes sobre a estrutura da navegação web contemporânea no contexto analisado. Ainda que os interva-

los observados variem conforme o tipo de website e a dinâmica específica de cada sessão, os números indicam padrões estruturais consistentes.

O acionamento de 10 a 20 domínios terceiros por sessão sugere que a interação do usuário com um único website não se limita ao domínio principal acessado. Ao contrário, cada carregamento de página ativa um conjunto de comunicações paralelas com serviços externos, configurando uma arquitetura distribuída de dependências técnicas. Essa multiplicidade de domínios evidencia que a experiência de navegação é mediada por uma rede de atores interconectados, muitos dos quais não são diretamente visíveis ao usuário.

A identificação de 4 a 7 categorias distintas de serviços indica que o rastreamento comportamental não se restringe a uma única função, como publicidade. As categorias observadas incluem analytics, publicidade programática, sincronização de identificadores, recomendação de conteúdo e outras camadas funcionais. Isso demonstra que a coleta de dados integra um ecossistema multifuncional, no qual diferentes tipos de serviços contribuem para a construção e operacionalização de perfis comportamentais.

O volume de 400 a 1200 transmissões médias por sessão revela a intensidade técnica das interações subjacentes à navegação. Mesmo sessões breves podem gerar centenas de requisições HTTP(S), refletindo tanto a complexidade das páginas modernas quanto a presença de scripts e bibliotecas que operam de forma assíncrona. Esse elevado número de transmissões amplia a superfície de circulação de dados e aumenta as oportunidades de coleta e correlação.

Particularmente significativo é o fato de que entre 50% e 70% das requisições estejam associadas a rastreadores. Esse percentual indica que uma parcela substancial do tráfego gerado durante a navegação não está diretamente vinculada à entrega de conteúdo, mas sim a mecanismos de monitoramento, mensuração e segmentação. Tal proporção reforça a caracterização da infraestrutura observada como orientada à coleta e processamento de dados comportamentais de maneira sistemática.

Em conjunto, os indicadores sugerem que o rastreamento comportamental constitui um componente estrutural da arquitetura web contemporânea, e não um mecanismo periférico ou ocasional. A presença consistente de múltiplos domínios terceiros, categorias funcionais diversas e elevada proporção de requisições associadas a rastreadores evidencia a consolidação de um ecossistema técnico voltado à extração e operacionalização de dados de navegação em larga escala.

4.3. Categorias de Serviços Identificados

A análise dos domínios acionados permitiu classificar os serviços observados em diferentes categorias funcionais:

- Redes de publicidade programática
- Plataformas de analytics e monitoramento de desempenho
- Serviços de sincronização e resolução de identidade
- Sistemas de recomendação de conteúdo
- Infraestruturas de leilão em tempo real

Essa diversidade funcional demonstra que a coleta de dados não atende apenas à exibição de anúncios, mas integra um ecossistema mais amplo voltado à modelagem comportamental contínua.

4.4. Persistência e Correlação Entre Sessões

Observou-se também a recorrência de determinados domínios ao longo de diferentes sessões de navegação, indicando persistência de mecanismos de rastreamento. Mesmo após a transição entre websites distintos, parte das entidades terceiras permaneceu ativa, sugerindo correlação interdomínios.

Esse comportamento indica a existência de mecanismos de acompanhamento transversal que permitem consolidar dados provenientes de múltiplas fontes, ampliando a capacidade de construção de perfis comportamentais agregados.

4.5. Tipos de Dados Transmitidos

A inspeção dos pacotes interceptados revelou a transmissão de diferentes tipos de informações, incluindo:

- Identificadores únicos associados ao navegador ou dispositivo
- Parâmetros relacionados a páginas visitadas
- Indicadores temporais e geográficos
- Dados relacionados a interações específicas

Embora nem todos os conteúdos estejam explicitamente legíveis devido à criptografia parcial e codificação de parâmetros, foi possível identificar padrões consistentes de envio de identificadores persistentes, sugerindo vinculação contínua entre sessões.

4.6. Encadeamento de Fluxos e Propagação de Dados

A análise temporal das requisições demonstrou que determinadas chamadas a domínios terceiros geram subsequentes comunicações com novos serviços, configurando cadeias de propagação de dados. Esse encadeamento evidencia que a coleta não ocorre de forma estática, mas dinâmica e reativa, ampliando progressivamente o alcance dos fluxos informacionais.

Esse padrão reforça a caracterização da infraestrutura observada como um sistema distribuído e interdependente, no qual múltiplos atores participam da circulação de dados comportamentais em tempo real.

5. Implicações Sociotécnicas e Regulatórias

Os resultados empíricos obtidos neste estudo indicam que a coleta de dados comportamentais na web não ocorre como um conjunto de eventos isolados, mas como parte de uma infraestrutura distribuída composta por múltiplos atores interconectados. Essa arquitetura técnica apresenta propriedades sistêmicas relevantes, incluindo persistência de rastreamento entre domínios, correlação de identidades entre dispositivos e integração com mercados automatizados de publicidade em tempo real. Tais características configuram um ambiente no qual dados pessoais deixam de ser meros subprodutos de interação digital e passam a constituir insumos estratégicos para sistemas preditivos e mecanismos de segmentação algorítmica.

Do ponto de vista sociotécnico, a existência dessa infraestrutura altera a dinâmica informacional entre usuários e plataformas digitais, produzindo assimetrias estruturais de conhecimento e poder. Enquanto organizações operadoras dessas redes possuem visibilidade detalhada sobre comportamentos, preferências e padrões de interação, indivíduos

tipicamente desconhecem a extensão das cadeias de compartilhamento de seus dados e não dispõem de meios efetivos para auditar ou controlar esses fluxos. Essa assimetria informacional reduz a capacidade de agência dos usuários e dificulta a formação de decisões informadas sobre sua própria exposição digital.

Outro aspecto relevante refere-se ao caráter cumulativo dos perfis comportamentais gerados por esses sistemas. Diferentemente de registros isolados de atividade, os dados coletados são agregados ao longo do tempo e correlacionados entre múltiplas fontes, possibilitando a construção de modelos preditivos capazes de inferir características pessoais, interesses e predisposições. Esse processo de inferência amplia o alcance da coleta original, pois não se limita a registrar comportamentos observados, mas passa a produzir estimativas probabilísticas sobre atributos não explicitamente fornecidos pelos usuários.

Essa capacidade de modelagem comportamental cria condições técnicas para formas avançadas de segmentação informacional. Conteúdos, mensagens e estímulos digitais podem ser distribuídos seletivamente com base em perfis individuais, o que permite adaptar comunicações a características cognitivas, emocionais ou ideológicas específicas. Embora tais mecanismos sejam amplamente explorados para fins comerciais, a mesma infraestrutura pode ser utilizada para direcionamento político, amplificação de narrativas específicas e disseminação estratégica de desinformação. Nesse contexto, a personalização deixa de ser apenas um recurso de conveniência e passa a representar um mecanismo potencial de influência comportamental.

As implicações regulatórias desse cenário são igualmente relevantes. No Brasil, a Lei Geral de Proteção de Dados estabelece princípios que orientam o tratamento de dados pessoais, incluindo finalidade, adequação, necessidade, transparência e responsabilização. Entretanto, a natureza distribuída e frequentemente transnacional dos ecossistemas de rastreamento dificulta a aplicação efetiva desses princípios. Dados podem ser coletados em um domínio, processados em outro país e compartilhados com múltiplos intermediários cuja identificação nem sempre é transparente. Essa fragmentação operacional cria desafios para fiscalização, responsabilização jurídica e exercício de direitos por parte dos titulares de dados.

Além disso, mecanismos tradicionais de consentimento apresentam limitações estruturais diante da complexidade dessas infraestruturas. Interfaces de aceitação de cookies e políticas de privacidade frequentemente não refletem a totalidade dos fluxos de dados envolvidos, o que compromete a efetividade do consentimento informado. Mesmo quando formalmente obtido, o consentimento pode não representar uma escolha genuinamente livre, dada a assimetria de informação e a dependência funcional de serviços digitais essenciais.

Os achados deste estudo sugerem que a infraestrutura de rastreamento comportamental deve ser analisada não apenas como um conjunto de técnicas isoladas, mas como um componente estrutural do ambiente digital contemporâneo. Trata-se de uma camada invisível de mediação que opera continuamente na circulação de informações online, influenciando quais conteúdos são apresentados, como são priorizados e a quem são direcionados. Essa camada exerce papel relevante na organização do espaço informacional digital e, conseqüentemente, possui implicações para autonomia individual, formação de opinião e dinâmicas sociais mediadas por tecnologia.

Diante desse contexto, investigações empíricas sobre fluxos reais de dados tornam-se fundamentais para ampliar a transparência sobre o funcionamento dessas infraestruturas e subsidiar debates públicos informados. Evidências observacionais permitem deslocar a discussão do plano exclusivamente teórico para uma análise baseada em operações concretas, contribuindo para o desenvolvimento de políticas, mecanismos de governança e instrumentos de supervisão mais adequados à complexidade técnica do ecossistema digital atual.

6. Conclusões e Trabalhos Futuros

Este estudo investigou empiricamente a infraestrutura de rastreamento comportamental que sustenta o ecossistema contemporâneo de publicidade digital, analisando fluxos reais de dados gerados durante interações de navegação em ambiente brasileiro. Os resultados demonstraram que mesmo sessões simples de acesso a websites ativam cadeias extensas de comunicação com múltiplos domínios terceiros, evidenciando a existência de uma arquitetura distribuída de coleta, correlação e redistribuição de informações comportamentais.

As análises mostraram que esses fluxos não constituem eventos isolados, mas sim processos sistemáticos que envolvem diferentes categorias de serviços, incluindo redes de publicidade, plataformas de analytics, sistemas de identificação e infraestruturas de leilão em tempo real. Essa estrutura técnica permite a construção contínua de perfis comportamentais agregados, ampliando a capacidade de segmentação e direcionamento informacional. Do ponto de vista científico, o trabalho contribui ao fornecer evidência empírica direta do funcionamento concreto dessas infraestruturas em contexto real de uso, complementando a literatura predominantemente conceitual sobre o tema. Além disso, a caracterização observacional dos fluxos de dados permite compreender o rastreamento comportamental como um sistema sociotécnico distribuído, no qual múltiplos atores participam da circulação e operacionalização de informações pessoais.

As implicações sociais e regulatórias desses achados são significativas. A existência de mecanismos persistentes de identificação e correlação entre sessões sugere a presença de assimetrias informacionais estruturais entre usuários e plataformas digitais, com potenciais efeitos sobre autonomia, privacidade e formação de preferências. No contexto brasileiro, tais resultados indicam a importância de abordagens regulatórias que considerem a natureza interconectada e transnacional dessas infraestruturas, bem como a necessidade de mecanismos de transparência mais adequados à complexidade técnica envolvida.

Como trabalhos futuros, pretende-se expandir a análise empírica para diferentes perfis de navegação, dispositivos e contextos de uso, bem como investigar o papel dessas infraestruturas na personalização de conteúdo informacional e na disseminação de mensagens direcionadas. Estudos adicionais também podem explorar metodologias automatizadas para mapeamento contínuo de fluxos de dados e avaliação comparativa entre ambientes regulatórios distintos.

Em síntese, compreender empiricamente a infraestrutura de rastreamento comportamental constitui passo fundamental para avaliar seus efeitos sistêmicos e subsidiar discussões informadas sobre governança de dados e arquitetura do ambiente digital contemporâneo.

References

- Acar, G. et al. (2014). The web never forgets: Persistent tracking mechanisms in the wild. In *ACM CCS*.
- Andrew, J. and Baker, M. (2021). The general data protection regulation in the age of surveillance capitalism. *Journal of Business Ethics*, 168.
- Couldry, N. and Mejias, U. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.
- Englehardt, S. and Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. In *ACM CCS*.
- Iqbal, U. et al. (2021). The web is watching: A comprehensive review of web tracking. *ACM Computing Surveys*.
- Marotta, L. et al. (2022). Rtb ecosystem analysis: Data flows in programmatic advertising. *IEEE Security and Privacy*.
- Pasquale, F. (2015). *The Black Box Society*. Harvard University Press.
- Wu, Y., Bice, S., Edwards, W. K., and Das, S. (2023). The slow violence of surveillance capitalism. In *ACM FAccT*.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*. PublicAffairs.
- Zuboff, S. (2022). Surveillance capitalism or democracy? the death match of institutional orders and the politics of knowledge. *Organization Theory*, 3(3).