

# CIBERSEGURIDAD Y PRIVACIDAD EN LA WEB: EVALUACIÓN INTELIGENTE (IA) DEL CUMPLIMIENTO NORMATIVO ANTE TECNOLOGÍAS DE SEGUIMIENTO

## COOKIES

### INVESTIGADORES

David Martínez<sup>1</sup>, Aniol Molero<sup>1</sup>, Eusebi Calle<sup>1</sup>, Dolors Canals<sup>2</sup>, Albert Jové<sup>3</sup>

### AFILIACIONES

<sup>1</sup>Institut d'Informàtica i Aplicacions, Universitat de Girona

<sup>2</sup>Departament de Dret Públic, Universitat de Girona

<sup>3</sup>Universitat Oberta de Catalunya

### INTRODUCCIÓN

**PANORAMA:** Generalización de las tecnologías de seguimiento y rastreo de personas vía web, entre ellas las "cookies".

**HIPOTESIS DE PARTIDA:** El cumplimiento de la normativa de privacidad sigue siendo escaso en el entorno digital.

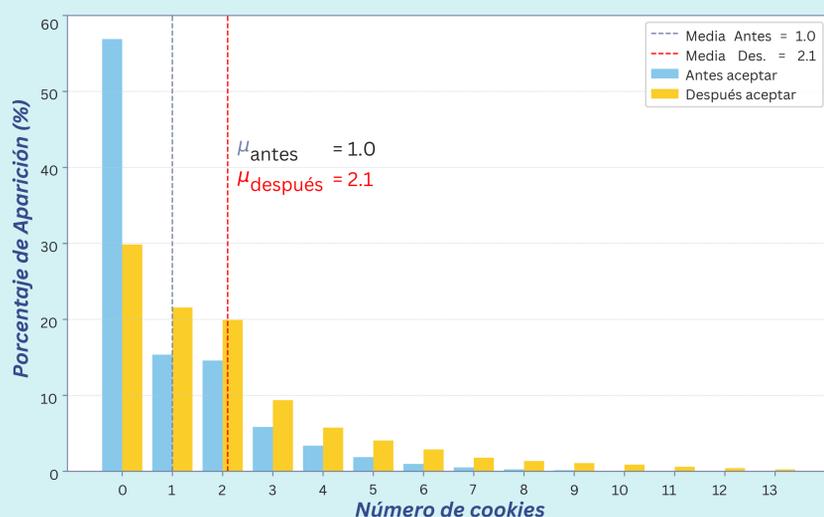
### OBJETIVO

Desarrollar **métricas confiables** para la **evaluación inteligente (algoritmos y herramientas basadas en IA)** de la integridad de los sitios web en términos de privacidad, diseñadas para detectar de forma masiva **cómo los sitios web obtienen el consentimiento de las personas usuarias** y qué tecnologías de rastreo emplean, incluidas aquellas ocultas e ilegales.

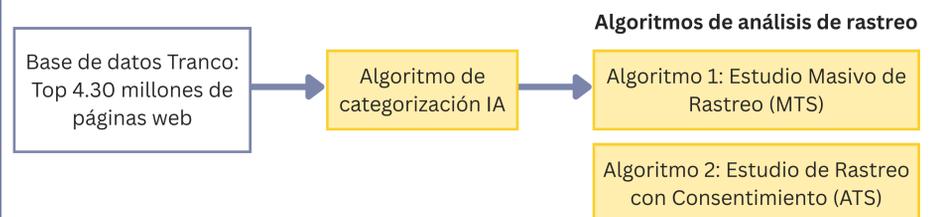
### METODOLOGÍA

Se utiliza el software *Website Evidence Collector* (WEC), publicado por el *European Data Protection Supervisor* (EDPS).

### DIFERENCIAS EN COOKIES DE RASTREO ANTES Y DESPUÉS DEL CONSENTIMIENTO



El promedio de cookies de rastreo por página web pasa de 1 antes de aceptar a 2.1 después, **más del doble**.



### RESULTADOS

**Modelo de IA** multietiqueta de **22 categorías** para la categorización de sitios web, alcanzando una **precisión del 97%**.

Se ha analizado el top **un millón de páginas** más visitadas (Tranco) via el algoritmo **Estudio Masivo de Rastreo (MTS)**, y **150.000 páginas** via el algoritmo **Estudio de Rastreo con Consentimiento (ATS)**.

- **Frecuencia de rastreo:** El **43%** de páginas web utilizan *cookies* de rastreo y el **63%** píxeles de seguimiento **sin el consentimiento de los usuarios**.
- **Concentración de poder:** Las 10 principales empresas propietarias del rastreo web son responsables del **81%** de las **infracciones**.

### IMPACTO CIENTÍFICO Y EN LA SOCIEDAD

A parte de los **resultados de la investigación** ya publicados [1] y su impacto en decenas de medios de comunicación nacionales [2], algunos se encuentran en fase de revisión para ser **publicados en revistas de gran prestigio internacional** [3].

### CONCLUSIONES

Esta investigación presenta **nuevos algoritmos abiertos, métodos** y un **modelo de categorización de IA** para realizar **análisis masivos y completamente automáticos** sobre el rastreo web y el cumplimiento normativo de *cookies* en relación con la obtención del **consentimiento válido** del usuario. Además, sirve como **punto de partida** para realizar **análisis masivos periódicos** del rastreo web en Internet, lo cual sería valioso para determinar si los niveles de rastreo cambian con el tiempo y **evaluar la efectividad de las regulaciones de privacidad**.

[1] Martínez, D., Calle, E., Jové, A., & Pérez-Solà, C. (2022). Web-tracking compliance: websites' level of confidence in the use of information-gathering technologies. *Computers & Security*, 122.

[2] Law&Trends (2023). Las webs más visitadas en España rastrean ilegalmente usuarios, según estudio. Disponible en: <https://www.lawandtrends.com/noticias/tic/las-webs-mas-visitadas-en-espana-rastrean-ilegalmente-usuarios-segun-estudio-1.html#gsc.tab=0>

[3] Martínez, D., Molero, A., Calle, E., Canals, D., Jové, A. (2025). Large-Scale Web Tracking and Cookie Compliance: Evaluating One Million Websites under GDPR with AI Categorization. *Journal of Network and Computer Applications* (en revisión).