

Manual de instalación y uso

AI Crawler Tracker Pro — WordPress plugin para identificar, analizar y controlar bots de IA

1) Requisitos

- **WordPress:** 5.5 o superior (recomendado 6.x).
- **PHP:** 7.2 o superior (compatible con 8.x).
- **Permisos:** rol con capacidad `manage_options` (administrador).
- **Cron de WP:** para tareas programadas (retención y correo diario).

Nota: El plugin **no borra datos automáticamente al desinstalarse** (por seguridad). Puedes activarlo/desactivarlo sin perder registros.

2) Instalación

A) Desde el ZIP

1. Entra en tu WordPress → **Plugins** → **Añadir nuevo** → **Subir plugin**.
2. Sube el archivo `ai-crawler-tracker-pro_v1.0.0.zip`.
3. Haz clic en **Instalar ahora** y luego **Activar**.

B) Vía FTP/SFTP (opcional)

1. Descomprime el ZIP.
2. Sube la carpeta `ai-crawler-tracker-pro/` a `/wp-content/plugins/`.
3. Activa el plugin desde **Plugins** en el admin.

Al activarse, el plugin crea su tabla (`{prefix}ai_crawler_logs`) y programa las tareas periódicas (retención diaria y, si lo habilitas, el email resumen).

3) Primeros pasos (configuración recomendada en 5 minutos)

Ve a **AI Crawlers** → **Ajustes** y revisa:

1. **Privacidad**
 - *Anonimizar IP:* activado (recomendado).
 - *Hashear IP:* opcional si necesitas máxima privacidad.
 - *Días de retención:* p.ej., 180.

2. Verificación DNS

- *Verificar bots por DNS*: activado (recomendado).

Confirma que GPTBot, Google-Extended, Perplexity, etc. vienen realmente de dominios oficiales.

3. Bloqueo

- *Bloquear si no verificado*: opcional.
- *Bots bloqueados*: marca los que no quieras que rastreen.
- *Bloquear por rutas*: una por línea (p. ej. `/wp-json/` o zonas sensibles).

4. Registro

- *Omitir estáticos*: actívalo si quieres centrarte sólo en **contenido real** (no CSS/JS/medios).
- *Días mostrados en dashboard*: por ejemplo, 30.

5. Alertas por email

- *Enviar email diario*: marca si quieres recibir cada mañana (08:00 hora WP) el **top de contenido** visitado por bots.

Guarda con “**Guardar ajustes**”.

4) ¿Dónde lo encuentro?

Tras activarlo, verás un nuevo menú **AI Crawlers** con tres secciones:

- **AI Crawlers** (listado “raw” de accesos)
 - **Dashboard** (resúmenes, tops, filtros y CSV)
 - **Ajustes** (configuración y bloqueos)
 - **Sobre mí** (autor, contacto y versión)
-

5) Uso diario

5.1 Listado general (AI Crawlers)

Ruta: **AI Crawlers** → **AI Crawlers**

- **Filtros**: por *bot* y por *estado* (permitido/bloqueado).
- **Ordenación**: por **Fecha** o **Ruta** (ASC/DESC).
- **Paginación** y **por página** configurable.
- **Exportar CSV (raw)**: descarga todos los campos (ideal para auditorías o BI).

Campos clave del log

- Fecha/Hora (hora local de WordPress)
- Bot (y host verificado, si aplica)
- Método y URL (con *query string*)

- Código de respuesta y **bytes servidos**
 - Referrer (si lo hay)
 - IP (anonimizada/hasheada según ajustes), *user-agent* y cabeceras seleccionadas
 - Estado (**Permitido/Bloqueado**)
-

5.2 Dashboard

Ruta: **AI Crawlers** → **Dashboard**

- **Gráfico diario** de hits en el rango elegido.
- **Top Bots (detallado)**
 - Columnas: *Bot*, **Veces accedido**, **Bytes servidos**, *Hits bloqueados*.
 - Ordenable, paginable y exportable a CSV.
- **Top Rutas (contenido real)**
 - Excluye estáticos comunes (CSS/JS/imágenes/medios).
 - Muestra **rutas más consultadas**, con **vecas accedido**, **bytes servidos** y *hits bloqueados*.
 - Ordenable, paginable y exportable a CSV.

Si en tu versión aparece el selector “**Rutas por bot**”, podrás filtrar **Top Rutas** para ver **qué bot** accedió a **qué URL** y cuántas veces, además de exportar **CSV por bot**.

5.3 Ajustes

Ruta: **AI Crawlers** → **Ajustes**

1. **Privacidad**
 - *Anonimizar IP*: oculta el último tramo (IPv4) o bloques (IPv6).
 - *Hashear IP*: guarda hash en lugar de IP en claro (añade privacidad a informes).
 - *Retención*: purga automática de registros antiguos.
 - *Salt*: cadena usada para hash de IP (guárdala en un lugar seguro).
2. **Verificación DNS**
 - Aplica *reverse DNS + forward confirmed*: se asegura de que la IP resuelve en un host permitido y que ese host apunta de vuelta a la IP.
3. **Bloqueo**
 - *Bloquear si no verificado*: devuelve **403** a bots que no superen la verificación.
 - *Bots bloqueados*: casillas para vetar por nombre (p. ej., Perplexity, GPTBot...).
 - *Bloquear por rutas*: una por línea (prefijos). Útil para áreas sensibles.
4. **Registro**
 - *Omitir estáticos*: centra el análisis en **contenido**.
 - *Días en dashboard*: controla el rango por defecto de informes.
5. **Alertas por email**

- Activa/desactiva el **resumen diario** de rutas más visitadas por bots de IA (enviado a `admin_email`).

5.4 Sobre mí

Ruta: **AI Crawlers** → **Sobre mí**

- Autor: **Arturo Yelmo** — **SEO y Desarrollo Web**
- Web: **arturoyelmo.com**
- Correo: help@arturoyelmo.com
- Versión: **1.0.0**
- El plugin está **firmado** en código y descripción.

6) Exportaciones a CSV

- **Raw** (desde “AI Crawlers”): todos los campos de cada hit.
- **Top Bots** (desde Dashboard): `bot_name`, `veces_accedido`, `bytes_servidos`, `hits_bloqueados`.
- **Top Rutas** (desde Dashboard): `ruta`, `veces_accedido`, `bytes_servidos`, `hits_bloqueados`.
- **Rutas por bot** (*si aparece en tu versión*): `bot`, `ruta`, `veces_accedido`, `bytes_servidos`, `hits_bloqueados`.

Codificación **UTF-8**. Abrible en Excel/Sheets; si ves caracteres extraños, importa el fichero indicando UTF-8.

7) Bloqueo y robots.txt

- Si marcas **Bots bloqueados**, el plugin añadirá al `robots.txt` dinámico las directivas `User-agent: <UA> + Disallow: /` para esos agentes.
- **Bloqueo por verificación**: si activas “Bloquear si no verificado”, cualquier UA que no pase la prueba de DNS recibirá **403**.
- **Bloqueo por rutas**: deny por prefijo (una línea por regla).

Recuerda: `robots.txt` es **voluntario**; el bloqueo por 403 es efectivo incluso si el bot ignora robots.

8) Email diario (opcional)

- Se programa a **las 08:00** (hora local de WordPress).

- Muestra el **Top de contenido real** (no estáticos) visitado el **día anterior**.
- Para activarlo: **Ajustes** → **Alertas por email** → **Enviar email diario**.

Si no recibes el correo, revisa que WP-Cron se esté ejecutando (alojamientos con tráfico muy bajo o cron deshabilitado pueden requerir un **cron del sistema** apuntando a `wp-cron.php`).

9) Buenas prácticas y rendimiento

- Activa “**Omitir estáticos**” para reducir ruido y volumen de logs.
 - Ajusta la **retención** (p. ej., 90–180 días) según el tamaño de tu web.
 - Si usas CDN/WAF (Cloudflare, etc.), recuerda que pueden alterar cabeceras o IPs.
 - Considera configurar “**IP del cliente real**” si tu hosting lo requiere.
 - Evita bloqueos agresivos sin revisar: puedes empezar con **verificación DNS** y **bloqueo por rutas** críticas.
-

10) Solución de problemas

No aparecen datos

- Asegúrate de que hay tráfico de bots (muchos bots respetan `robots.txt`).
- Comprueba que **no** has marcado “Omitir estáticos” si sólo recibes accesos a recursos.
- Desactiva temporalmente cachés a nivel de servidor para probar.

Email diario no llega

- Verifica la dirección `admin_email`.
- Revisa SPAM.
- Comprueba que WP-Cron se ejecuta (desactívalo/actívalo para reprogramar; o llama manualmente a `wp-cron.php`).

Temo bloquear Googlebot

- No bloques “Googlebot” normal. El plugin separa **Google-Extended** (uso para IA). Revisa tus necesidades antes de marcar bloqueos.

Quiero borrar todos los datos al desinstalar

- Por defecto **no** se borran. Si sabes lo que haces, puedes editar el método `uninstall()` para eliminar la tabla y opciones.
-

11) Actualizaciones

- Sube el **nuevo ZIP** y WordPress sustituirá los archivos (o actualiza por FTP).
 - La tabla y los datos existentes se mantienen.
 - Revisa el **changelog** y re-guarda Ajustes si añadimos opciones nuevas.
-

12) Licencia y contribuciones

- **Licencia:** GPL (GPLv2 o posterior). Puedes **usar, estudiar, modificar y redistribuir** el plugin.
 - **Contribuciones:** estaré encantado de recibir **mejoras, ideas y parches**.
 - Contacto: help@arturovelmo.com — arturovelmo.com
-

13) (Avanzado) Extender firmas de bots

Puedes añadir o modificar firmas de bots mediante el filtro

`ai_crawler_tracker_bot_signatures`:

```
add_filter('ai_crawler_tracker_bot_signatures', function($sig) {  
    // Añadir un bot personalizado:  
    $sig['MiBotIA'] = ['/(?:MiBotIA)/i', ['.mibotia.com']];  
    return $sig;  
});
```

Formato: ['Nombre visible' => ['regex_del_user_agent',
['.dominio_permitido_1', '.dominio_permitido_2']]]

14) Preguntas frecuentes (FAQ)

¿Puede identificar a un bot que miente sobre su identidad?

Ayuda la **verificación DNS** (reverse + forward confirmed). Si miente o usa infraestructura ajena, puedes **bloquear por ruta** o por UA conocido.

¿Consume recursos?

El registro es ligero y las tablas del panel están **paginadas**. Además, puedes excluir estáticos.

¿Afecta al SEO?

No interfiere con Googlebot *clásico*. Tú decides bloquear **Google-Extended** y otros agentes de IA.

15) Contacto

- **Autor:** Arturo Yelmo — SEO y Desarrollo Web
 - **Web:** arturoyelmo.com
 - **Email:** help@arturoyelmo.com
-

Anexo: mapa rápido de menús

- **AI Crawlers** → **AI Crawlers**: listado completo + CSV (raw).
- **AI Crawlers** → **Dashboard**: gráficos + Top Bots + Top Rutas (+ *Rutas por bot* si está disponible).
- **AI Crawlers** → **Ajustes**: privacidad, verificación, bloqueos, registro y email diario.
- **AI Crawlers** → **Sobre mí**: autoría y versión.