Guide: Tracking LLM-Sourced Traffic in GA4

AI tools like ChatGPT, Perplexity, and Gemini are changing how people discover content. But LLM-driven visits often go unrecognized in standard analytics. This guide outlines how to track traffic from large language models in GA4 so you can better understand what content is being surfaced and how users engage.

1. Create a Custom Exploration or Acquisition Report

Choose one of the following approaches:

a. Exploration (recommended for deeper analysis):

Go to Explore → + Blank

- Name your exploration “LLM traffic analysis”
- Add the following dimensions:
  - session source/medium
  - page referrer (or) page path + query string
- Add these metrics:
  - sessions
  - engaged sessions
  - engagement rate
  - conversions, etc.
- Add rows to differentiate the data display:
  - session source/medium (or individually as session source and session medium)
  - page referrer
  - page path + query string

Sample sources:

- *bendyourmarketing.com*
- *digitnetix.com*
- *hocdigitalsolutions.co.uk*

b. Traffic acquisition report:

- Navigate to Reports → Acquisition → Traffic acquisition
- Click Add filter and apply a regex on session source / medium
    - Regex in this report is limited to 250 characters

Sample sources:

- *linkedin.com*
- *reddit.com*
- *grandcrudigital.com.au*

2. Apply a regex filter to capture all LLM referrers

Use the "matches regex" filter option with a comprehensive pattern. Here's a base regex to adapt:

```
(?i).*(chatgpt\.com|chat\.openai\.com|perplexity(\.ai)?|gemini\.google\.com|bard\.google\.com|copilot\.microsoft\.com|claude(\.ai)?|edgeservices\.bing\.com|deepseek|you\.com|poe\.com|qwenlm\.ai|huggingface\.co|doubao\.com|writesonic\.com|copy\.ai|bnngpt|nimble\.ai|aitastic\.app|iask\.ai|chat-gpt\.org).*$
```

- (?i) makes it case-insensitive
- Covers major and emerging LLM platforms
- Can be extended as new LLMs appear by inserting new string before the final end-parenthesis (ex. |openrouter\.ai|magical\.team|useblackbox\.io|ai-coustics\.com|chinchilla\.ai)

Sample domains:

- *chatgpt.com*
- *perplexity.ai*
- *claude.ai*

3. Add dimensions & metrics for insights (on Acquisition report)

- Dimensions:
    - session source/medium (identifies the LLM)
    - page path + query string or page referrer (shows which content is being surfaced)

- Metrics:
  - sessions
  - engaged sessions or engagement rate
  - average session duration
  - conversions or key events

Sample sources:

- *chatgpt.com*
- *perplexity.ai*
- *claude.ai*

4. Save, schedule & surface on dashboard

- Save your exploration or customization
- Add it to your GA4 Library or Dashboards for quick access
- Schedule recurring reviews to update the regex as new LLMs gain traction

5. (Optional) Enhance with GTM + custom events

To capture deeper interactions—such as chatbot clicks, prompt usage, or API-driven visits—leverage Google Tag Manager:

- Set up click triggers for outbound LLM links or chat widget activity
- Send a custom event (e.g., llm_click) with a custom dimension like llm_source
- Register in GA4 and analyze via Explorations or Looker Studio

Why This Matters

- ↗ **Visibility into AI-driven traffic:**
  LLM referrals may be miscategorized (e.g., direct, referral) if not explicitly filtered
- ↗ **Content optimization:**
  Discover which pages LLMs consistently reference—and enhance those assets
- ↗ **Attribution clarity:** Properly credit LLM-sourced visits and conversions, avoiding "unassigned" buckets

Quick Checklist

☑	Choose exploration vs. traffic report

☑	Apply regex filter for session source/medium
☑	Add relevant dimensions & metrics
☑	Save report & add to dashboard
☑	Implement GTM custom event tracking (optional)

What's Next

With this setup, you'll gain robust visibility into LLM-sourced visits—including which LLMs are driving traffic, how users engage, and what content wins in AI discovery. The system is future-ready: just update your regex as new models gain momentum.

With this setup, you'll gain clearer visibility into LLM-driven traffic and content performance. It's easy to maintain—just update the regex as new models emerge.

Contact Us.

✉ letstalk@amsive.com

↗ www.amsive.com