Residential Proxies as the Next Front in the AI Wars

Originally published Wednesday, September 16, 2026 at https://www.unite.ai/residential-proxies-as-the-next-front-in-the-ai-wars/

In order to stop AI's knowledge-base from stagnating or corrupting, frontier AI models need to keep ingesting vast quantities of current data from the internet.

In this respect, the major AI providers, and their avid competitors, are currently meeting two obstacles: a) the publishers, social networks, platforms and other sources of fresh data now realize what this access is worth, and are looking to monetize it; and/or b) domains are now getting so hammered by relentless AI scraper-bots looking to close the 'knowledge gap' that the backlash against this abuse is becoming quite extreme.

From the providers' perspective, one possible solution for the latter issue is to require that a user be registered with the domain, and logged in before they can access that sweet, sweet data. In this way, request volume can be evaluated against an authenticated user rather than an IP address (see *'IP Seeing You'*, below).

Of course, coincidentally, this scenario is a data-broker's dream, since it forces people to de-anonymize themselves.

This was one of the fundamental objections against the UK government's 2026 legislation requiring Brit-based users wishing to access 'adult'-rated site to identify and authenticate themselves through third-party platforms.

Here, the net result was deemed by many to be a CCP-style 'chilling effect' on previously-anonymous users, and the beginning of tacit mandatory online authentication in the UK, rather than the law's avowed intent to protect children from NSFW content.

So in a world where logins are the norm, if not mandatory, the State benefits from logistical or even forensic data, as well as a populace that's a little more circumspect when they're online; and the advertisers and monetizers are finally relieved of their decades-long battle to associate a user with a persistent ID, and exploit that user's data in a variety of ways, from targeting, to inclusion in statistical modeling, and more potential applications besides.

IP Seeing You

There are alternate methods of identifying a returning user than their IP address, such as various forms of browser fingerprinting, among other approaches.

But since there's no connection or access *at all* without an IP address, it can't easily be addressed by browser updates, or other software-based methods intended to protect the privacy or identity of a user. Therefore the IP address retains moderate credibility as a 'cheap' solution to controlling and monitoring user access.

Since the IPV4 address pool is long since exhausted, and its replacement struggles for widespread adoption, alternate IPs are at a premium; therefore VPN providers tend not to let users rotate their IP address much.

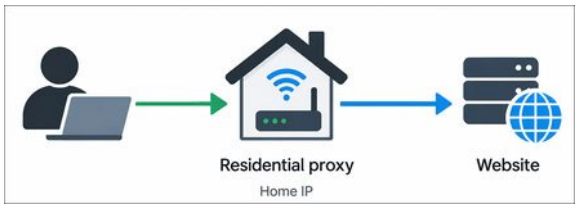
In most cases, the domestic VPN user picks a country of origin that they want to be seen as browsing from, and then they get a pretty stale IP address – one that has almost certainly been entered into the various blacklists that portals use to keep out non-national users, or for other reasons.

The third-party identity provider platform Yoti (used in the UK's age verification scheme) notes:

> *'[Many] major online platforms already block known VPN traffic. Services like BBC iPlayer, Netflix, and others use established techniques to detect and prevent access from commonly used or suspicious VPN IP addresses.*
>
> *'These include maintaining updated blocklists and monitoring for high-volume traffic from known IP pools. Even as some VPN providers offer more advanced or harder-to-detect services, these tend to be expensive, creating another layer of friction and financial deterrent – especially for children and teens.'*

About that 'expense': if you're willing to spend considerably more cash than a standard VPN subscription, a residential proxy side-steps nearly all these obstacles – at least for the moment.

A residential proxy is the IP address of a broadband user, just like you or me, who is either selling their 'safe'-looking location to a commercial company in order that the company's customers can piggy-back on that IP to avoid blacklists; is inadvertently providing this service, unpaid, because it came with an app for an appliance; or is providing the service because they've been hacked into a botnet.



Using a residential proxy, all your traffic is routed through a domestic, consumer-level IP address, making blocks and interventions less likely to occur.

Most resprox 'hosts' sell access via a wide range of apps, led at the moment by the venerable Honeygain, with earnings averaging between $.10 and $.80 USD per GB of traffic.

In any case, if one IP-renter on a resprox account gets IP-blocked by a provider that they are targeting, they can switch to a new IP that may have a whole bunch of historical issues from previous tenants but *not that user's particular issue*.

In other words, previous actions on that IP address may have caused it to be blocked by Reddit or other social networks, or any other number of providers; but if the new tenant's target is not among these blocks, they can start all over again with that target. If it is, the user can just try a different residential IP.

As for the host, their own inevitably and increasingly polluted IP address can often be renewed by interrupting the connection for a while (and in most cases will be rotated automatically, at intervals); additionally, the costs and potential consequences of permanently blacklisting any single IP address mean that most such blocks are eventually lifted.

The Age of Age-Checks

With age-gating a fact of life in the UK, imminently in the EU and Australia, and intermittently around the US, VPN usage – the primary means of avoiding national online age-gates – is already coming into focus as a target for legislation, state oversight, or even prohibition – as is the case in countries such as China and the Russian Federation.

Two years ago, most of us probably did not know or care that VPNs existed (except that our company's IT department may have installed one on our laptops, for secure access to a corporate network). Now, at least in the UK, EU, and Australia, it's a technology with far greater public awareness.

The *next* such 'open secret' that seems likely to surface – and to become a target for legislation, committees and oversight – over the course of the coming twelve months is the residential proxy.

The growing market in web-scraping services usually offers residential proxy access as a 'premium' service. Bright Data offers a 798GB bandwidth residential proxy package at an eye-watering $2k a month; FireCrawl charges $8-15 USD per GB; and Apify around $8 USD.

Many of the other major providers obfuscate their charges, with users directed to 'contact sales' – and effectively, it could be argued that all the advertised lower tiers are merely upsell trailers for that discussion – because data-hungry AI companies have 'life-changing' money to throw at the problem.

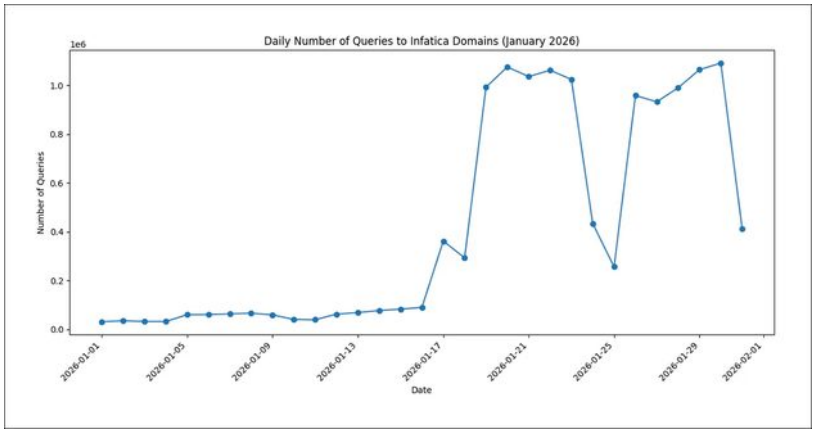
Above the Radar..?

Though residential proxies are far from an illicit activity at the moment, like all technologies that facilitate anonymization, they are beginning to come under fire as a favored tool for cybercriminals, notably in Japan.

Any threat to residential proxies would seem to have potential implications for the AI companies increasingly making use of them: a recent (June 2026) report from cybersecurity company Infoblox posits that AI has become a major driver of residential proxy demand.

Analyzing DNS traffic across its customer base, Infoblox found monthly queries to residential proxy domains rising from roughly 400 billion to more than 500 billion between January 2025 and April 2026 – an increase of about 25%:

CONTINUED ON NEXT PAGE



Spikes in the total number of residential proxy-related traffic points across the course of 2026 . Source

The company stated at the time:

> *'There are likely several explanations for this: certainly, the rise in AI-related training, which often requires scraping websites, is a major driver of residential proxy demand. Residential proxies bypass many anti-scraping measures, as the traffic appears to be coming from the devices of real people.'*

This is a similar upward trend to prior surges of interest around VPNs as age-gating laws emerged, and is beginning to take the practice out of the 'corporate geek' domain and into potential public focus.

Cloudflare characterizes the new generation of AI-focused web-scrapers as 'designed to blend in', observing *'They can rotate IP addresses through residential proxies, generate human-like user agents, and mimic plausible browsing patterns'*. The company also reports that crawling for the purpose of obtaining training data represents 80% of all AI bot activity in this period.

A June 2026 report from Include Security revealed the extent to which scraping behemoth Bright Data (which currently boasts the availability of 400m+ IPs across 195 countries) incorporated an unauthenticated public endpoint into a range of consumer devices, to provide residential proxy services through the installation of apps (i.e., third-party developers included Bright Data into their apps in order to monetize them) .

And in October of 2025, KrebsOnSecurity reported that the 'record-smashing distributed denial-of-service (DDoS)' botnet Aisuru had moved tactics from traditional DDoS to residential proxies:

> *'[A] glut of proxies from Aisuru and other sources is fueling large-scale data harvesting efforts tied to various artificial intelligence (AI) projects, helping content scrapers evade detection by routing their traffic through residential connections that appear to be regular Internet users.'*

Since fraud and other forms of cybercrime activity are increasingly being associated with residential proxy use, this may be what eventually takes it off the market; and because the service is substantially provisioned through apps and SDKs distributed via major app platforms, a requirement for hosting platforms (iOS, etc.) to ban or modify these so that resprox hosting is not among the offered services, would greatly undermine the provisioning of this class of service.

Targeting standalone portals/apps as well could reduce the entire residential proxy scenario to skeleton coverage – and presumably raise the prices considerably.

The question is, will very pro-AI nation states – such as the US – be willing to cut AI companies off in a way that will almost certainly *not* affect Chinese scrapers, which (presumably) will operate according to state need rather than legislative guidelines? The current mood in the US at least, suggests not.

Conclusion

Residential proxies are essentially VPNs that actually still work, because the agile IP addresses offered are not blacklisted and stale; and there is notable evidence that AI companies are resorting to their use, as fortifications against AI web-scrapers become more widespread and more elaborate.

Some of the furious pace of AI-based web-scraping – including up to 39,000 requests a minute to a single domain, and the assimilation and subsequent burning of books – has perhaps been aimed at curating the sum of human knowledge to date – a 'one-time catch-up', if you like.

However, the future lies in gaining and retaining access to information generated *from this point on* – social media posts, academic studies, timetables, blogs, corporate sites, and millions of other sources of ongoing data output – the continuity of the collective knowledge and thought of the human race.

Therefore one wonders to what heights a the cost of a GB of residential proxy traffic can climb before even the well-heeled frontier AI companies balk, or else decide to scrape more selectively, finally inside the confines of some kind of budgetary restriction.

That price will depend on the extent to which the most attractive data platforms either negotiate access or successfully repel boarders.

One final thought – if I were running a trillion-dollar AI company increasingly dependent on an external residential proxy-based scraping service, I would consider it a rational economy either to actually buy and operate the best-placed such company, or else establish my own resprox network from scratch (presumably via a maze of shell companies).

First published Wednesday, September 16, 2026

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai