

WHITE PAPER

# HOW TO ESTIMATE IMPRESSIONS ACCURATELY FOR MEDIA MENTIONS

---

This white paper is written by Magnus Stahl, data scientist at Hypefactors.

It's about why the best possible estimation of impressions for online media mentions is important and how Hypefactors' work with applied AI improves the accuracy of this number.

---

# How to estimate impressions accurately for media mentions

**Impressions is an important quantity in modern media monitoring. Efficient and accurate solutions should use best-in-class data and applied AI to estimate impressions.**

The amount of information on the web is growing, and quickly. Organizations that wish to monitor their public relations are presented with new challenges. At the same time, the rate at which conversations spread - both globally and digitally - simply requires that organizations can identify and respond to vast amounts of mentions quickly, rationally and coherently.

To solve this challenge, it is absolutely vital that the people with the skills and knowledge to analyze how their organizations should react to mentions at scale, have access to data that is:

1. Gathered in a fully automatic way
2. Processed real-time and filtered by intelligent algorithms
3. Quantified by a state-of-the-art system that can determine the impact of a mention

At Hypefactors we recognize that a crucial part of media monitoring is the ability to distill the vast amount of information produced on the web, and aggregate it into simple but powerful metrics. The choice of metric is of critical importance since it shapes and defines the message that can be interpreted and communicated to stakeholders in other parts of the organization. At the same time, it has to capture the increasingly complex landscape of earned media. Simply put, we want to tell you what people say about your brand, but also to who and how many.

It is the “how many” part which we will review in this whitepaper. This quantity is called impressions. Impressions describes how many individual views a mention might have. It’s not a count of unique individuals. Like a paid media ad, exposing the same user multiple times will count for multiple impressions. We will describe how Hypefactors quantifies impressions on the general web, and how AI can help estimate how many have interacted or seen a mention

Impressions describes how many individual views a mention might have.

about your brand. It's essential simply because better impression estimation leads to better calculations of what a mention can be worth for your brand.

## General Web Traffic Estimation

On the web, figuring out how many might have seen a mention about your brand is a daunting task. Not only do we at Hypefactors monitor millions of data sources every day; we also have to quickly and efficiently be able to add new sources and intelligently estimate how many people might interact or view a specific mention on a source.

Normally, knowing the number of impressions for a website is easy if you have access to an analytics tool that's deployed directly on the website. The problem is that you don't have this access for all the relevant media sites where the mentions about your brand are generated, and thus we have to be able to look at any website and estimate the traffic.

To make matters worse, the Internet is fundamentally decentralized, so there's no central control points which can be surveilled in order to monitor and quantify all web traffic. To estimate who visits any website, we have to rely on aggregating data from many different sources such as: panel data, ISP data, measurement data from apps and sites and other public sources of data. And we also have to deduce the complicated interplay between the various sources of data in order to identify unique users and their behaviours.

**This is why we at Hypefactors have developed an impression-estimation strategy based on applied AI and state-of-the-art data.**

## Impressions with applied AI

Quickly summarized, AI (or in this case, the subfield called machine learning) is the discipline of creating mathematical models which can estimate functions from data by learning patterns. AI can be leveraged in a situation where an unknown quantity (impressions) must be estimated from a variety of data sources such as: traffic data, website structure, information about a countries' internet behaviour, user statistics and other variables. By inputting all these variables together, we can try to find statistical relationships between the data and determine correlations between articles where we already know the impression count, and how the input data looked for those articles. This technique is known as regression analysis and is one of the techniques which we utilize to improve the estimation of impressions.

The Internet is fundamentally decentralized, so there's no central control points which can be surveilled in order to monitor and quantify all web traffic.

AI can be leveraged in a situation where an unknown quantity (impressions) must be estimated from a variety of data sources

## Getting the Best Data

There's no good AI without good data. With the current knowledge in AI, it's simply impossible to estimate functions or numbers efficiently without first getting hold of quality sources of data from which we can estimate. Which is why we at Hypefactors have teamed up with the leading industry provider in traffic estimation, SimilarWeb. SimilarWeb collects traffic data from hundreds of millions of devices everyday and serves it in aggregated form to a wide array of companies, including market leaders such as eBay, Google and Microsoft. SimilarWeb provides us with the large raw quantity of data we need from visitors around the world to gather a coherent image of the worldwide traffic on the web.

## Estimating Impressions from Web Traffic

We have developed a proprietary analysis algorithm which relies on extensive data collection and AI-powered filtering to estimate impressions even more accurately.

There is of course a difference between how many people visit a site, and how many impressions a specific mention on that site will have. To further zoom in on this quantity, we have developed a proprietary analysis algorithm which relies on extensive data collection and AI-powered filtering to estimate impressions even more accurately. To better explain this, imagine having a review of a particular product published on a blog X with 50.000 unique visitors every month, and on a news site Y with the same amount of visitors. Immediately a number of questions arise:

- What are the differences between X and Y?
- How does users interact with articles on each page?
- Which articles are there?
- How often do they change?
- What type of content is presented?

For example, the probability that some users might not even see the mention on site Y is much higher since the rate at which articles produced is higher. And not every visit to a page might even be human - a certain quantity of traffic to each site will be entirely automated!

It is for these reasons that our algorithm also continuously monitors and analyzes data sources on top of the data we receive from SimilarWeb. This allows us to not only estimate how many people might have seen a mention, but also take into account the probability that a user clicked on the article also.

One of the unique outcomes of this algorithm is that we can accurately classify media outlets represented on the web into a media type: news sites, which tend to publish hourly or daily, magazine that publish regularly but necessary daily, and blogs, that usually publish less frequently.

## The Numbers

Now we'll move on to describing the impression-related metrics we enrich mentions of your brand with. As mentioned above, these metrics are a part of the total calculation of the Hypefactors Value, quantifying the financial value of media coverage, but they are also shown to help users get a sense of the scope of an individual mention.

### **Visits Pr. Month:**

An auxiliary metric which shows how much monthly traffic the specific website receives. This metric is meant to give the user a general sense of the size of the website.

### **Impressions (OTS):**

The central metric. This metric determines how many views your mentions could have seen. Which is why it's also known by its more accurate term, Opportunity To See (OTS). We estimate it by several factors. We use our general knowledge of the media channel that produced the mention, such as it's media type, country of origin, structure and other parameters. We also analyze the structure of the website. Finally, we pair these unique insights with the state-of-the-art data provided by SimilarWeb that specifically details the incoming traffic to the mention.

We are able to give you the most specific estimation of not just who saw a mention, but who clicked on it, increasing the exposure of your brand.

### **Estimated Clicks:**

This is our most detailed metric, and is available in the case where the website that produced the mention allows us to gather the necessary data needed to estimate it. With this metric, we combine the above metrics with a structural analysis of the website built on extensive periods of monitoring it. With this metric, we're able to give you the most specific estimation of not just who saw a mention, but who clicked on it, increasing the exposure of your brand.

To summarize, we've covered why and what impressions are for earned media mentions, and how we use applied AI and the best data we can get to estimate them. We have described varieties of the base impression metric, and our relationship with an industry-leading data provider. Finally, we'll continue to innovate by exploring newer estimation techniques and getting continuously better quality data.

Hypefactors is a Copenhagen based tech company helping organizations all over the world to unlock the business potential of PR. We combine data, analytics, technology and tools to provide a unified and easy-to-use experience. We provide an AI-based PR automation platform, including all the tools to power the PR workflow and all the facts to document the PR results. All-in-one and beautifully simple.

Hypefactors deliver the transparency that companies have been missing in order to understand the value of their PR and communication investments. Incl. the financial value of every single piece of media coverage and a lot of other facts to prove and improve the Earned Media results.

[www.hypefactors.com](http://www.hypefactors.com)  
[info@hypefactors.com](mailto:info@hypefactors.com)

**hypefactors**   
Unlock the business potential of PR