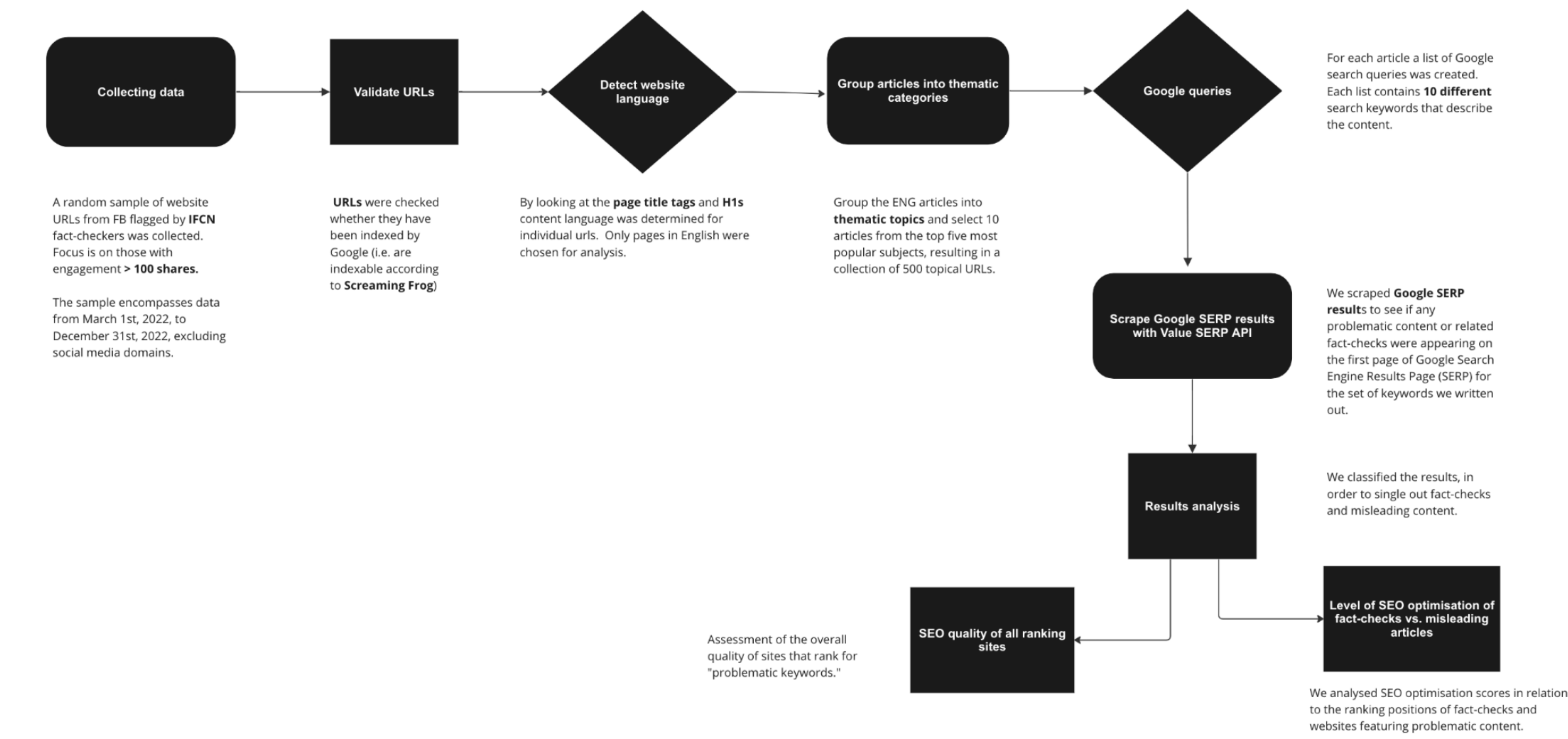


# SEO Effect

The competition between fact checks and ‘fake news’ in Google results.

## RESEARCH PROTOCOL

### Protocol



## Data collection

The research commences by selecting a set of 500 random website URLs that were identified by fact-checkers who participate in the Meta (3PFC) program. All URLs within this selection were shared on Facebook, either within a post or a linked post, with each of them receiving over 100 shares (that we use as a proxy for engagement).

Social media domains and URLs from video platforms were excluded from this sample, making it a faction of Meta's 38.099 entries. The time frame for this data collection spans from March 1st, 2022, to December 31st, 2022.

The sample was made available to us through Meta's URL Shared Dataset.

## Methodology

(1) SEO Screaming Frog was used to crawl the pool of URLs in the data set, to narrow down the analysis to pages that Google deems worthy of indexing. By looking at the page title tags and H1s content language was determined for individual URLs. Only pages in English were chosen for analysis. As a result of slimming down our sample to most relevant results, we ended up with a set of 138 URLs in English, of which majority were from US registered domains.

(2) After reviewing the set of URLs retained, the next step was to label them to reflect the topics they covered. We have analysed the contents of the URLs and assigned them the following labels: COVID, CLIMATE, US LIBERALS, US ELECTIONS, and UKRAINE (meaning the war in Ukraine).

(3) Subsequently, we selected the initial 10 articles from each category and reverse-engineer them to derive possible queries they might be ranking for in Google. We did that by looking at their SEO tags, H1s and unique way of phrasing things. Following this process, we got a list of 500 queries.

(4) We then used that list of queries to scrape Google search results. This returned a list of 5,000 new URLs (and respective rankings), of which only a fraction matched the original website domains. We then again labelled the list by marking fact-checking content, content that was disputing fact-checking narratives, and mainstream media content. Everything else was labelled as "other".

(5) The next step was to give a ranking, and an optimisation score to the set of 5,000 URLs we got in return. We did it by running SEO Site Checkup tool on the domains of these URLs. The tool takes into consideration the following features:

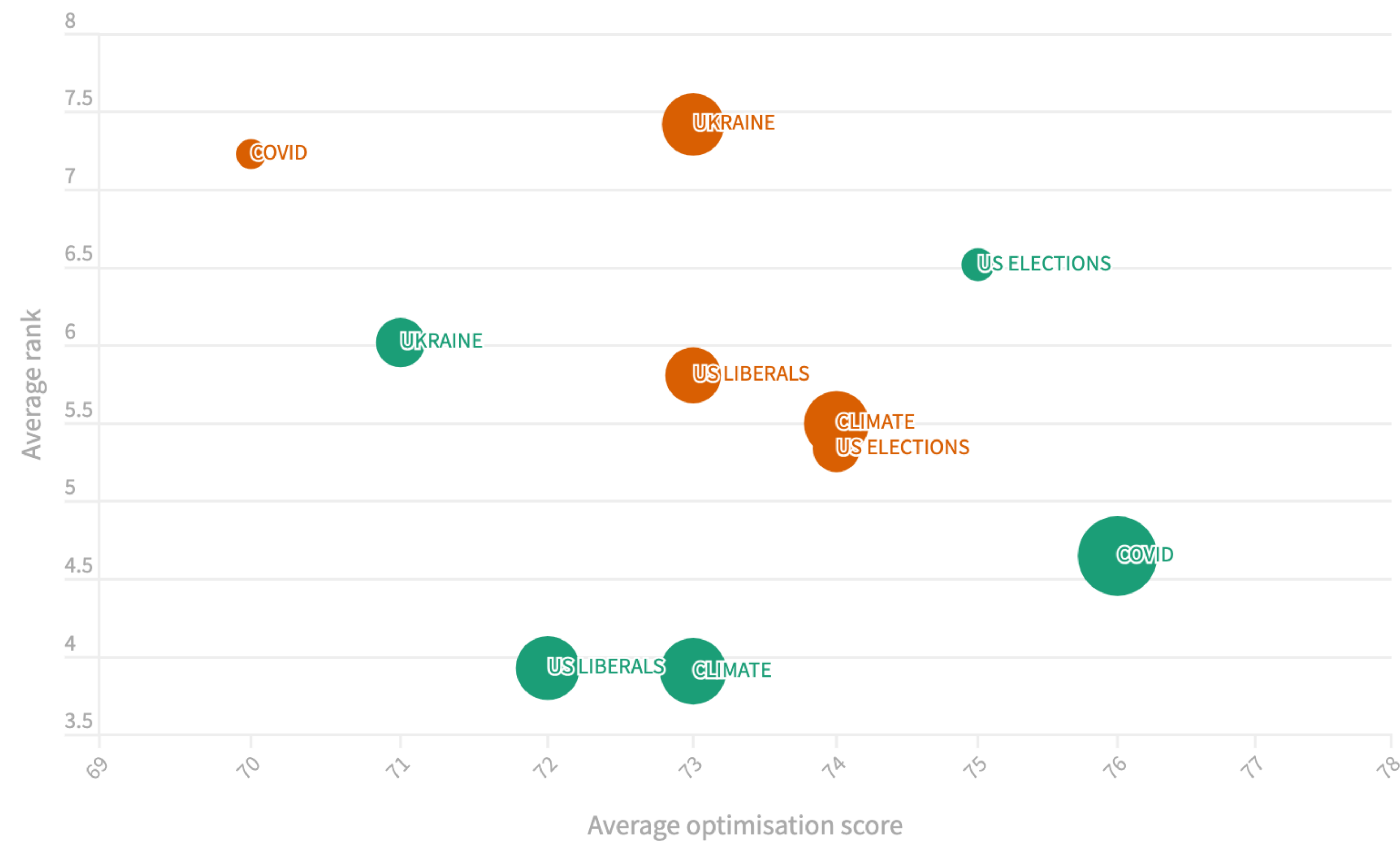
- Common SEO issues (presence of optimised meta titles, the use of social media tags, sitemap, presence of robots.txt rules, presence of the sitemap.xml file, url structure, GA code presence etc.)
- Page speed optimisations (CSS catching, DOM size etc.)
- Server and security (SSL, canonicalisation, protocols etc.)
- Mobile usability
- Advanced SEO (structure data markups, various types of webpage tags etc.)

Based on the presence or absence of these SEO attributes and adherence to Google Webmaster guidelines, a domain is assigned a score. The standard optimisation benchmark is 78 (based on 100 most visited websites in the US). Any score below this indicates incomplete or lack of optimisation efforts.

## VISUALISATIONS AND ANALYSIS

## The competition between fact checks and ‘fake news’ in Google results

Label fact-check problematic content



## Summary and Findings

### Finding 1

When assessing the ranking data and levels of site optimisation, we can observe an evident competition between fact-checks and narratives they try to correct. However, on average, fact-checking content and debunking articles tend to achieve higher rankings (4.6 vs 6.1 on average). They are also slightly better optimised (74 vs 73).

This can largely be attributed to the stronger performance of content relating to US LIBERALS and CLIMATE. On average, both types of content secure positions between 4th and 3rd in the SERP. Despite their average optimisation scores, ranging from 72 to 73, being slightly lower than the benchmark score of 78.

### Finding 2

In relative terms, it seems that Google is deliberately placing disinformation about UKRAINE further down in the SERP. This is despite the fact that the problematic content surrounding the conflict in Ukraine originates from sources that normally would be deemed as authoritative (such as duma.ru and tass.com). These sources also have on average an optimisation score comparable with fact-checks concerning CLIMATE but nonetheless appear much lower in Google SERP.

### Finding 3

Content aimed at rectifying misinformation about COVID has the highest average optimisation score of 76. This category also has the largest number of ranking URLs, totalling 99. In contrast, content that challenges the safety and effectiveness of Covid-19 vaccines ranks much lower, hovering around the 7th position. Google also displays fewer instances of such content, a mere 14 web pages. As a result, the visibility of these vaccine-related materials is notably lower, likely due to their comparatively modest SEO optimisation levels, around 70. This is also indicative of Google's broader commitment to targeting anti-vaccine narratives.

### Finding 4

The pages that enjoy the highest visibility scores are those featuring fact-checks that correct false information concerning US LIBERALS. Conversely, fact-checks addressing misleading narratives related to US ELECTIONS tend to perform less effectively on average than the articles that undermine the credibility of the US voting system.

### Finding 5

Using SEO Site Checkup indicators as a guide, it appears, on average, Google would consider fact-checking sources more deserving of higher ranking compared to domains that present contrary narratives. Generally, this assumption is validated as fact-checking content indeed tends to rank higher. Nonetheless, it also appears that the impact of Google moderation practices are in place. Content that challenges problematic narratives related to UKRAINE, CLIMATE, and US LIBERALS ranks significantly lower than the sites that seek to counter these narratives, regardless their average optimisation levels being almost similar.

## PARTICIPANTS

Richard Rogers (UvA) , Kamila Koronska (UvA), Denis Teyssou (AFP)