

ETHICAL CONCERNS OF SEARCH TECHNOLOGY: SEARCH ENGINE BIAS

Mirka Pastierová

*Comenius University, Faculty of Arts,
Department of Library and Information Science*

Abstract

Purpose – Search engines are a dominant part of our everyday activities and lives. These tools support decision-making, play a crucial role in constructing knowledge, and have a significant impact on our individual and social behaviour. The paper is aimed at the search engine bias problem as one of the important ethical issues associated with search technology algorithmic design and development.

Design/Methodology/Approach – Conceptual analytical method and critical approach are applied to a problem of search engine bias. Based on analysis, typology of specific problems and solutions are summarized and characterized. ANT (actor-network theory) (Latour, 2005) was used to introduce model consisting of search engine bias problem actors present in a complex bidirectional relations.

Results – The focus of this conceptual article is mainly on search engine bias in connection with manipulation techniques such as SEO (search engine optimization) methods and paid results. Problems of Google as a gatekeeper, personalization, and biased algorithmic design are also further analysed in more detail. As a result of the analysis, possible solutions to the search engine bias problem are categorized and discussed.

Originality/Value – This work provides new insights into search engine bias problem in the context of recent technological development and trends. The current methods, principles, and frameworks as a solution to the ethical issues of search engine technology are summarized.

Keywords: search engines, search engine bias, ethical issues, SEO, paid results, personalization, algorithmic design and development

INTRODUCTION

Information search activities and search engines are deeply embedded in our culture, society, and everyday life. Search engines determine what we know about this world and at the end shape our decisions and behaviour based on provided tools, information, and recommendations. In the era of algorithmic culture, implications of ranking and recommender algorithms in social contexts are evident. For that reason, the call for fair and unbiased search engines is more important than ever. Especially in a situation where there is predominantly one big player, Google, on the market.

Search engines have evolved from primitive systems for retrieving records and documents to adaptive and proactive agents powered by the most recent AI (artificial intelligence) technology. As stated by Bostrom (2016), “Google search engine is the greatest AI system that has yet been built”. Alongside the development of AI and other related search technologies, emerging ethical issues are relevant on multiple levels.

Crucial ethical concerns related to the search engine technology can be organised into the following broad categories (Hinman, 2005; Tavani, 2012):

- problem of algorithm (search engine bias, the problem of opacity/non-transparency),
- personal privacy and informed consent,
- monitoring and surveillance, and
- censorship and democracy.

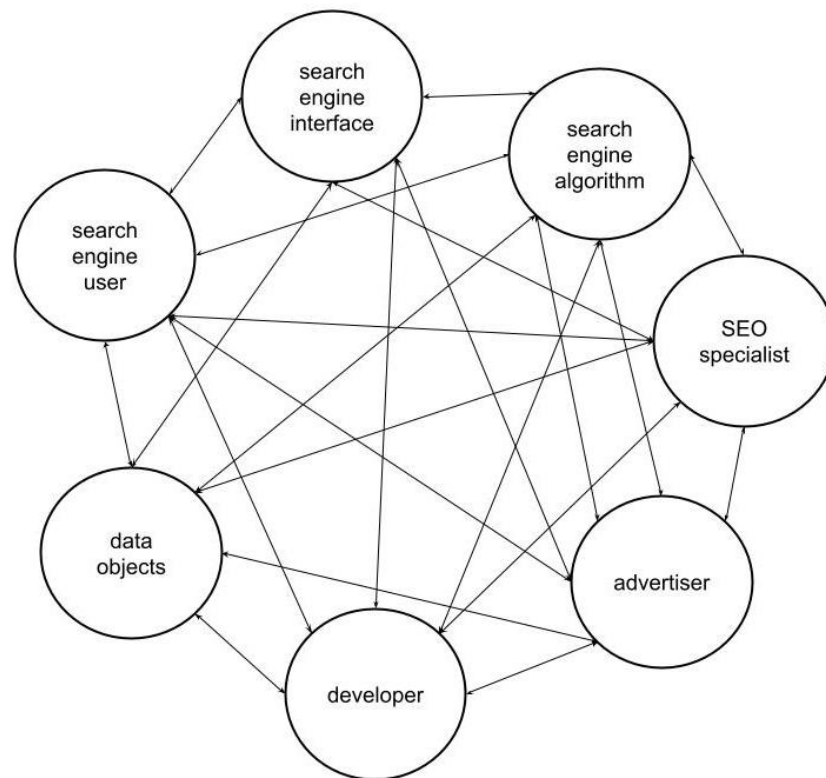
The focus of this article is on *search engine bias*, but as it relates to other issues listed above, we do not try to isolate it and find some relations between them. Search engine bias is a part of a wider cluster, a problem of algorithm. It might seem that it is independent of human intervention, but the opposite is true. Many of the search engine ethical issues stem directly from human subjective judgments, value systems, and mental models. On the other hand, together with unstoppable development of search technology, and especially machine learning, new ethical challenges such as accountability, control, transparency, etc. are emerging continuously.

1 SEARCH ENGINE BIAS

Search engines are one of the most important tools for accessing digital content and presenting it to a user on the basis of his/her actual (or proactively predicted) information needs. Nowadays variable modes of interaction are available to end users starting with (more common) text, voice, image, gestures, etc. and their combinations in a form of multisearch (Raghavan, 2022). Various algorithmic models are implemented to process, index, store, and present data via search engines. Algorithmical architectures differentiate one search engine from the other, they create competitive advantage, and most of the time their particular parts are protected by intellectual property rights.

It is also important to note that traditional search engines (such as Google, Bing etc.) do not index web content in its entirety. They access only the surface web, and most of the web content that is hidden from regular search engines is called deep (invisible) web. Sometimes it is interchangeably termed the dark web, which is incorrect, because the dark web is just a part of the deep web that includes illegal content.

Based on our analysis and previous work (Pastierová, 2021a, 2021b, 2021c) we summarized network model of search engine bias problem (Fig 1). Model applies actor-network theory (ANT) (Latour, 2005) which we have already discussed in more depth recently (Pastierová, 2021c). Network model includes humans acting in many roles such as search engine user, developer, advertiser, SEO (search engine optimization) specialist, etc. Another element of the interconnected network is the search engine itself that contains an algorithm defining specific features, including values and the user interface. Data objects and structures are part of the search engine bias problem, too. These actors are a part of constantly shifting networks of relationships depending on a specific social, political, economic contexts and situation.



social - political - economic contexts

Fig. 1 Search engine bias problem actors

We will reflect on specific ethical issues regarding these actors and contexts in further analysis which applies the following structure of search engine bias problems:

- Google as a gatekeeper,
- SEO methods,
- paid results,
- biased algorithmic design,
- outcomes of personalisation.

1.1 Google as a gatekeeper

The Google search engine remains the dominant player on the market for organic search and paid placement. The worldwide market share of Google as of July 2022 was 91.4 % followed by Bing with only 3.3% and Yandex 1.49% (Search, 2022).

Google is the first choice for users because, with the help of continuous investments in marketing, it managed to be top of mind. The word “google” also became part of the casual narrative when searching for information using search engines in general. Users have individual motivations regarding Google preference over alternative search engines, but the main reason behind its dominance is a long list of

business contracts with distributors. Google became a monopoly gatekeeper because, in many cases, it prohibits distributors from dealing with alternative search providers to secure its position as a default search engine (Goodison, 2020).

Google's position as a monopoly and powerful arbiter of public information is critical because it shapes the way knowledge is socially constructed. It can have a significant impact on our behaviour and priorities. It can potentially, as any other marketing tool, transform our values aiming at creating artificial needs for more goods and supporting consumer culture. Search engines can become tools of manipulation, control, and isolation on multiple levels. They can also help to polarise and control social groups, weaken the society as a whole, and create an arena for escalating unacceptable behaviours.

One of the extreme examples could be the case of Dylan Roof who murdered nine people in Charleston in 2015 (Hersher, 2017) that was reasoned by attorney who argued that Google shaped Roof's beliefs. Roof followed another case of racial profiling of Zimmerman from 2012 who shot and killed Trayvon Martin. Roof believed that Zimmerman was on the right and typed keywords "black on white crime" into Google and this specific act changed the trajectory of his life following the white supremacist path. At the time, Google also generated racist keywords using the autocomplete feature. So hypothetically, if Roof used Google typing 'black on', he could get related suggestions such as 'black on white crime', 'black on white violence', etc.

1.2 SEO methods

SEO (search engine optimization) is a methodology and a practical process designed to help a specific website be visible at the highest possible position among organic results (unpaid listings) on the first SERP (search engine result page).

In a study by Nielsen Norman Group, named Search Meta-Analysis Project, conducted in 2019, researchers found that in 20% of the 377 search instances, users looked at only one result on SERP (Liu, 2020). The Sistrix study (Southern 2020) using data from over 80 million keywords and billions of search results showed that the click-through rate (CTR) of search engine results decreased dramatically from the first organic result with an average of 28.5% to a second result with 15% and third with only 11% CTR.

Search behaviour has radically changed together with the gradual introduction of new search engine features such as knowledge panel, featured snippets, 'people also ask' etc. Users tend to spend more time on SERP and sometimes even get their answer right away, so they do not need to leave the result page. That is the main reason behind the trend of declining CTR and could also have a radical impact on website organic traffic.

Ethical SEO principles are embedded in so-called white-hat SEO methods. These are directly recommended by SEO specialists, on the contrary black-hat SEO methods are recognized and (most of

the time) penalised by search engines. Understanding the actual search engine ranking factors is crucial in the effective SEO process. What is quite interesting is that it takes a lot of testing to define them exactly because no search engine makes them publicly available. Speaking of black-hat SEO, some of the old-fashioned methods are still present (such as keyword stuffing, cloaking, thin content etc.), but they usually do not work because they are easily identified by a search engine robot. At the end there are cases where ranking algorithm criteria were changed based on unethical SEO practices. For example, the use of relevant keywords that describe content in metadata was previously often misused to manipulate the ranking of the results. Now they have so little significance that some SEO practitioners do not use them at all.

Overall, we can conclude that ethical SEO is still a powerful tool to manipulate search engine results and achieve higher rankings that contribute to search engine bias issues at the end.

1.3 Paid results

The ultimate problem with search engines is the obvious conflict between user interests and monetization techniques of search engines such as paid placement. It is important to understand that Google is an advertising business and is sponsored by advertisers. In 2021, Google's revenue increased to 256.7 billion US dollars, from which 209.49 billion US dollars was made up by advertising revenue (Annual 2022). The Google Ads platform is steadily growing and as more businesses are going online, this trend is not going to change in the near future. In this case, search engine bias is present due to targeting ad content based on behavioural signals and personal user information (Pastierová, 2021). On the other hand, forms for accepting Google Ads can be easily manipulated to attract users to ads. They can use keywords based on the popularity of search traffic and do not have to be strictly based on the relevance to the actual ad to a website content. This can be confusing for a user, especially if distinguishability of paid ads is limited.

Most of the results on SERP now consist of commercial offerings, which are many times not possible to differentiate from organic results easily. Schultheiß and Lewandowski (2021) conducted a study with 100 participants, recorded click and gaze behavioral data by completing tasks and followed by questionnaire. The results showed that the ads were viewed intensely because they were prominently displayed to gain visual attention. Most of the time, users without any knowledge of ads selected actual offerings instead of organic results without even realising it. Schultheiß and Lewandowski (2021) noted that these ads should be appropriately labeled so that users can identify them as paid ads.

1.4 Biased algorithmic design

Search engine algorithms are designed to rank and prioritise results depending on many factors such as user model, relevance of the information to user intent but also actual search traffic and popularity of

a specific search result. Ranking algorithms can inherently produce a problem of search engine bias in a process of weighting specific factors (and moreover including results that are paid by advertisers).

Noble (2018) points out that bias is present in search engines based not only on aggregating user behavioural data, but also that it is embedded directly in code. She coined the term 'technological redlining', which refers to new forms of algorithmic profiling based on race or sex in decision-making tools. Possible consequences of such practices are social inequality of marginalised groups with impact on democracy as it is.

Throughout previous years, there were many cases of Google having to deal with critiques of search results biased by racist, sexist, and other offensive values (Sundin et al., 2022). Some instances of potentially discriminatory narratives of Google autocomplete and results were described by Cadwalladr (2016). For example, she tried a question 'are Jews' and received, among other suggestions, a question 'are Jews evil?' which leads to links with neo-nazi content. Of course, these widely medialized cases prompted Google into action, so it changed its autocomplete suggestions. But still these instances of hate speech and discrimination have serious implications for our democracy. Meanwhile, Google does not take any responsibility for that and argues for search results as a reflection of actual content available on the Internet and usually blames users.

Human interventions are present in the process of making editorial choices (Goldmann, 2006) about what data to include or exclude in a search engine database. Another issue is that the mental models of designers and their value systems are embedded in design. Together with development of machine learning techniques, and especially self-supervised machine learning algorithms, less human control is needed for intelligent systems to work. This implies that not only less is known about algorithmic features (co called black box algorithms) but also more attention needs to be paid on the issue of transparency of AI on its own. The ethical problem is also inherently present in the mechanics of machine learning algorithms which are training on specific data sets. Training data could be distorted or biased at the beginning, which has a direct influence on the algorithm output and following decisions or recommendations.

Political and social bias is present in the algorithmic design of search engines in countries that censor search results. These techniques lead to filtering or blocking specific content to align with the political regime and comply with censorship laws. Google Search was blocked by the 'Great Firewall' in China in 2002 and later continued to provide censored search services between 2006 and 2010 (Murphy, 2016, Gallagher, 2018, New, 2010). As it was announced based on leaked Google confidential documents, it plans to launch a special project "Dragonfly" that would allow algorithms to "blacklist websites and search terms about human rights, democracy, religion and peaceful protest" (Gallagher, 2018). Not only that censored websites will be potentially filtered, but also results from banned search queries will show no results. Google's endeavours to launch the Dragonfly application in China represent a potentially big

revenue stream, but on the other hand raises many moral and ethical concerns. Google wants to provide services for everyone, but in this case at the expense of joining forces with the oppressor of human rights and free speech.

1.5 Outcomes of personalization

In the past, ranking algorithms were designed to calculate results and present them on a basis of “consensus relevancy” (Pitkow, et al. 2002). It was assumed that the information is relevant to a specific user based on computing the relevancy for the entire population. This one-size-fits-all approach was overcome by personalization techniques using contextualization and individualization.

Personalization was supposed to be one of the solutions to a problem of search engine bias and a tool for reducing it (Goldman, 2006), but it brought some unexpected outcomes. Personalization of search results is executed with the help of extensive user data collection including ID, search history, e-mail address, and localization to mention just a few. These data significantly help to bring more relevant personalised results tailored for the specific needs of a user. On the other hand, it involves a spectrum of ethical problems. One of them is that search engine results are restricted, and therefore, also biased by information, which are suggested to be relevant depending on previous user behavioural patterns and interests. This phenomenon, present not only in search engine results but also in social media (Facebook, Twitter, etc.), is known as the “filter bubble” (Pariser, 2011). The filter bubble effect causes isolation in a loop of one-sided knowledge and limits us from expanding into new learning territories.

2 POSSIBLE SOLUTIONS TO THE SEARCH ENGINE BIAS PROBLEM

Solving search engine bias problems is currently a major responsibility of companies providing search services. There is certain pressure developed by the media, academia and also end users, especially members of marginalised communities. At the same time, not all the problematic instances are solved.

Search engine bias recommendations for a solution could be found in various areas and on multiple levels starting with:

- theoretical frameworks,
- practical design methods,
- algorithmic features in search engines,
- ethical codes of leading organisations,
- legal frameworks,
- national and transnational initiatives such as government frameworks and guidelines.

The main initiatives regarding the application of ethical principles are provided by models and design principles developed by theorists in the fields of computer ethics and technology design (Tavani, 2012). The specific field of IT ethics also encourages initiatives in line with fair and unbiased search engine design. One of the theoretical frameworks implementing human ethical values in the process of information system design is value-sensitive design (VSD) (Friedman, Kahn, and Borning 2008, Friedman, Hendry, 2019).

More recently a specific category of alternative tools emerged, anonymous and private search engines (Pastierova, 2021) such as DuckDuckGo (2022), StartPage (2022), SwissCows (2022) etc. They eliminate the problem of filter bubbles and do not personalise search results because these tools do not collect personal user information nor create user profiles. Some of them also provide features to prevent third parties from tracking and monitoring users when exiting the SERP.

Google has responded to a search engine bias problem and continually changed different features of the ranking algorithm. Of course, there are still many problems unsolved, but there are some interesting projects and efforts aligning with the principles of inclusion. More recently, Google introduced Google Real Tone which brings skin tone equity across Google products (Raghavan 2022). Google implemented the existing Skin Tone Scale developed by Ellis Monk (Skin 2022) to create more inclusive products and eliminate problems of colour bias and colorism. The Google Real Tone feature enables to filter images according to a skin tone. For example, results for search query “bridal makeup” were mainly about tutorials for white females, using this new feature will enable to choose results according to an actual skin tone of a user. This is just to mention one feature, which is a part of a wider set of initiatives and principles focused on inclusivity of Google Products (Designing, 2022).

The ACM Code of Ethics and Professional Conduct (2022) was developed to provide guidance on ethical principles important in decision making for professionals who use computing technology in an impactful way. Among the fundamental ethical principles is respecting diversity and human rights. Fairness should also be promoted at all times and professionals should avoid creating technologies to discriminate or oppress people on the 'basis of age, color, disability, ethnicity, family status, gender identity, labor union membership, military status, nationality, race, religion or belief, sex, sexual orientation, or any other inappropriate factor'. (ACM, 2022)

On the level of EU governance, the Ethics Advisory Group (EAG) was established to provide ethical guidance. Many activities of the High-Level Expert Group on Artificial Intelligence (AI HLEG) are also focused on ethical issues. AI HLEG drafted guidelines for human-centered and trustworthy AI systems called Ethics Guidelines for Trustworthy AI: Shaping Europe's Digital Future (Ethics, 2018). One of the main points is that trustworthy AI should respect ethical principles and values.

One of the more polarising viewpoints is that search engines should be legally regulated to become neutral. In contrast, regulation by state or government can create other issues of search engine bias connected to censorship (as we know it from China).

CONCLUSION

The problem of biased algorithmic design continually creates new challenges and is a cause that search engine technology is not objective, neutral, or value-free. If further development in the areas of search engines is not going to change, we will continue to follow this path in the future.

The ethical issue of search engine bias needs to be better understood so that developers and especially users have the choice to make informed decisions. Users also need to know underlying mechanisms of result ranking and recognize that in many cases monetization models have priority over actual user needs.

Based on the above-mentioned concerns, it is possible to conclude that search engine technology is currently far from being neutral nor objective. Subjectivity is inherent in algorithmic design and embedded in indexing and ranking methods. Search engine bias generates ethical problems based on manipulating results, shaping user knowledge, individual behaviour, decisions, and experiences. Another issue is that of the conflicting interests of search services providers as ad companies.

Partial solutions to a search engine bias problem are already in action, but many remaining concerns need to be addressed. A more holistic approach to solving this problem is needed. In the end, there is still hope that together with evolving intelligent technologies more guidelines and frameworks will be implemented with a common aim to align search engine development with ethical values and principles. Meanwhile, it is important to understand that it is not possible for frameworks and guidelines to reflect on all the emerging issues because of divergent character and the dynamic pace of technological development.

ACKNOWLEDGEMENT

The article was developed in the framework of the VEGA 1 / 0360 / 21 project Social representations of ethical challenges of the digital information revolution.

REFERENCES

- ACM Code of Ethics and Professional Conduct. (2022) *ACM*. Available from: <https://www.acm.org/code-of-ethics>.
- ANNUAL Report pursuant to Section 13 or 15(d) of the securities exchange act of 1934 (2022). Alphabet, 2022. Available from: https://abc.xyz/investor/static/pdf/20220202_alphabet_10K.pdf.
- BOSTROM, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- CADWALLADR, C. (2016). Google, Democracy and the Truth about Internet Search. The Observer, 4 December 2016, *The Guardian*. Available from: <https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook>.
- DESIGNING Inclusive Products for Everyone - Google. (2022) Available from: <https://about.google/belonging/in-products/>.
- DUCKDUCKGO. (2022). *Homepage*. Available from: <https://duckduckgo.com>.
- ETHICS Guidelines for Trustworthy AI: Shaping Europe's Digital Future (2018). Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- FRIEDMAN, B., P. KAHN, and A. BORNING. (2008). Value-sensitive design and information systems. *The Handbook of Information and Computer Ethics*, K. E. Himma and H. T. Tavani (eds.), Hoboken, NJ: John Wiley and Sons, pp. 69–101.
- FRIEDMAN, B., HENDRY, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press. ISBN 9780262351706.
- HERSHER, R. (2017). What Happened When Dylann Roof asked Google for information about race? *NPR*, 10 January 2017. Available from: <https://www.npr.org/sections/thetwo-way/2017/01/10/508363607/what-happened-when-dylann-roof-asked-google-for-information-about-race>.
- GALLAGHER, R. (2018). Google Plans to launch a censored search engine in China, leaks reveal. *The Intercept*. August 1, 2018. Available from: <https://theintercept.com/2018/08/01/google-china-search-engine-censorship/>.

GOLDMAN, E. (2006). Search Engine Bias and the Demise of Search Engine Utopianism. *Yale Journal of Law and Technology*. 188 (Spring 2006). Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=893892.

GOODISON, D. (2020) Google Is “A Monopoly Gatekeeper:” DOJ Antitrust Lawsuit. *CRN*, 20 Oct. 2020. Available from: <https://www.crn.com/news/google-is-a-monopoly-gatekeeper-doj-antitrust-lawsuit>.

LATOUR, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: OUP Oxford, 2005. ISBN 9780199256044. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=211608&lang=sk&site=ehost-live&scope=site>

LIU, F. (2020). The Love-at-First-Sight Gaze Pattern on Search-Results Pages. *Nielsen Norman Group*. Available from: <https://www.nngroup.com/articles/love-at-first-sight-pattern/>.

MURPHY, G. (2016). Search Engine Bias: An Analysis from the Index. *Search Engine Watch*. Available from: <https://www.searchenginewatch.com/2016/02/17/search-engine-bias-an-analysis-from-the-index/>.

NEW Approach to China: An Update. (2010). *Official Google Blog*. Available from: <https://googleblog.blogspot.com/2010/03/new-approach-to-china-update.html>.

NOBLE, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018. Available from: <http://ebookcentral.proquest.com/lib/uniba-ebooks/detail.action?docID=4834260>.

PARISER, E. (2011). Beware online "filter bubbles". *TED Talk*. Available from: https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.

PASTIEROVÁ, M. (2021a). Aktuálne trendy vyhľadávania informácií. *Knižnica*. No. 2. Vol. 2021. Available from: https://www.snk.sk/images/Edicna_cinnost/Casopis_Kniznica/2021/2021_07_06_01_Mirka_Pastierova.pdf

PASTIEROVÁ, M. (2021b). Problémy vyhľadávania informácií v kontexte súkromia. *ITlib: Informačné technológie a knižnice*, 2021, pp. 24–30. Available from: <https://doi.org/10.52036/1335793X.2021.1-2.24-30>.

PASTIEROVÁ, M. (2021c). Trends in digital information search. *Knížničná a informačná veda XIX*, 2021, pp. 109–119. Available from:

https://fphil.uniba.sk/fileadmin/fif/katedry_pracoviska/kkiv/zbornik_KIV/kiv29_pastierova.pdf.

PITKOW, J. et al. (2002). Personalized search. *Communications of the ACM*. 45, 9 (September 2002), 50–55. Available from: <https://doi.org/10.1145/567498.567526>.

RAGHAVAN, P. (2022). *Google I/O keynote*. Available from:

<https://io.google/2022/program/8e80903f-955f-4a5b-9118-b0ce4acdb0e6/>.

SEARCH Engine Market Share Worldwide. (2022). *StatCounter Global Stats*. Available from:

<https://gs.statcounter.com/search-engine-market-share>.

SCHULTHEIß, S., and D. LEWANDOWSKI. (2021). Google search results—They’re all the same, right? *Information Matters*. Vol.1, Issue 11. Available from:

<https://r7q.22f.myftpupload.com/2021/11/google-search-results-theyre-all-the-same-right/>

SKIN Tone Research @ Google. (2022). Available from: <https://skintone.google/get-started>.

SOUTHERN, M. G. (2020). Over 25% of People Click the First Google Search Result”. *Search Engine Journal*. 15. July 2020. Available from: <https://www.searchenginejournal.com/google-first-page-clicks/374516/>.

STARTPAGE. (2022). *Homepage*. Available from: <https://www.startpage.com/>.

SUNDIN, O., D. LEWANDOWSKI and J. HAIDER. (2022). Whose relevance? Web search engines as multisided relevance machines. *JASIST*. Volume 73, Issue 5, May 2022

SWISSCOWS. (2022). *Homepage*. Available from: <https://swisscows.com/>.

TAVANI, H. (2012). Search Engines and Ethics. *Stanford Encyclopedia of Philosophy*. August 2012. Available from: <https://meinong.stanford.edu/entries/ethics-search/index.html>.

ABOUT THE AUTHOR

Mirka Pastierová

She is a research assistant at the Department of Library and Information Science, Faculty of Arts, Comenius University, Bratislava, Slovakia. Her current research interests are focused mainly on information search, search technology, human-computer interaction, and cognitive science. In her teaching she concentrates on HCI, information retrieval, information organization, and SEO. She has participated in various research projects focusing on information behaviour, human-agent interaction, and educational technology. Currently, she is a part of research team dealing with a topic of ethical challenges of digital information revolution.

Dept. of Library and Information Science, Faculty of Arts, Comenius University Bratislava, Gondova 2, 818 01 Bratislava, Slovakia

ORCID: 0000-0001-8056-9421

E-mail: mirka.pastierova@uniba.sk