Digital Platform Services Inquiry - September 2024 report revisiting general search services

Issue Paper

18 March 2024

Submission to the Australian Competition & Consumer Commission

17 April 2024

Professor Mark Sanderson, Professor Daniel Angus, Dr Aaron Snoswell, Dr Dana Mckay and Dr Johanne Trippas

ARC Centre of Excellence for Automated Decision-Making and Society
School of Computing Technology, RMIT University

Contents

The ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) welcomes the opportunity to comment on the Australian Competition and Consumer Commission (ACCC) Digital Platform Services Inquiry – September 2024 report revisiting general search services (Issues Paper).

About ADM+S

The ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) is a cross-disciplinary, national research centre which commenced operations in mid 2020. ADM+S has been established and supported by the Australian Research Council (ARC) (CE200100005) to create the knowledge and strategies necessary for responsible, ethical, and inclusive automated decision-making. Focus areas for ADM+S research are news and media, social services, health and transport. ADM+S brings together nine of Australia's leading universities, and more than 80 researchers across the humanities, social and technological sciences, together with an international network of partners and collaborators across industry, research institutions and civil society. More information about the ADM+S, our researchers and research projects can be found on our website: **www.admscentre.org.au**

The authors also wish to acknowledge Professor Axel Bruns (ADM+S at QUT) for his valued input into this submission.

About the School of Computing Technologies, RMIT University

The School of Computing Technologies is a centre for digital innovation, world class research, and education in STEM via streamlined collaboration across relevant disciplines. Researchers conduct world leading research in artifical intelligence, search and recommendation, machine learning, data analytics, distributed computing and cybersecurity. More information about the School of Computing Technologies can be found at **www.rmit.edu.au/about/schools-colleges/computing-technologies**

We thank the ACCC for the opportunity to provide this submission. Should you wish to discuss any of the issues raised in this submission, please do not hesitate to contact the ADM+S Centre at adms@rmit.edu.au

Responses

Our submission uses insights from our research at the ADM+S and considers the following questions raised in the issues papers: questions 8, 15, 16, 18, 23 and 24.

8. What are the most effective methods of sharing click-and-query data? How could the privacy and security risks associated with the sharing of click-and-query data be mitigated?

One of the most comprehensive click-and-query datasets that is currently available is Meta's URL Shares dataset which is available to approved researchers through the Social Science One initiative. According to Meta (**https://transparency.fb.com/en-gb/researchtools/other-datasets/**):

"The URL Shares dataset summarizes the demographics of those who viewed, shared and otherwise interacted with web pages (URLs) shared on Facebook starting January 1, 2017 up to and including October 31, 2022. URLs are included in the dataset if shared (as an original post or reshare) with "public" privacy settings more than 100 times. Social Science One hosts applications for access to the URL Shares dataset. Researchers whose applications are approved by Social Science One will be granted access once per quarter following completion of onboarding with Meta."

The dataset enables the systematic study of not only the hyperlinks that are posted (shared) by users of Meta's Facebook platform, but also the interactions with these links by other users of the platform, including clicks, shares, reactions, and more. The inclusion of view and interaction data is key to enabling observability of this URL sharing ecosystem on the platform. Critically, it is not just the transmission side (posting of links by users) that is important in understanding the ecosystem of URL sharing on Facebook, but also the reception and interaction with these links by other members of the platform. Such audience reception and interaction data informs how Rieder and Hofmann (2020) 'propose the concept of observability as a more pragmatic way of thinking about the means and strategies necessary to hold platforms accountable' (p. 3)[1]. Relating this to search, observability suggests that a full understanding of search cannot be gained by information on search rankings alone, but also in connecting such information with how users interact with and click through the links provided.

While many platforms have made moves in recent years towards offering greater transparency regarding platform operations, these transparency initiatives often stop short of providing the audience reception and interaction data above, leading to a thin form of transparency that ultimately limits platform accountability[2]. The concept of observability is therefore key to ensuring that any transparency data provided by platforms can be contextualised and

[1]Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, 9(4), 1-28.
[2]Crain, Matthew. (2018) The Limits of Transparency: Data Brokers and Commodification, *New Media & Society,* 20(1), 88–104.

understood, that we can 'observe' how platforms operate on and through users.
A key example of click-and-query data that would significantly strengthen oversight of digital platforms and improve public accountability would be data on platform-based advertising. As a counterexample to the moderately well-designed and provisioned URL Shares dataset, Alphabet and Meta presently only provide a thin form of transparency via their advertising datasets and dashboards which provide copies of ads (the ad creative, and basic page/vendor data) that are currently running on their platforms. While these transparency dashboards enable a basic understanding of the kinds of ad messages that are circulating on these platforms, there is an absence of data on who has clicked through, interacted with, or otherwise encountered this advertising content online[3]. This lack of advertising observability frustrates attempts to understand the wider reception of this content, which is crucial if we are to address forms of harmful advertising content[4], provide assurances as to the accuracy of platform-provided metrics on advertising, or simply understand how ad placement works inside search and other platform services. It is for these reasons that members of the ADM+S ran a multi-year study on advertising observability (**https://www.admscentre.org.au/adobservatory/**) that enrolled citizen scientists to donate the ads they encountered online[5]. While results from this study have enabled greater insights into advertising reception, it would be more scalable and effective if advertising data (including clicks, interactions, etc.) was made available using a model similar to the Meta URL shares dataset above.

In terms of privacy and security risks, the Social Science One model, the governance model applied to Meta's URL Shares data, is appropriate for dealing with many of these issues. The URL Shares dataset addresses privacy and security risks through aggregation, anonymization, differential privacy, and secure cleanroom data access environments.

Users are described using aggregated demographic buckets, such that a user can be known by the country they are from, age bracket (e.g. 35-44), and gender. Gaussian noise is added to data tables to help protect individual privacy and prevent discoverability of users in the table. The dataset can also only be interrogated within a secure cloud environment after researchers have submitted a research plan, which includes institutional-approved ethics plan(s), that is reviewed by Social Science One. This cleanroom environment allows full interrogation of the dataset, however it also prevents download of raw data and tables, as only downstream analysis and results can be exported, not raw data. While this may frustrate some research, it is a fair compromise while this model is stress tested by the various international teams that currently have access.

[3]In the limited case of political ads we do get some additional data on who the ad may have been targeted towards (age range and M/F gender), and an estimate of how many people have seen the ad.
[4]Parker, C., Albarrán-Torres, C., Briggs, C., Burgess, J., Carah, N., Andrejevic, M., ... Obeid, A. (2023). Addressing the accountability gap: gambling advertising and social media platform responsibilities. *Addiction Research & Theory*, 1–7. https://doi.org/10.1080/16066359.2023.2269852
[5]Burgess, J., Andrejevic, M., Angus, D., & Obeid, A. K. (2022). The Australian Ad Observatory: Background Paper - ADM+ S Working Paper Series. *Analysis & Policy Observatory*. https://apo.org.au/node/318616

User information

user_id	country	agecat	gender	ppacat
user1	DE	35-44	F	NULL
user2	US	25-34	M	-1
user3	BR	25-34	F	NULL
⋮	⋮	⋮	⋮	⋮

User activity

url_id	user_id	timestamp	action
url1	user1	1483247834	like
url2	user1	1483266874	click
url1	user2	1483276812	share
⋮	⋮	⋮	⋮

JOIN ON user_id

GROUP BY (PRIVATE) SUM $\left(\sum\left[\mathbb{1} + N(0,\sigma^2)\right]\middle|\mathbf{X}\right)$

url_id	year-month	country	agecat	gender	ppacat	view	click	share	like	...
url1	2017-01	DE	35-44	F	NULL	548	44	-4	68	⋮
url1	2017-01	US	25-34	M	-1	4736	199	111	152	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Breakdown Table

url_id	clean_url	share_title	...	hate_speech	top_country
url1	www.url1.com/hi.html	Hello	...	0	DE
url2	www.url2.com/hi.html	Hi	...	4	US
⋮	⋮	⋮	⋮	⋮	⋮

URL Attributes Table

Accessible data

Data anonymization and aggregation by Meta URL Shares data, (from: Facebook Privacy-Protected Full URLs Data Set (version 10.1), by Messing, Solomon; DeGregorio, Christina; Hillenbrand, Bennett; King, Gary; Mahanti, Saurav; Mukerjee, Zagreb; Nayak, Chaya; Persily, Nate; State, Bogdan; Wilkins, Arjun).

15. To what extent do consumer-facing LLM-based chatbots compete with general search services at present?

At this stage of deployment of such services, one can only speculate on exactly how much chatbots will compete with general search services. One data point that we can highlight, however, is work by researchers at RMIT University who have studied the content of a large number of prompts that were volunteered from a set of crowdsource workers using the Google Gemini chat bot. When the paper[6], is published a 1K collection of prompts collected during this research will be made publicly available for study.

Conclusions of this study include that nearly 50% of the dataset prompts collected are 'one-turn' - no conversation. This result highlights that these 'chatbot' systems have not yet achieved a genuinely conversational state. Their current mode of interaction, focuses on direct commands and task execution. The interaction lacks the natural, fluid dialogue of human conversations, indicating a gap in achieving genuine interactive information retrieval and conversational exchanges. Many users in the study were able to vary their interaction style to target the platform's intended use, while 36% of users had an average prompt length of 15 words or less, indicating the use of a "web-search" interaction style characterized by short queries.

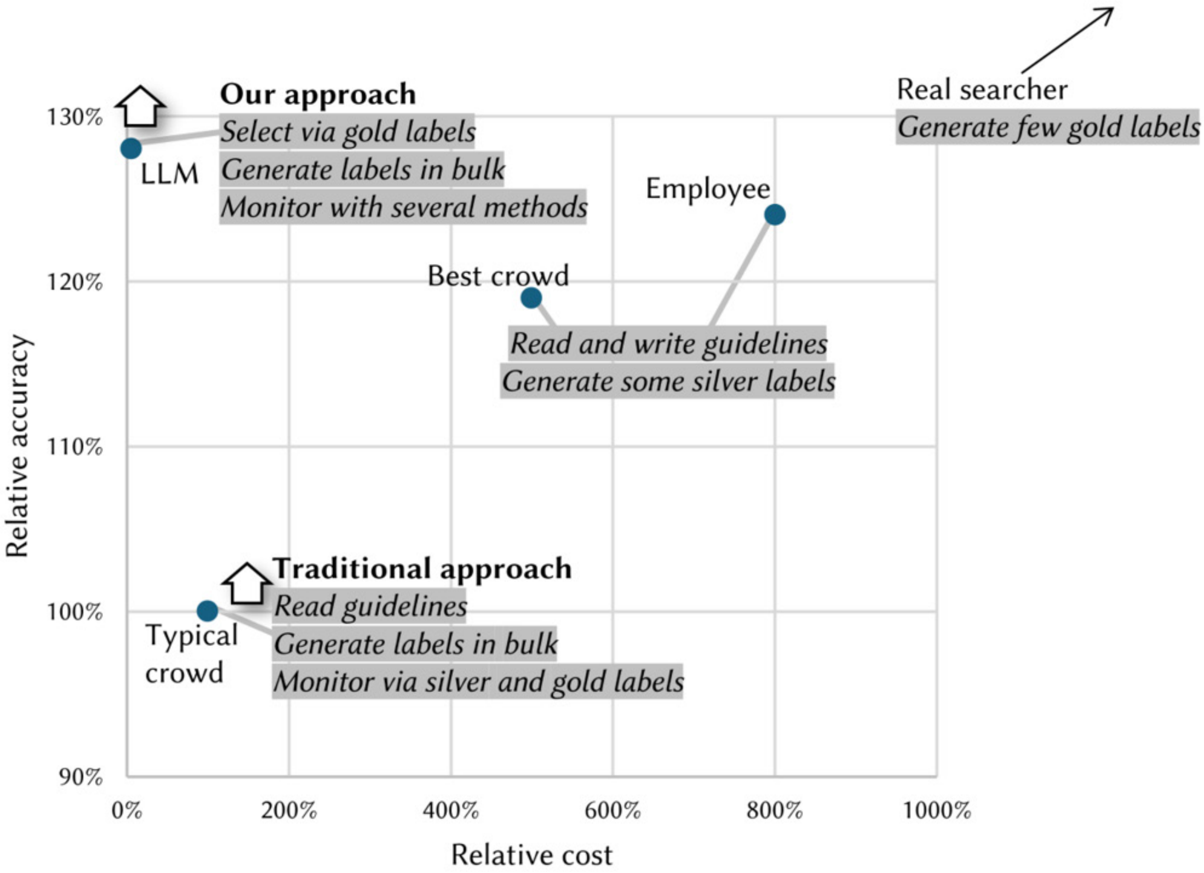
Based on this study, one might be drawn to conclude that chat bots do present competition with general search services.

16. How has generative AI been integrated into search engine services so far? In terms of their utility and effectiveness in finding information for consumers, how do they compare with general search services?

The most striking integration of generative AI into search engine services has happened behind the scenes. In a tech talk in May 2023 hosted at RMIT University[7], Microsoft Bing revealed it was using the LLM, GPT-4, to create most of the label data that records the relevance of documents to queries. The labels are used to train the machine learning algorithms of their standard Bing web search engine. Until this announcement, it was assumed that such a labelling task could only be done by humans. In the case of Bing, they had been using a large team of crowdsource workers. The GPT-4 labelling process they use replaces nearly all human assessment.

Subsequent to the talk, a paper was published[8] detailing the labelling process further. The following figure (taken from figure 5 in the paper) shows the very low cost and relative high quality of using an LLM labelling approach compared to a traditional crowdsourcing approach.

[6] Johanne R. Trippas, Sara Fahad Dawood Al Lawati, Joel Mackenzie, and Luke Gallagher. 2024. What do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*, July 14–18, 2024, Washington, DC, USA. ACM, New York, NY, USA, 5 pages. https://www.johannetrippas.com/papers/trippas2024what.pdf
[7] https://x.com/IR_oldie/status/1659413086007328768
[8] Thomas, Paul, Seth Spielman, Nick Craswell, and Bhaskar Mitra. "Large Language Models Can Accurately Predict Searcher Preferences." In *2024 International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 2024. https://www.microsoft.com/en-us/research/publication/large-language-models-can-accurately-predict-searcher-preferences



Ever since this innovative work became known, other search companies are looking to copy what Microsoft Bing has done. We know in personal conversations with local Australian search companies that they are actively exploring such an innovation.

The use of generative processes to enhance the training of retrieval systems has led to other research examining replacing humans in the testing of search engines. This includes work conducted by researchers and students working at ADM+S[9] to automatically generate queries that different users might enter into a search engine.

The most widely discussed means of integrating generative AI into search has been with the creation of multiple retrieval augmented generation systems, so called RAG systems. The best known of these is Bing Copilot. The extensive examination of RAG architectures[10] is an attempt by generative AI researchers to minimise the problems of hallucination by exploiting the relative accuracy of search engines in identifying relevant and authoritative content in response to a question or query. In essence the request issued to a RAG system is translated into one or more

[9]Alaofi, M., Gallagher, L., Sanderson, M., Scholer, F., & Thomas, P. (2023, July). Can generative LLMs create query variants for test collections? an exploratory study. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1869-1873). - https://dl.acm.org/doi/10.1145/3539618.3591960
[10]Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, pp.9459-9474. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

queries which are run against a traditional search engine. The top matching results are retrieved by the search engine and then given to a generative language model to produce some form of synthesised content. RAG has become a very widely studied architecture.

Other well known RAG based search engines include Perplexity or Claude, Google of course is also developing its own search solution that will incorporate generative methods.

There is another notable trend in the work of search engine developers in the area of so-called vector retrieval. Here LLMs are used to represent documents and queries in a more abstract semantically oriented fashion. The use of this aspect of LLMs has been found to enhance the accuracy of traditional search. An example of this in the commercial market was the recent announcement by RedBubble that they have integrated factor retrieval into parts of their search engine using a technology created by Marqo. RedBubble developers presented their work at a tech talk in Melbourne in April 2024[11] where they described use of a foundation model that is a mixture of image and text analysis that is enabling a higher quality search on their art search services. A number of other search companies based in Australia have privately told us that they are actively pursuing adaptations of search technologies that incorporate large language models or generative AI processes.

18. Will the integration of generative AI into search engine services lead to new or additional monetisation strategies in general search services beyond an advertising-based model?

Like other aspects of the emerging generative AI ecosystem, the impact on search engine monetisation strategies is uncertain.

The integration of generative AI into search may augment the predominant advertising monetisation strategy - for instance where product placements are injected or integrated into conversational chat interfaces (as is done in Bing's GenAI search) this can make the distinction between organic and sponsored results less clear. It is not yet clear what flow-on impacts this might have on e.g. advertiser ROI (e.g. reduced user click-through could occur if the generative AI model summarises the contents of advertiser pages), advertiser trust in the search platform (e.g. generative AI confabulations visible alongside advertiser content might risk brand damage), or consumer trust in search results (e.g. consumers might decide not to trust any links from a generative AI enabled search platform). Another conceivable flow-on effect is that the metrics by which exposure is measured (e.g. views, click through - which are tied to advertiser payment) might need to be adapted for generative AI interfaces such as conversation.

Similar trends have played out with Amazon Alexa's smart assistant / voice platform since ad-based services began to be introduced circa 2018. For instance, research on this emerging smart-speaker advertising economy found that the increased interactivity of smart-speaker

[11]Meeting held in April 2024 https://www.meetup.com/melbourne-search/

advertisements can lead to higher brand and product recognition[12]- it is likely that generative AI integration will have analogous yet-to-be-discovered effects on advertisement effectiveness, design, and consumer use of search platforms.

The ability of generative AI systems such as LLMs to 'fill in the blanks' and make broad inference leaps might increase the viability of privacy preserving advertisement-funded search engines, such as DuckDuckGo . This search engine places sponsored search results based on search keywords, avoiding extensive user tracking. LLMs might conceivably enrich the ability to infer high relevance sponsored product placements from such limited data as keywords. For example, text-to-image generative AI platforms already routinely feature services that convert natural language queries to keywords, and vice-versa. This sort of prompt optimization model could in principle be applied to the problem of inferring ad relevance based on limited search queries while preserving user privacy. This might in turn increase the appeal of privacy-preserving advertisement funded search platforms.

The current search engine ecology has a limited number of 'premium' platforms where users pay to experience an advertising-free service (such as kagi.com). It is possible that the integration of generative AI services will increase the perceived value of this sort of search platform to a degree that paid ad-free services might see further uptake. However, it seems more likely at this point that any advantages a paid service might offer other than privacy (e.g. better user experience, novel generative AI enabled search methods) will readily be adopted or even improved upon by larger incumbent platforms with access to larger digital ecosystems (i.e. Google combining web search with video and location search through YouTube and Maps integration).

23. How easily can digital platform services integrate with generative AI and expand into providing general search? Would such expansion have any impact on competition among general search services?

It is worth remembering that the use of LLMs is expensive just simply in terms of compute cost. This impacts the slowness of generative AI. A key consumer demand of a search service is that it responds very quickly to any query issued. It is known that Google and Bing aim to respond to queries in under a quarter of a second. Research has found that if the engines do not respond that quickly, consumers use their search services less often[16]. While generative AI is being used in testing and tuning of the Bing general search engine, its use in the actual search process is a challenge. It is exceptionally hard to achieve quarter second response times when generative AI is part of the search system. This speed issue remains a key challenge, one that has not yet been technologically solved.

[12]Park, Kyuhong, et al. "Alexa, tell me more! The effectiveness of advertisements through smart speakers." *International Journal of Electronic Commerce,* 26.1 (2022): 3-24.
[13]https://duckduckgo.com/
[14]Mo, Wenyi, et al. "Dynamic Prompt Optimizing for Text-to-Image Generation." arXiv preprint arXiv:2404.04095 (2024).
[15]https://kagi.com/
[16]Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. 2013. Online controlled experiments at large scale. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). Association for Computing Machinery, New York, NY, USA, 1168–1176. https://doi.org/10.1145/2487575.2488217

24. How do consumers evaluate the quality of general search services?

The key metric for those looking for information from search engines is that they can find what they need, and fast[17]. This is true regardless of what type of search they are engaging in–navigational (looking for a particular website), transactional (looking to complete a web-enabled task, such as watching a video or finding a gaming server), informational (looking for information from one or more sites)[18], or browsing based (looking for a collection of information to browse)[19]. Searchers show some level of confirmation bias in assessing search results–preferring search results that align with their beliefs, but this is not universal and can be mitigated[20].

While fast results that conform to existing beliefs and provide immediately useful information are the main evaluation criteria searchers use to assess search services, there are secondary features that are under-recognised but valued by searchers. A diversity of search results allows searchers to learn new things, experience serendipity[21], enhance their creativity[22] and change views[23], each of which searchers have described as useful or even delightful. As such, prioritising only fast, immediately useful results will not serve all needs, and is likely to have flow on effects for creativity, polarisation and productivity.

[17]Connaway, L. S., Dickey, T. J., & Radford, M. L. (2011). "If it is too inconvenient I'm not going after it:" Convenience as a critical factor in information-seeking behaviors. *Library & Information Science Research*, 33(3), 179-190. https://doi.org/10.1016/j.lisr.2010.12.002
[18]Broder, A. (2002). A Taxonomy of Web Search. ACM SIGIR Forum, 36(2), 3-10.
19McKay, D., Chang, S., Smith, W., & Buchanan, G. (2019). The Things We Talk About When We Talk About Browsing: An Empirical Typology of Library Browsing Behavior. J*ournal of the Association for Information Science and Technology*, 70(12), 1383-1394. https://doi.org/10.1002/asi.24200
[20]Rieger, A., Draws, T., Theune, M., & Tintarev, N. (2021). This Item Might Reinforce Your Opinion: Obfuscation and Labeling of Search Results to Mitigate Confirmation Bias. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, Virtual Event, USA. https://doi.org/10.1145/3465336.3475101
[21]Makri, S., Blandford, A., Woods, M., Sharples, S., & Maxwell, D. (2014). "Making my own luck": Serendipity strategies and how to support them in digital information environments. *JASIST*, 65(11), 2179-2194. https://doi.org/10.1002/asi.23200
[22]McKay, D., Makri, S., Chang, S., & Buchanan, G. (2020). On Birthing Dancing Stars: The Need for Bounded Chaos in Information Interaction. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, Vancouver BC, Canada. https://doi.org/10.1145/3343413.3377983
[23]McKay, D., Makri, S., Gutierrez-Lopez, M., Porlezza, C., Macfarlane, A., Cooper, G., & Missaoui, S. (2024). I'm the same, I'm the same, I'm trying to change: Investigating the role of human information behavior in view change. *Journal of the Association for Information Science and Technology*.