Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers
Sheffield, UK / 30 Oct - 2 Nov 2024

SCROLL, PRINT, ALGORITHMICALLY CLUSTER: A CO-ANALYSIS APPROACH TO EXPLORE THE INTERPLAY BETWEEN USERS, PLATFORMS AND ALGORITHMIC MODELS ON INSTAGRAM

Nicholas Carah
The University of Queensland

Maria Gemma Brown
The University of Queensland

Rani Tesiram
The University of Queensland

Hine Tapairu Kahukura
The University of Queensland

Lisa Enright
The University of Queensland

Kiah Hawker
The University of Queensland

Background

Instagram turns 15 in 2025. For users who have been on the platform for much of that time, their profiles are a biography of their own lives, their visual practices and the platform itself. Instagram is ‘more than one thing’: a platform, a database of images and videos, an interface for posting, swiping and tapping, algorithmic models that classify and curate (Leaver et al. 2020). Instagram has industrialised and platformised the everyday practices of photography for its billions of users (Authors). Our posts and stories are both an archive of our lives and visual practices, but also part of the historical socio-technical process of assembling image data sets for training machine vision systems.

Suggested Citation (APA): Carah, N., Brown, M.G., Tesiram, R., Kahukura, H., Enright, L., & Hawker, K. (2024, October). *Scroll, Print, Algorithmically Cluster: A co-analysis approach to exploring the interplay between users, platforms and algorithmic models on Instagram*. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

This paper presents the final part of a multi-year project where we use a combination of cultural and computational methods to explore the relationships between our everyday image-making practices and the algorithmic models of instagram (Authors). In the earlier phases we developed a purpose-built machine vision system that approximated the capacity of Instagram to cluster images together based on visual patterns. We have 'critically simulated' the capacity of machine vision models to recognise and reproduce the patterns in the visual cultures of Instagram including images posted on hashtags relating to cultural events like music festivals, food and dining scenes, and intimate publics surrounding alternative art, culture and fashion. We have also explored the capacity of machine vision models to recognise visual patterns in advertisements on Instagram. Our aim in this larger project is not only to describe the visual practices of users, but also to enable a speculative exploration of how Instagram uses machine vision to inform the recommendation engines that curate our feeds and target advertising.

The Study

In this paper we present findings from a study with 25 participants who have used Instagram for five or more years as part of their professional or creative practices. Participants include photographers, designers, musicians, influencers and others who work in the cultural industries. We begin by asking participants to download their archive of images they have posted to their Instagram profile or stories and share it with the researchers, using the personal information download tools provided by Meta. Their collections range in size from 500 up to more than 10,000 posts and stories.

We process their archive by randomly selecting 500 images we print as photographs and uploading all their images to our purpose-built machine vision system (Tan 2020). Our system consists of a three-step data processing pipeline: (1) embedding, where images are mapped into numeric feature vectors, (2) clustering, where images are associated with each other based on similarities perceived by the algorithms by computing distances in their numerical embeddings, and (3) creating a visualisation, where clusters of images are presented in an interface for exploration and analysis.

Figure 1 shows the visualisation interface featuring a cluster of images from the Instagram profile of one of the researchers. On the left hand side are clusters of images, with the number in the pink box indicating how many images are in that cluster. If the user clicks on a cluster, they are taken into a second layer of clusters. As they move through the clusters they can see how the machine vision system has grouped their images in increasingly fine-grained ways.

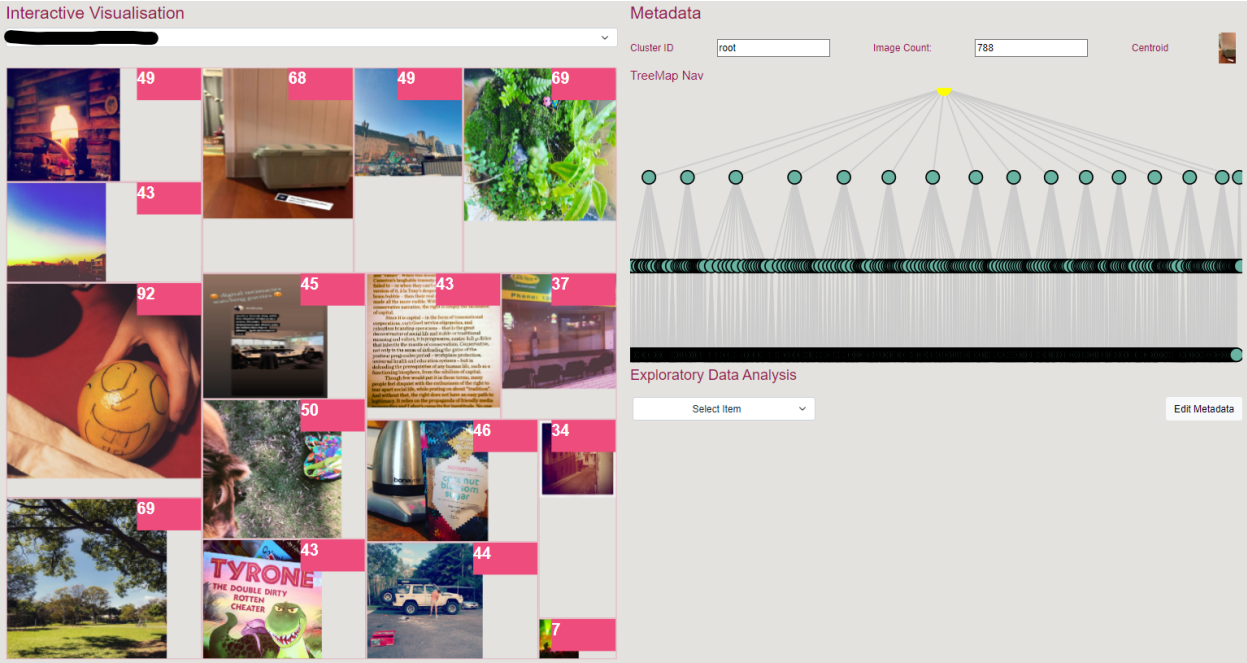

Figure 1: the user interface of our machine vision system

Once the images are printed and clustered, we meet with the participant for a co-analysis interview that unfolds in three steps.

Firstly, we undertake a 'scrollback' through their Instagram profile (Robards and Lincoln 2017). Through this process we explore with them what they post to their Instagram profile, whether their practices have changed over time, and how that relates to changes in their lives, professional or creative practices, and the platform.

Secondly, we give participants the stack of 500 photographs we have printed from their profile and stories posts. Where the first step only explores the archive of posts on their profile, this step introduces the ephemeral posts from their stories feed that are provided by Instagram in the data download. We ask participants to organise their images in any way they like. This process takes time, and as an order emerges, we discuss with participants their choices that prompts reflection on their digital archive, disconnected from the chronological order of their profile.

Thirdly, leaving their manually-sorted images on the table we ask participants to open and explore the machine visualisation, describing as they go how they think the algorithmic model has organised their images, whether they can explain its logic, and whether its clustering reflects their own.

Scrolling, printing and algorithmically clustering Instagram posts

Our novel integration of data donation, computational and scrollback interview methods brings to the fore the multiple uses and values of images on Instagram. By working with participants to download, scroll, print and algorithmically cluster their instagram profiles and stories we develop an intimate form of platform biography (Burgess and Baym 2020). This situated method of co-analysis contributes to efforts to investigate the

algorithmically curated and ephemeral nature of Instagram's culture and platformization (Bainott et al. 2021, San Cornelio and Roig 2020). Participants describe how their practices of expressing themselves, often over the course of a decade of their lives, are entangled with the development of platform interfaces, features, algorithms and business models.

By moving back and forth between their profile, the printed out images and the machine vision system we explore their images on, off and underneath the platform interface. The process stimulates the algorithmic imaginary (Bucher 2017, Bishop 2018) of participants as they offer a range of creative, thoughtful and frustrated responses to how the curation of their images by algorithmic models reflects and diverges from their own practices. Some organise their printed images chronologically, sorting them into periods that reflect their lives; some around objects, scenes or places like images; some around practices like travel, work, and social events. Then when they move to the machine visualisation some express a sense of eeriness or creepiness about patterns in their images the machine vision system sees that they hadn't recognised. Some are struck by the childlike and banal nature of the model's choices. They offer rich descriptions of how machine vision doesn't 'get' narrative or context, even though it successfully identifies patterns. Some express a sense that their Instagram profile has two 'audiences': the people they imagine viewing and interacting with their stories, and the algorithmic systems that sort and cluster them.

Over the course of a decade and a half our everyday practices of making and sharing images have been integrated into the industrial process of training machine vision models. This training period has been a formative stage in the emerging development of models that can synthetically generate and augment images at scale. Machine vision tools are useful not only for exploring platform-scale questions about classification and recommendation, but also as part of participatory approaches that elicit and develop the algorithmic imaginaries of Instagram users. We propose that the intimate platform biographies we develop with participants are valuable in making sense of how our image making and sharing practices are not only archives of our personal lives, but archives for the industrial process of training machine vision models.

References

Bainotti, L., Caliandro, A., & Gandini, A. (2021). From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods. *New Media & Society*, *23*(12), 3656-3676.

Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New media & society*, *21*(11-12), 2589-2606.

Bucher Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information Communication & Society* 20 (1): 30–44.

Burgess, J., & Baym, N. K. (2022). *Twitter: A biography*. NYU Press.

Cornelio, G. S., & Roig, A. (2020). Mixed methods on Instagram research: Methodological challenges in data analysis and visualization. *Convergence*, *26*(5-6), 1125-1143.

Leaver, T., Highfield, T., & Abidin, C. (2020). *Instagram: Visual social media cultures*. John Wiley & Sons.

Robards, B., & Lincoln, S. (2017). Uncovering longitudinal life narratives: Scrolling back on Facebook. *Qualitative Research*, *17*(6), 715-730.

Tan, X.Y. (2020). Image Machine. *GitHub.* Retrieved from: https://github.com/qut-dmrc/ImageMachine