What scientific research programme does DSA Article 40 enable?

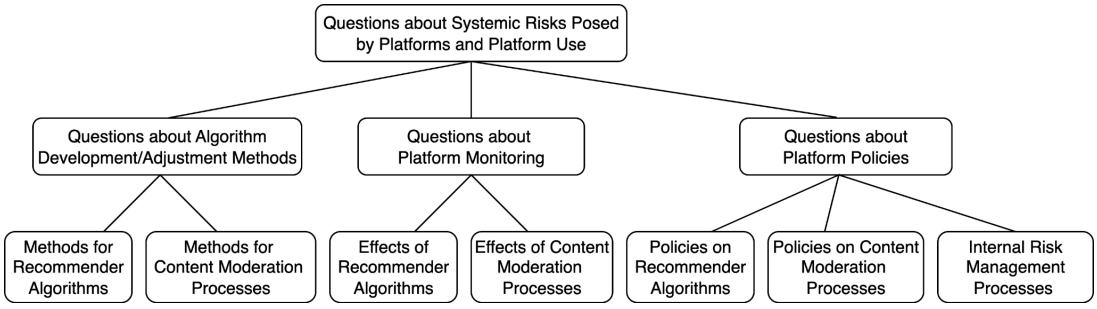
Alistair Knott[1,2], Przemyslaw Grabowicz[1,3], Dino Pedreschi[1,4], Susan Leavy[1,3], David Eyers[1,5], Andrew Trotman[1,5]

[1]Social Data Science Alliance [2]Victoria University of Wellington [3]University College Dublin [4]University of Pisa [5]University of Otago

In this seminar we will introduce a new grouping of researchers, the Social Data Science Alliance (SDSA), which was established earlier this year, to help coordinate the studies of Very Large Online Platforms and Search Engines (VLOPSEs) conducted under the new access provisions defined by the DSA's Article 40. There are already several venues for this coordination, including the DSA Observatory at the Universiteit van Amsterdam, the DSA40 Collaboratory at the Weizenbaum-Institut in Berlin, and the European Centre for Algorithmic Transparency (ECAT) in Madrid. The SDSA's focus, alongside these initiatives, is to help establish the broad outlines of the new area of science that will be founded on the new access methods created by the DSA's Article 40. The Social Data Science Alliance currently has 79 individual members, and 5 supporting organisations; details can be found here.

Our first project was to survey members, to find out what research questions they regard as the most important ones to ask under Article 40(4). In our seminar we will present the results of this survey. Our aim is twofold. First, we *enumerate* the research questions provided by our respondents, and note areas of consensus. Respondents suggested many questions—there are well over a hundred. Second, we *structure* the set of research questions, to sketch the shape of the scientific research programme our respondents collectively identify. Questions are of several types and subtypes, and are about platform users and platform content of several types. We propose a *taxonomy* of research questions reflecting this structure.

The top level of our taxonomy is shown here. All suggested questions are about systemic risks, of course, as access under DSA 40(4) is for the purpose of studying systemic risks. Our



taxonomy focuses on risks associated with three basic *activities* conducted in platforms, each relating to platform algorithms. Our focus is on algorithms and policies, because platforms are controlled by them.

- Technical teams **develop** and **deploy** algorithms, and regularly train and adjust them. The methods for training and adjusting algorithms define the actions available to the platform, to influence how the platform operates.
- The platform is intensively **monitored**, to observe the *effects* of the deployed algorithms on its traffic and users, through a variety of rigorously empirical methods. This complex monitoring operation provides *feedback*, which informs how the algorithms are adjusted.
- At the highest level, platform executives create **policies** about *how* algorithms should be optimised and adjusted, in the light of feedback. These policies define how the platform should *ideally* operate, and how it should respond to contingencies as they arise.

As may be clear from this analysis, we view platforms as complex dynamical systems. We think of platform management quite technically, as the task of *controlling* a complex dynamical system, through actions taken on platform algorithms, that retrain them, or adjust them.

As is quite conventional (see, e.g., Allen and Lawson, 2024), we distinguish two key types of platform algorithm.

- **Recommender algorithms** push content at platform users, ranking items for presentation in content feeds, or lists of search results.
- **Content moderation algorithms** withhold or contextualize content from users, by identifying harmful or inappropriate content in various categories, and moderating it in various ways, e.g., by adding contextual information.

In our taxonomy, each of the three top-level activities regarding algorithms is further subdivided into activities regarding recommender algorithms, and activities regarding content moderation algorithms. These algorithms are separately *developed* and *adjusted*. Their effects are separately and jointly *monitored*. (What content is *removed* or *moderated*, for which users, in which situations? What content is *seen*, by which users, in which situations?) Finally, *policies* are formulated for how recommender and content moderation algorithms should be adjusted, in the light of feedback from monitoring. These activities define the high-level structure of our taxonomy. We believe they are helpful in organising the set of research questions provided by our respondents.

There's one other useful component of structure in our taxonomy. Respondents had questions about many specific categories of user ('young adults', 'immigrants', 'people in economic precarity'), content ('dangerous speech', 'quality news content') and world events ('pandemic', 'election', 'political crisis'). Questions often feature particular combinations of items from these sets. To represent these combinations productively, certain intermediate nodes in our taxonomy use variables that generalise over items of each type. The typology of user groups (U), content categories (C), and world event types (E) can then be defined separately.

In fact, many respondents asked substantive questions about how *platforms* classify users, content categories and world event types. In many cases, platforms are likely to have the richest operational category definitions. Formulating precise research questions often requires

knowledge of these operational definitions. An outstanding question for discussion is whether platforms should provide category definitions as part of the ‘catalogues’ they supply to prospective researchers, under the terms of the DSA’s Delegated Act on Data Access.