The Algofeed project. A methodological proposal to assessing the effects of algorithmic recommendations on platformized consumption

Gabriella Punziano[1] , Alessandro Gandini[3] , Alessandro Caliandro[2] , Massimo Airoldi[3] , Giuseppe Michele Padricelli[1] , Suania Acampa[1] , Domenico Trezza[1], Noemi Crescentini[1] , Ilir Rama[3]

[1] Department of Social Science, University of Naples Federico II, Italy, [2]Department of Political and Social Science, University of Pavia, Italy, [3] Department of Social and Political Science, University of Milan, Italy.

Abstract

This paper outlines the construction phase of the consumption profiling, algorithmic awareness, and digital literacy tool within the ALGOFEED project. ALGOFEED, funded as an Italian Research Project of Significant National Interest (PRIN), seeks to illuminate the socio-cultural impacts of platform-based feedback loops while developing a novel methodological and theoretical framework for the sociological examination of consumer culture algorithms. The project aims to produce unique empirical insights into cultural consumption patterns on digital platforms in Italy. Specifically, ALGOFEED investigates the effects of automated content recommendations on platforms like TikTok and YouTube on individual and collective cultural consumption behaviours over time. The primary objectives include acquiring distinctive empirical findings on platform-driven cultural consumption in Italy, advancing a fresh methodological and theoretical approach for studying algorithmic consumer culture sociologically, and fostering algorithmic awareness among Italian consumers through impactful diffusion activities grounded in a transformative paradigm.

In this study, the authors introduce the empirical tool designed to detect the feedback loop and define its components: digital skills, Platform Usage Type, and Algorithmic Awareness.

Keywords: *Algofeed; Algorithmic Recommendations; Algorithmic Awareness; Digital Skills; Cultural Consumption*

1. Introduction

The ALGOFEED project is an Italian Research Projects of Significant National Interest (PRIN) which has the general objective of shedding light on the socio-cultural effects of platform-based feedback loops, developing a new methodological and theoretical framework for the sociological study of consumer culture algorithms capable of producing unique empirical results on the cultural consumption platform in Italy. Feedback can be defined as "the property of being able to adjust future conduct by past performance" (Wiener 1989: 33). In the case of interactions between platform users and recommender algorithms, these feedback-based learning happens on algorithmic outputs expose consumers to selections of content they "may also like"; on the other hand, based on consumers' input data, machine learning systems iteratively decide how to update their future outputs, aiming to get aligned with consumers' expectations. This dynamic process of mutual influence between recommender algorithms and platform consumers will likely establish "feedback loops" (Jiang et al., 2019; figure 1). The social science literature has widely discussed platform-based feedback loops. Research has mostly focused on how online algorithms risk producing a polarized public opinion (Moller-Hartley et al., 2021), often overemphasizing the power of algorithms in a technologically deterministic fashion (Bruns, 2019). Conversely, the active role of the "human in the loop" has received less attention – apart from recent works highlighting how platform users may "resist" algorithmic power and showcase various levels of "algorithmic awareness" correlated with socio-demographic variables (Gran et al. 2020). Recent articles stress how the effects of platform-based feedback loops on online users go well beyond the political sphere, and directly concern consumer culture and habits more broadly intended (Fourcade and Johns, 2020). Scholars in the context of cultural consumption studies have noted how recursive interactions between consumers and recommender systems are likely to strengthen past consumption patterns, eventually "normalizing" them (Hallinan and Striphas 2016). This techno-social process is believed to induce the unaware adaptation of consumer tastes and identities to automated recommendations. Yet, sociological research still rests, for the most part, on theoretical speculations. This lack of empirical evidence on platforms' "feedback culture" is due to epistemological and methodological limitations: platform algorithms are opaque "black boxes" that are "immune from scrutiny" (Pasquale, 2015) due to both technical and corporate reasons. Moreover, algorithmic outputs are highly personalized and changeable over time, and thus very difficult to track. In the end, users have a limited awareness of algorithmic systems' activities, further complicating the study of their interactions with them. For all these reasons, scholars and practitioners have frequently launched calls for "opening the black box" and making platform algorithms accountable (Ananny and Crawford 2018). Considering this background, the main aim of this research project is to fill this empirical gap. How does the automated content recommendation on digital platforms affect users' individual and collective consumption patterns over time? To answer this research question, a mixed methods strategy will shed light

on the socio-cultural effects of platform-based feedback loops. Given the breadth of the topic addressed, preparing a plan for empirical advancement involving a case study limited to the Italian context was necessary. In this regard, attention focused specifically on platformed entertainment and music consumption feedback loops. From a methodological standpoint, the research path articulates through a sequential mixed-methods research design (Teddlie & Tashakkori, 2011) aimed at empirically investigating a) the joint evolution of individual entertainment and music consumption trajectories and personalized algorithmic recommendations over time, b) aggregate content recommendation trends at the platform level, and c) Italian consumers' awareness and understandings of platform-based recommendation systems. To better understand how algorithmic systems recursively shape cultural consumption, longitudinal and cross-platform examinations (Rogers 2017) capable of grasping at once individual consumers' trajectories and understandings and the macro unfolding of consumption and recommendation trends (Airoldi 2021) are required. The research focuses on two widespread video sharing platforms presenting personalized content recommendations, that is, YouTube and TikTok. The mixed-methods research design combines three concatenated operational steps of data collection: 1) a short pre-tracking survey, 2) longitudinal digital tracking, and 3) qualitative follow-up interviews. Illustrated the general methodological path of the project, in this work, we will specifically present point 1, which is the empirical tool for detecting the feedback loop and the dimension of the tool. The three dimensions are: Digital Skills, Platform Usage Type and Algorithmic Awareness.
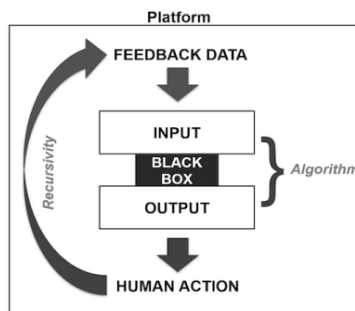
Figure 1 - Feedback loops. Source: our elaboration

2. The construction of the pre-tracking survey questionnaire

The construction of the questionnaire   is aimed at profiling the sample across three specific dimensions of significance: the first concerns Digital Skills (DS), a framework on digital competencies widely experimented in literature: despite extensive reference to some existing scales in the literature, many indicators have been custom-built or modified for the research project. The second operationalizes dimensions related to the use of social platforms and their cultural consumption (Platform Usage Type, PUT). The third macro-dimension pertains to

Algorithmic Awareness (AA), specifically the questionnaire section dedicated to questions probing individuals' understanding of the algorithmic processes governing the Internet and social media platforms. This latter area, originally encompassed within the DS domain, has been separated from it conceptually because it serves as a "bridge" between skills and usage experiences. Therefore, as described later, a dedicated section has been allocated to it in the document. The empirical advancement concerning the first dimension of analysis is aimed at describing the framework on digital competencies with a specific focus on the operational phases in defining the universe of skills that users possess to interact with the world of social media; the second aims to describe the operational steps that led to the operational definition of sub-dimensions and indicators regarding the type of platform usage and media consumption practices; finally, the third concerns the description of the dimensions put in place to identify the relationship between the user and algorithmic processes, across different aspects: cognitive, procedural, behavioural, and affective. The objective is to construct   the type of user defined by the experience with social networks, particularly the use of the platforms under study, TikTok and YouTube.

2.1. The construction of the pre-tracking survey questionnaire

The Strategic Program for the Digital Decade guides Europe's digital transformation establishes objectives in digital skills, digital infrastructure, digitalization of businesses, and public services. With particular reference to digital skills, the European framework DIGCOMP (figure 2) establishes a standard reference for digital skills by identifying and describing what it means to be competent in using digital technologies. DIGCOMP outlines key competencies essential for effective interaction with technology in various areas of life, such as education, work, and social participation. It aims to promote and enhance digital skills at all levels. According to the European framework, digital competence is "the ability to use digital technologies confidently and critically". According to this framework, and following Vuorikari, Kluzer and Punie (2022) digital skills are divided into four main areas (see figure 2).

Figure 2 – Digital Competence (soure DigiCompEdu 2.21)

A summary is shown in the table below.

Table 1. Areas and indicators of the dimension DS

Content Creation Skills	Digital Communication and Interaction Processes	Security
1. Creation and Modification Of Content Produced By Third Parties: • Ability to create and modify content such as texts, images, videos, music. • Level of k nowledge of software tools for content editing (e.g., text editor, video and image editing software). • Frequency of modifying third-party content (e.g., for work, hobbies). • Level Of knowledge of licenses applicable to the use of digital content (e.g., creative commons licenses).	*1. Creation and management of a profile/account on digital platforms:* • Frequency of creating or updating profiles on various digital platforms (e.g., social media, forums). • Level of detail and care in profile maintenance (e.g., type of personal information entered in the profile, frequency of publication, type of content posted, periodic profile updates). • Knowledge/Use of profile privacy and security settings (e.g., profile visibility controls, post privacy settings, contact and friend management, block and blacklist: data sharing controls with third-party apps; location settings, cookie and tracking management, etc.). *2. Knowledge of good practices in online communication:* • Knowledge of netiquette rules (evaluable through concrete examples such as: do you recognize yourself in this sentence?).	*1. Knowledge of online security practices and data protection:* • Level of knowledge of good online security practices (secure passwords, encryption, safe browsing, phishing recognition, etc.). • Browsing modes (private or incognito offered by browsers). • Frequency of applying online security practices in daily life (e.g., use of VPNs, password managers, website certificate verification). • Ability to implement personal data protection measures (e.g., encryption, backup, privacy settings). • Identification of potential risks (such as phishing messages).

---

[1] Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

The areas of this first dimension are addressed in the survey form through specific questions operativized with accurate scaling techniques. Using Likert scales, participants are first asked how frequently they browse the internet, which social platforms they use, and for what purposes. Moreover, through several questions, participants are asked to self-assess their proficiency in typical operations performed while browsing the internet. Also, within the dimension related to digital skills, participants are asked, through simulations of content consultation, to distinguish between messages created by individuals or bots, and to judge which types of behaviors they consider appropriate to adopt on social media.

2.1.1 Platform Usage Type

The construction of questions regarding the platform's type of use and cultural consumption is inspired by the work of Boyd and Ellison (2007) on the structure, objectives, and user types of Social Networks. The articulation of dimensions and their respective indicators in this specific section took into consideration two critical concepts in the study of social media and related practices: "contexts" and "users" (Bennato 2008; Boccia Artieri and Marinelli, 2018). The "contexts" represent the usage environments of the users and are linked to the tools of users' digital experience and the time spent. Essentially, it concerns their presence on Social Networks (General Platform Presence) and, specifically in the research context, their usage experience of YT and TT platforms (Specific Platform Presence). In the former, indicators are defined primarily to detect the essential characteristics of the user: in addition to the device usually used to access online platforms, they are asked about the main social networks they are subscribed to, the type of subscription activated, and the possible extent of expenditure incurred for content consumption. The second dimension is defined by indicators (Table 2) that address this aspect more specifically, focusing precisely on the two platforms of interest. The second macro-dimension of the questionnaire ("users") is oriented toward users and their modes of consuming and creating content and managing social networks. Specifically, three dimensions have been defined: Interactions, Networking, and Contents. The first investigates the prevalent modes of interaction with content produced by others, or towards preferred and less appreciated content. This would also allow observing some proxies related to the phenomenon of filter bubbles, presumably characterized by attitudes of constant acquiescence (Spohr, 2017). The second dimension deals with the breadth of contact networks on the two platforms, inbound (followers) and outbound (pages, users, and channels "followed"). The third dimension pertains to Contents, i.e., the mere enjoyment about prevalent themes and the creation - if any - of content, focusing on the propensity to realize or consume viral content. In this regard, a question about the user's role on the platform through a continuum from "simple content observer" to "professional content creator" better clarifies this semantic dimension.

Table 2. Areas and indicators of the dimension PUT

Contexts of Use	Users (The "Prosumer" User and Their Network)
1 General Platform Presence	*1 Interactions*
• Most used devices for browsing	• Primary mode of interaction with content (like, share, comment, private resharing)
• Purpose of Internet usage	• Sentiment of interaction (support, neutral observation, criticism, conflict, etc.)
• Most used Social Networks (SN)	
2 Specific Platform Presence	*2 Networking Network breadth*
• Type of YT/TT Account	• Number of Followers
• Membership seniority	• Number of Following
• YT/TT access time	• Number of Channels followed
• Time spent	*3 PCC Contents*
	• Prevalent consumption categories
	• Prevalent creation categories
	• Virality: Engagement for viral content (consumption/creation)

The operational definition of the described indicators leads to the development of questions within the questionnaire that focus on the longevity of users' social media presence on the platforms chosen as case studies (YouTube and TikTok). Through scaling techniques, participants are asked if they have an active account on both platforms, how frequently they access them, and how often they produce content. They are also asked about the number of followers per platform and the type of content they view or produce (video streaming, podcasts, vlogs, challenges, etc.) by theme (politics and society, TV series and shows, travel, wellness, recipes, cars, etc.).

2.1.2 Awareness Algorithmic

Algorithmic awareness (AA) is a concept that refers to individuals' understanding and awareness of how algorithms operate and are employed in various digital and social contexts.

In a single set of questions operationally defined through a Likert scale, questionnaire participants are asked to indicate their level of agreement with the characterizations of the indicator presented in   Table 3; the works of Zarouali, Boerman, de Vreese (2021) and Felaco (2022) were used on AA dimensions which can be articulated as follows:

Table 3. Areas and indicators of the dimension AA

cognitive dimension	procedural dimension	behavioral dimension	affective dimension
1. Awareness of content filtering operation. •Algorithms are used to recommend multimedia content on the platform. •Algorithms are used to prioritize certain multimedia content over others. •Algorithms are used to personalize specific content on the platform.	*1. Awareness of automated decision-making process.* • Algorithms are used to display multimedia content on platforms based on automated decisions. • Algorithms do not require human judgments in deciding which multimedia content to display on the platform. • Algorithms make automated decisions on which content I can see on the platform. *2. Awareness of risks and ethical issues.* • It is not always transparent why algorithms decide to display certain content. • The content recommended by algorithms on the platform may be subject to human biases and stereotypes. • Algorithms use personal data to recommend specific content on the platform, which affects online privacy. • Algorithms may exclude a user from seeing content other than their preference profile.	*1. Awareness of human-algorithm interaction.* • The content recommended by algorithms on the platform depends on users' online behavior on that platform. • The content recommended by algorithms on a platform depends on users' online behavioral data. • The content recommended by algorithms on a platform depends on users' available online data. *2. Development of bottom-up tactics.* • Optimizing web content to achieve a higher placement in the news feed. • Choosing to follow or unfollow certain accounts or hashtags to influence the composition of the news feed. • Modifying or creating content in specific ways to be favored by recommendation algorithms. • Exploiting feedback loops to reinforce specific patterns or outcomes.	*1. Positive, negative, indifferent emotional reactions.* • Feeling frustrated when not understanding why an algorithm shows certain content. • Feeling curious when encountering unexpected content suggested by algorithms. • Feeling indifferent about why algorithms select certain content. *2. Critical reflection generated by unexpected outcomes:* • Being prompted to reflect on algorithmic logics when the process suggests unexpected content. • Tendency to ignore unexpected outcomes of algorithm recommendations without seeking to understand the reason.

3. Conclusion

The paper aimed to illustrate the operational phase of the methodological framework at the first level of the ALGOFEED study, highlighting the importance of developing appropriate research tools to investigate the complex dynamics of interaction between users of digital platforms and

recommendation algorithms. The developed survey represents a significant step towards a thorough understanding of how users' technical and digital skills and platform usage practices influence and are influenced by algorithmic recommendations. A fundamental aspect emphasized in the paper is the importance of identifying and selecting representative user profiles that can provide a comprehensive overview of various interaction experiences with recommendation algorithms. This approach enabled us to grasp the complexity of the relationships between users and algorithms and analyze the various dimensions involved in detail. The empirical path useful for achieving this goal consists of creating synthetic indices, each corresponding to ideal-typical user profiles. This process is facilitated through analysis techniques aimed at developing models based on the combination of indicators selected during the operational data collection phase through interrogation, such as PLS Path Modeling. Therefore, the objective is to establish a typology categorizing online users types and magaing their exposure to feedback-loop processes. This pathway becomes viable upon integration with the remaining phases of the research. Subsequently, during the tracking and the interview stages, insights into the cultural consumption patterns of these user types emerge, facilitated by an understanding of their reactions to recommendations and their level of algorithmic awareness, obtained through the application of regression and clustering analysis. Within the scope of the initial research phase, the objective is to formulate three synthetic indices through the combination of indicators of dimensions related to digital skills and PUT. By combining these indicators (Figure 3), the synthesis of the audience-oriented user is postulated. This profile represents a user with medium-low content processing capabilities engaging on digital platforms as part of a diverse audience seeking entertainment content. Conversely, at the opposite end of the spectrum, we find the creator user - a user with high content processing capabilities carefully produced and promoted for professional and career advancement purposes in the digital creation sector. In the middle lies the uncertain user, straddling the characteristics of the two previous types, with a medium level of content processing capabilities and currently in a phase of consideration regarding a potential professional investment in the sector.
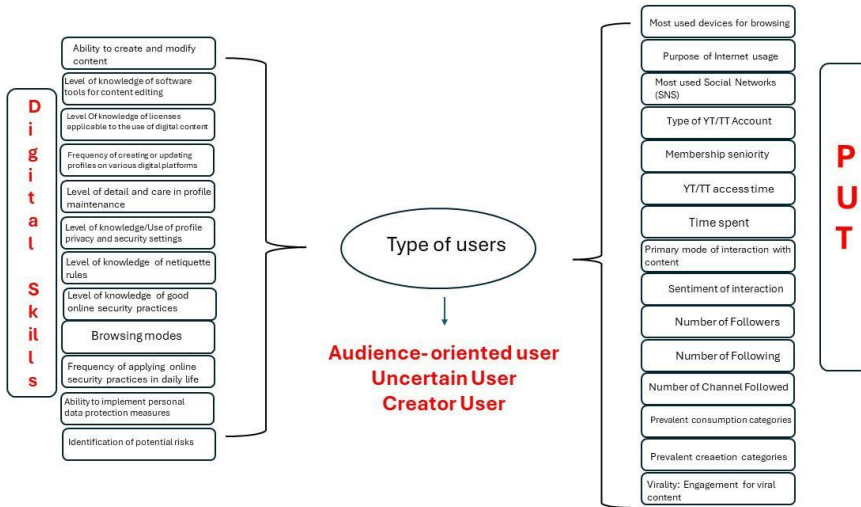
Figure 3 – The typological scheme obtained by the PLS-P

The crucial part of this paper was methodological: the operational definition of dimensions as digital skills, algorithmic awareness, and cultural consumption on social platforms that required a rigorous review of existing literature tools and careful methodological reflection. The project's subject matter contains some unexplored knowledge spaces (e.g., digital skills related to social media usage), making it essential to consider new dimensions and indicators for profiling our objectives.

The study broke down platform usage into the contexts of social platforms, subscription methods, and user experiences. It emphasized the necessity of considering users' platform presence, digital navigation skills, and familiarity with social media. Algorithmic awareness was identified as a critical area, highlighting its complexity and the nascent state of research tools for operationalization. The analysis covered cognitive, procedural, behavioral, and emotional aspects of users' evaluation and response to algorithmic recommendations in digital environments.

In conclusion, while it is possible to say that, in the context of our project, it is still too early to assess the tool's actual reliability and the validity of the concepts and indicators defined, this exploratory proposal, which aims to open debates that would allow for reflection aimed at improving the presented measurement tool, remains of decisive importance.

References

Airoldi M. (2021). The Techno-Social Reproduction of Taste Boundaries on Digital Platforms: The Case of Music on YouTube. Poetics, 89.

Ananny M., Crawford, K. (2018). Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. New Media & Society, 20(3), 973–89.

Bennato, D. (2008). Le professioni del web, in Marco Pedroni e Paolo Volontè (a cura), La creatività nelle professioni, Bozen-Bolzano University Press, Bolzano, pp.75-88, 75.

Boccia Artieri, G., & Marinelli, A. (2018). Introduzione: piattaforme, algoritmi, formati. Come sta evolvendo l'informazione online. Problemi dell'informazione, 43(3), 349-368.

Bruns A. (2019). Are Filter Bubbles Real? Polity.

Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press.

Felaco, C. (2022). Lungo la scala di generalità: le dimensioni della consapevolezza algoritmica. Sociologia Italiana.

Fourcade M. & Johns F. (2020). Loops, ladders and links: the recursivity of social and machine learning. Theory and society, 49(5), 803-832.

Gran A.B., Booth P. & Bucher T. (2020). To Be or Not to Be Algorithm Aware: A Question of a New Digital Divide? Inf., Comm. & Society, 24(12), 1779-1796.

Hallinan B. & Striphas T. (2016). Recommended for You: The Netflix Prize and the Production of Algorithmic Culture. New Media & Society, 18(1), 117–37.

Jiang R. et al. (2019). Degenerate Feedback Loops in Recommender Systems. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 383–90.

Møller Hartley, J., Bengtsson, M., Schjøtt Hansen, A., & Sivertsen, M. F. (2021). Researching publics in datafied societies: Insights from four ap-proaches to the concept of 'publics' and a (hybrid) research agenda. New Media & Society, DOI: 14614448211021045

Ragnedda, M. (2018). Conceptualizing digital capital. Telematics and informatics, 35(8), 2366-2375.

Rogers R. (2017). Digital methods for cross-platform analysis. The SAGE Handbook of Social Media, 91-110

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. Business information review, 34(3), 150-160.

Teddlie C. & Tashakkori A. (2011). Mixed methods research. The Sage Handbook of Qualitative Research, 4, 285-300.

Wiener N. (1989). The Human Use of Human Beings: Cybernetics and Society. Free Association Books.

Zarouali, B., Boerman, S. C., & de Vreese, C. H. (2021). Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). Telematics and Informatics, 62, 101607.