

How to investigate algorithmic-driven risks in online platforms and search engines? A narrative review through the lens of the EU Digital Services Act

Cecilia Panigutti
cecilia.panigutti@ec.europa.eu
European Commission, Joint Research
Centre (JRC)
Ispra, Italy

Delia Fano Yela*
delia.fano.yela@ri.se
Research Institute of Sweden (RISE)
Lund, Sweden

Lorenzo Porcaro*
lorenzo.porcaro@uniroma1.it
Sapienza University of Rome
Rome, Italy

Astrid Bertrand
astrid.bertrand@ec.europa.eu
European Commission, Joint Research
Centre (JRC)
Brussels, Belgium

Josep Soler Garrido
josep.soler-garrido@ec.europa.eu
European Commission, Joint Research
Centre (JRC)
Sevilla, Spain

ABSTRACT

Algorithmic systems are increasingly integrated into digital services, where they shape user experiences, mediate access to information, and influence decision-making processes. As their societal impact grows, the field of *algorithm auditing* has emerged in recent years as a specialized area of study and practice focused on examining how these systems operate and assessing their broader impact on society. On the regulatory side, the European Union's Digital Services Act (DSA) introduces new obligations for digital services, including provisions that specifically address the algorithms employed by very large online platforms and search engines. This paper seeks to establish a connection between the field of algorithm auditing and the provisions of the DSA that specifically address algorithmic systems. We analyze the existing algorithm auditing literature and develop a taxonomy of *study designs*, defined as structured plans used to investigate risks associated with the use, design, and functioning of digital services's algorithms. We categorize these study designs based on their objectives in the broad risk management framework of the DSA, and identify key methodological steps and best practices. In doing so, this paper offers a guide for researchers, auditors, and digital services aiming to adopt rigorous approaches to algorithm auditing within the framework of the DSA.

*Work done when the author was at European Commission, Joint Research Centre (JRC)

Disclaimer: The views expressed in this article are purely those of the authors and may not, under any circumstances, be regarded as an official position of the European Commission.



This work is licensed under a Creative Commons Attribution 4.0 International License. *FAccT '25, June 23–26, 2025, Athens, Greece*
© 2025 Copyright held by the European Union.
ACM ISBN 979-8-4007-1482-5/2025/06.
<https://doi.org/10.1145/3715275.3732052>

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence; Machine learning**; • **Applied computing** → **Investigation techniques**.

KEYWORDS

algorithm auditing, digital services, systemic risks, online platforms, DSA, Trust and Safety, narrative review

ACM Reference Format:

Cecilia Panigutti, Delia Fano Yela, Lorenzo Porcaro, Astrid Bertrand, and Josep Soler Garrido. 2025. How to investigate algorithmic-driven risks in online platforms and search engines? A narrative review through the lens of the EU Digital Services Act. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, June 23–26, 2025, Athens, Greece. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3715275.3732052>

1 INTRODUCTION

Digital technologies have become deeply integrated into our everyday lives. Many of us rely on digital services to connect with friends and family, shop for goods and services, and stay informed on the latest news and events [5, 19, 35, 71]. Behind the scenes, digital services employ algorithms to personalize our online experiences, influencing the news we consume [7, 23], the content in our feeds [18, 111], and the products and services we are recommended [44, 54, 118]. Furthermore, these algorithms drive targeted advertisement [104], facilitate seamless interactions [123, 124], and enforce community guidelines [2, 75]. Consequently, the influence of algorithms extends beyond mere convenience and affects the choices we make, the information we encounter, and the products we consume. This sparks concerns regarding *algorithmic-driven risk*, i.e. risks associated with the use, design, and functioning of algorithms. For example, the dissemination of harmful content [46, 47], the amplification of disinformation [82, 109], the distortion of democratic processes and civic discourse [20, 61, 101], the perpetuation of harmful biases [13, 40, 67], and the impact of algorithms on users' mental and physical well-being [45, 77, 96].

On the regulatory side, the EU has taken a proactive stance to address the challenges posed by algorithmic-driven risks in the

digital landscape with the introduction of the Digital Services Act (DSA) [34]. The DSA establishes a set of due diligence obligations to all intermediary services that offer their services to recipients in the EU. In particular, the DSA applies to *online platforms*¹ and *online search engines*² employing an asymmetric approach that applies higher standards of due diligence to services with a larger user base³ reflecting their greater influence and potential to cause systemic risks. These specific categories of digital service providers are referred to as Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs)⁴ and are subject to a supervised risk management framework that requires them to assess⁵ and address⁶ the systemic risks that the use of their services and systems, including algorithmic systems, pose to society. In this context, the DSA identifies several algorithmic systems to which VLOPs and VLOSEs must pay particular attention. These include their recommender systems, content moderation systems, systems that apply and enforce their terms and conditions, systems used for the selection and presentation of advertisements, and their data-related practices⁷.

The systemic risks identified in the DSA include the dissemination of illegal content and any risk that has a negative effect on fundamental rights, civic discourse, electoral processes, public security, gender-based violence, minors, mental health, and physical well-being. Many of these risks can be either exacerbated or directly mediated by algorithmic systems. There is, in fact, a notable overlap between the algorithmic-driven manifestations of these risks and the concerns explored within the field of *algorithm auditing*, which is specialized area of study and practice aimed at evaluating harms emerging from the use of algorithms [6, 21, 74, 97, 113]. Algorithm auditing encompasses a variety of methodological approaches designed to examine how these systems operate, assess their impacts on society, and identify potential biases, discriminatory outcomes, or other unintended consequences. As such, it offers a valuable framework for addressing the DSA's call for comprehensive risk assessments and meaningful mitigation strategies.

Various stakeholders, such as researchers, practitioners, affected communities, policymakers, and industry experts approach algorithm auditing from diverse perspectives, each bringing different methodologies to the table. Previous efforts to classify algorithm auditing approaches have predominantly focused on *how* data was collected for auditing (e.g. using an API, web scraping, or synthetic accounts) [6, 50, 97]. While this framing was appropriate in a research environment defined by limited access to platform-held data,

the Digital Services Act (DSA) introduces a series of mechanisms that expand the possibilities for auditing algorithmic systems, including new data access rights for independent researchers⁸, a Transparency Database that collects content moderation decisions made by digital service providers⁹, and advertisement repositories¹⁰. These tools create new opportunities to generate insights on systemic algorithmic-driven risks and invite a reconceptualization of audit practices beyond their data collection techniques.

In addition, VLOPs and VLOSEs must undergo annual *independent audits*¹¹ to assess compliance with DSA obligations, including those imposing to identify, analyze, and mitigate algorithmic-driven risks. In order to carry out such compliance audits effectively, audited providers must grant compliance auditors meaningful access to their algorithmic and IT systems, including testing environments, as a necessary precondition for a valid and comprehensive audit [32]. Such compliance audit process can involve testing and probing digital services' algorithmic systems. While this mirrors practices in algorithm auditing, the availability of internal information in a compliance context introduces the need to identify algorithm study designs that can effectively inform the methodological choices of compliance auditors. This is particularly important given that most, if not all, of the existing algorithm auditing literature focuses on methodologies developed for external, rather than internal, access to data and systems.

In response to these developments, our categorization departs from prior classifications and instead organizes algorithm audit studies according to their objectives, that is, the purposes they might serve within the broader risk management framework of the DSA. This structure enables us to first define audit study designs conceptually, as structured plans for investigating algorithmic-driven risks, and then discuss how such designs can be operationalized under different data access conditions. We therefore intentionally separated our taxonomy from the discussion on data collection strategies to emphasize their role as adaptable tools applicable to different study designs. We use the *narrative review* [107] approach to explore relevant literature and identify four categories of algorithm audit study designs, each focused on a distinct objective: uncovering and raising awareness of previously unknown algorithmic-driven risks (*risk-uncovering*), identifying key parameters that influence algorithmic behavior and associated risks (*reverse-engineering*), examining how digital services' interface design choices may exacerbate these risks (*interface design*), and, finally, rigorously measuring the influence of the algorithm on the risks (*risk-measuring*). For each category, we outline key steps, provide methodological recommendations, and identify best practices.

The remainder of the paper is organized as follows. In Section 2, we position our contribution within the broader scientific literature on algorithm auditing, situating it among ongoing efforts to categorize methodologies in this field and to clarify related terminology. Section 3 outlines our methodology, the narrative review approach, and specify our criteria for inclusion and exclusion of related literature, and the process by which we extracted and developed the taxonomy. Section 4 forms the core of the paper: here, we present

¹Article 3(i) defines an online platform as "a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation;"

²Article 3(j) defines an online search engines as "an intermediary service that allows users to input queries in order to perform searches of, in principle, all websites, or all websites in a particular language, on the basis of a query on any subject in the form of a keyword, voice request, phrase or other input, and returns results in any format in which information related to the requested content can be found;"

³those with more than 45 million active recipients of their services in the EU, that is, a number equivalent to 10% of the Union population.

⁴<https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses>

⁵Article 34 and recitals 79-85 on risk assessment

⁶Article 35 and recital 86-90 on risk mitigation

⁷Article 34(2)

⁸Article 40

⁹<https://data.europa.eu/data/datasets/dsa-transparency-database?locale=en>

¹⁰Article 39

¹¹Article 37

our taxonomy of audit study designs. For each category of study design, we describe its objectives, the actors who tend to conduct it, its main characteristics and methodological approaches, the kind of evidence it generates, best practices, illustrative examples from the literature, and the related area of relevance to DSA obligations. In section 5 we explore various strategies for data collection, outlining their respective advantages and limitations, and we examine the new tools introduced by the DSA in this context. Finally, Section 6 offers discussion and concluding remarks.

2 RELATED WORK AND TERMINOLOGY

Recent work seeks to extract valuable insights from research on algorithm auditing to identify lessons applicable to auditing in the context of EU regulations [9, 74, 78, 79]. This paper builds on these efforts by emphasizing the relevance of algorithm auditing methodologies in the context of the DSA for various stakeholders, including third-party auditors, digital services, and researchers.

The concept of algorithm auditing. As outlined in the introduction, algorithm auditing involves diverse practices aimed at evaluating the harms caused by the use of algorithms, as well as their transparency, accountability, and fairness. Sandvig et al. [97] were early proponents of the concept of algorithm auditing, drawing on principles from social science’s *audit studies*, which are field experiments designed to reveal instances of discrimination [39, 116]. They introduced study designs specifically aimed at identifying algorithmic discrimination, with a particular emphasis on algorithms used by online platforms and external access to the data and algorithms under investigation. Since then, the term algorithm audit has expanded to encompass the examination of various problematic algorithm behaviors beyond discrimination. These social science inspired studies are extremely relevant for our analysis as they are very much aligned with the *socio-technical* evaluations [60] (evaluations that recognize that technical considerations are inseparable from the social context in which the systems are designed and operate) useful to assess the systemic risks outlined in the DSA. We therefore borrow from the terminology used in [97] and use the term *algorithm audit study design* to refer to *structured plans for investigating risks related to the use, design, and functioning of algorithmic systems in digital services*.

Related fields. The scientific literature on algorithm auditing is closely related to that of *fairness in machine learning*, *explainable and interpretable AI*, and *red teaming for AI*, the latter particularly in the context of Generative AI. These areas share a common concern for identifying biases and discrimination within algorithms [13, 72] and ensuring the reliability and technical transparency of algorithmic systems [36, 43, 81, 84, 95]. However, such studies often treat algorithmic models as the primary unit of analysis, examining internal properties, biases, and performance in a controlled experimental environment [91, 103]. In contrast, we consider studies that investigate algorithmic-driven risks stemming from models that are integrated into digital services, which is more directly aligned with the scope of the DSA.

The algorithm auditing ecosystem. The existing literature distinguishes between various types of algorithm audits based on the auditor’s relationship to the entity being audited, as well as the level of access and cooperation involved. These distinctions are

reflected in the terminology used to classify different types of algorithm audits. For example, many authors distinguish between *first-*, *second-*, and *third-party* algorithm audits based on the auditor’s independence from the audit target [21], or between *internal* and *external* audits based on the level of access to the algorithmic systems being examined [90]. External audits are sometimes referred to as adversarial audits [3, 31, 113], as they are conducted without cooperation from the digital service or access to its internal data.

In this paper, we identify three key stakeholders that can benefit from the methodological best practices outlined in our taxonomy: *researchers*, *third-party auditors*, and *digital services*. Digital services typically conduct first-party internal audits of their own algorithms. Researchers often engage in third-party external audits. The category of third-party auditors includes all organizations that carry out independent audits of digital services’ algorithms, regardless of their level of access to internal systems.

This last category includes *compliance auditors*. Compliance audits involve independent assessments to ensure adherence to legal obligations [17]. In the context of the DSA, compliance auditors must assess compliance with all due diligence obligations under the regulation. However, in this paper, we focus specifically on provisions directly linked to the design and functioning of a platform’s algorithmic systems. Indeed, such compliance audits may incorporate algorithmic audits similar to those discussed in this paper [32].

3 METHODOLOGY

We used a *critical narrative review* methodology, as defined by Sukhera et al. [107], to examine methods for investigating algorithmic-driven risks in digital services. We selected this approach for its flexibility and interpretative nature, which allowed us to incorporate a wide range of studies including grey literature such as reports and news articles. This narrative review was conducted through a structured process organized into five key phases:

- (1) **Development of interpretative framework.** Our review is grounded in the DSA, with a focus on its provisions related to algorithmic systems and the associated risks, particularly those outlined in Article 34 on Risk Assessment. This article identifies four key algorithmic-driven risk factors: platform recommender systems, content moderation systems, ad selection algorithms, and platform design. These factors contribute to risks such as the spread of illegal content, harm to fundamental rights, civic discourse, elections, public security, gender-based violence, public health, protection of minors, and the person’s well-being. This framework guided our selection and synthesis of relevant literature.
- (2) **Identification of relevant studies.** We began by identifying seminal papers on algorithm auditing¹² and relevant keywords related to DSA risks and algorithmic factors. We then used these to refine our search queries on Google Scholar and Scopus, and supplemented our search by incorporating the references of identified papers (including relevant grey literature references). This process generated a comprehensive corpus of 300 studies that formed the basis for our analysis.

¹²such as surveys of relevant algorithm auditing methodologies

- (3) **Screening for relevance.** We collected basic metadata, including abstracts, to conduct a manual relevance screening. We applied two criteria to select papers for full-text review: the study had to present a methodology for investigating algorithmic-driven risks in digital services or provide a concrete example of such an investigation. A total number of 40 papers were included in the master corpus for full review. All authors participated in this screening process, and we held regular meetings to discuss and resolve any borderline cases, ensuring a consensus on the selected papers.
- (4) **Full-text review and reference mining.** Each paper in the master corpus underwent a full-text review, during which reviewers filled out a review form capturing key insights, methodological strengths, and limitations. Additionally, reference mining was conducted by identifying potentially relevant papers from the references cited in the reviewed studies. These additional papers were incorporated into the master corpus to ensure comprehensive coverage.
- (5) **Final analysis and synthesis.** This process comprised a series of meetings where all authors discussed key findings of their individual analyses, focusing on identifying overarching themes, methodological trends, and summarizing insights. As a result of our analysis, we introduce a novel taxonomy that systematically organizes these methodologies (§4).

4 TAXONOMY OF STUDY DESIGNS

The taxonomy focuses on *algorithm audit study designs*, defined as structured plans to investigate risks associated with the use, design, and functioning of digital services' algorithms. The proposed categorization is based on the purpose that each study design might serve within the broader risk management framework of the DSA.

We present an overview of the four main categories in our taxonomy (Table 1): 1) *risk-uncovering* studies, whose focus is on uncovering and raising awareness on previously unknown algorithmic-driven risks 2) *reverse engineering* studies, whose focus is on understanding the functioning of the algorithms integrated into digital services, 3) *interface design* studies, whose focus is on risks stemming from digital service's interface design choices, and 4) *risk-measuring* studies, whose focus is on quantitatively ascertaining algorithmic-driven risks. Each category is explored through its main steps, best practices, and illustrative examples, as well as its relevance to the DSA and how it can support researchers, third-party auditors, and digital services.

Although the identified categories provide a structured approach to investigate algorithmic-driven risks, different stakeholders can adapt and combine them to achieve a more comprehensive assessment of a specific algorithmic-driven risk. For instance, a *risk-measuring study* often builds upon the findings of a *risk-uncovering study*, which has identified a new risk associated with the use of an algorithmic system, and is further informed by the insights gained from a *reverse-engineering study*, which can identify the key functionalities driving the algorithm's behavior. Alternatively, a *risk-measuring study* designed to measure the effect of a specific design feature on users would typically build on a *interface design*

study that would have identified this feature as potentially problematic. Therefore, the proposed taxonomy should be viewed as a flexible framework, comprising modular components that can be combined and tailored to examine algorithmic-driven risks.

4.1 Risk-uncovering study

Definition and scope. The risk-uncovering study employs an investigative approach to identify and expose risks stemming from the use, design, and functioning of digital services as experienced by users in real-world usage, with the objective of raising awareness of previously unknown harmful behaviors of algorithmic systems.

Who usually conducts it and why. These studies are usually carried out by investigative journalists, civil society organizations, and impacted communities or individuals. Risk-uncovering studies often serve as a foundational step in the broader process of evaluating algorithmic-driven risks, aiding in the discovery of "unknown unknowns".

Main characteristics and approaches. Unlike other types of audit studies, which investigate a particular risk that has been already identified beforehand, this audit type is triggered by a chance encounter or unexpected observation by users as they interact with algorithmic systems in real-world settings (also known as *situated use* [100, 106]), leading to the discovery of previously unknown algorithmic-driven risks. We have identified two types of risk-uncovering studies in the literature that differ in their level of organization:

- **User-driven audit.** A bottom-up and user-driven approach where individuals or groups of users engage in an audit following an incidental exposure to problematic algorithmic behavior in their everyday use of the service. These are spontaneous (collective or individual) efforts to hypothesize and test an observed algorithmic behavior during the course of their ordinary interactions with the service¹³.
- **Investigative report.** A detailed and in-depth examination of a specific issue, event or situation, typically involving potential misconduct, wrongdoing or unethical practices. These reports are often conducted by civil society organizations such as investigative journalists, NGOs, advocacy groups, and community-based organizations. This type of investigation can be organized and systematic.

Type of evidence generated. Risk-uncovering studies generate evidence that varies in scope and rigor, ranging from anecdotal accounts of personal experiences and testimonies from users to rigorous, quantitative and qualitative evidence gathered through systematic investigation. At one end of the spectrum, anecdotal accounts provide valuable insights into problematic algorithmic behaviors, while at the other end, investigative reports conducted by journalists or organizations offer a more comprehensive and rigorous examination of systemic issues. Regardless of the level of rigor, this evidence is crucial in identifying previously unknown risks, harmful algorithmic behaviors, or unintended consequences associated with algorithmic systems.

¹³This subcategory of our taxonomy is inspired by the work of Shen and colleagues [100]

Table 1: Overview of the proposed taxonomy

Name	Objective	Typical evidence	Main areas of relevance to DSA obligations	Resources analyzed
Risk-uncovering study	Uncover and raise awareness of previously unknown VLOPs and VLOSEs algorithmic-driven risks.	Qualitative and quantitative evidence in the form of news articles, reports, and blog posts, ranging from anecdotal accounts to systematic analyses.	Obligations related to risk identification, analysis and mitigation: Art. 34 and 35	[14, 16, 27, 30, 46, 47, 56, 59, 86, 93, 100, 102, 109]
Reverse engineering study	Understand VLOPs and VLOSEs' algorithmic system's internal parameters, functionalities or processes.	Quantitative evidence of the functionalities and processes of VLOPs and VLOSEs' algorithmic systems.	Obligations related to profiling and transparency: Art. 26(3), 28(2), 27, and 38	[1, 4, 10, 55, 57, 58, 99, 117, 119]
Interface design study	Critically assess and identify User Interface (UI) and User Experience (UX) design elements that might exacerbate algorithmic-driven risks.	Qualitative evidence in the form of expert assessments and insights on user experience and behavior.	Obligations related to the identification, analysis and mitigation of risks stemming from UI and UX design choices: Art. 34 and 35.	[26, 28, 64, 76, 80, 120, 121]
Risk-measuring study	Systematically and rigorously measure and understand VLOPs and VLOSEs algorithmic-driven risks.	Quantitative and statistically robust evidence on the impact of VLOPs and VLOSEs' algorithmic systems on specific risks.	Obligations related to risk identification, analysis and mitigation: Art. 34 and 35.	[22, 37, 48, 49, 51–53, 85, 87, 89, 92]

4.1.1 *Main steps and best practices.* Building on the steps outlined in [100], we identified three key stages of the risk-uncovering audit study:

- (1) **Initiation.** Individuals or groups encounter instances of harmful algorithmic behavior while using the digital service.
- (2) **Hypothesizing and testing.**
 - **User-driven audit.** Users respond to observed behavior by forming *folk theories* [25, 29, 42] which are informal, intuitive explanations of how the algorithm works, based on anecdotal evidence or intuition. Unlike scientific hypotheses, these folk theories are not systematically developed or tested, but rather emerge and are tested through users' everyday interactions with the system. Such testing usually lacks systematic rigor, as it is not guided by any experimental design.
 - **Investigative reports.** The process may involve a combination of data analysis, expert interviews, user testimonials, or an examination of the algorithm's design and implementation.
- (3) **Awareness raising.**¹⁴
 - **User-driven audit.** Initiators share their findings with others through platform channels or community spaces. News and media sources may become aware of the audit and publish articles highlighting the issues it uncovers.
 - **Investigative reports.** A news article or a report outlining the findings is published and shared with the public.

4.1.2 *Illustrative examples.* An example of a user-driven, bottom-up audit of the algorithmic system utilized by Booking.com to rate businesses based on user reviews is presented in [30]. In this case, the platform's users observed that the algorithm had a tendency to inflate low review scores (**step 1**), prompting them to hypothesize and test its behavior submitting varied scores (**step 2**), and ultimately raising awareness about the perceived bias by strategically modifying their review practices (**step 3**). An example of investigative reports are the risk-uncovering studies carried out by the Wall Street Journal and Bloomberg on TikTok's recommender system and its impact on children's mental health [14, 108]. In this case, journalists became aware that TikTok's recommender system seemed to create *rabbit holes* of potentially harmful content for children and teenagers (**step 1**). This prompted them to hypothesize and test the algorithm behavior through the use of automated accounts [108], and interviewing experts and impacted users [14] (**step 2**), leading to the publication of a series of news articles aiming at raising awareness on the problem to the general public (**step 3**).

4.1.3 *Main areas of relevance to DSA obligations.* The findings of risk-uncovering studies can increase the understanding of algorithmic-driven risks for various stakeholders in the DSA ecosystem. Specifically, **researchers** can leverage these findings to identify areas for further systematic evaluation and pinpoint the most impacted groups. **Digital services** can also benefit from the findings of these studies to identify their blind spots. They can also facilitate end-user engagement through design interventions that foster user discussion and participation [62, 100]. Additionally, by actively recruiting and guiding users in these exercises, digital services can tap into

¹⁴The *awareness raising* and *hypothesizing and testing* steps may occur simultaneously or in an iterative fashion as more information is progressively uncovered and shared.

their collective insights and experiences, ultimately leading to more comprehensive and effective risk management (Art. 34 and 35). Lastly, **third-party auditors** can also leverage these findings to ensure that uncovered risks are addressed in VLOPs and VLOSE's annual risk assessment reports (Art. 34) and assess the effectiveness of proposed mitigation measures (Art. 35).

4.2 Reverse engineering study

Definition and scope. The scope of a reverse engineering study involves uncovering and analyzing the internal mechanisms and behaviors of digital services' algorithmic systems integrated in digital services without direct access to the underlying code or model. It focuses on understanding how the system processes inputs to produce outputs by systematically manipulating variables and observing the effects. This process often follows a rigorous experimental design and employs statistical tests to measure and validate the resulting insights. This step is often necessary to gain a deeper understanding of the algorithmic system and later characterize its role in the emergence of algorithmic-driven risks (risk-measuring audits, §4.4).

Who usually conducts it and why. This type of study is typically conducted by researchers. Their objective is often to identify crucial factors to be controlled in subsequent experiments designed to measure associated risks (§4.4).

Main characteristics and approaches. There are two types of approaches to reverse engineering audits:

- **Black-box approach.** Auditors treat the algorithmic system as a black box, observing only its input-output behavior. They conduct controlled experiments by varying specific inputs or attributes (e.g., location) to understand how these changes affect the algorithm's behavior (e.g., search result rankings).
- **Grey-box approach.** Auditors have partial knowledge of the system's internal workings. By using their domain expertise, available documentation, and any accessible internal information, they make informed guesses about the intermediate steps and processes in the algorithmic pipeline. They then perform controlled experiments to manipulate these intermediate components, helping to identify which one is responsible for the observed behavior.

Type of evidence generated. Reverse-engineering studies generate evidence that helps the auditor understand the algorithmic system's behavior and decision-making processes. Such evidence involves measurable data from controlled experiments, such as the impact of specific inputs on outputs. This evidence can be statistically rigorous, especially when supported by hypothesis testing and other statistical analyses to validate findings.

4.2.1 Main steps and best practices.

- (1) **Scope determination** Define the scope of the study by selecting the digital service and its specific features or functionalities under investigation.
- (2) **Hypothesis formulation**
 - **Black-box reverse engineering audit.** Formulate hypotheses on the impact of certain attributes/inputs on the algorithmic system's behavior.

- **Grey-box reverse engineering audit.** Formulate hypotheses about the potential relationships between specific intermediate processes or components and the algorithmic system's behavior.

- (3) **Experimental design definition.** Plan and design experiments to test the formulated hypotheses by identifying the variables to manipulate and the methods for controlling and varying inputs, in particular:

- **Measured outcome.** Identify the most appropriate attribute that measures system's behavior (e.g., changes in recommendations).
- **Treatments.** Identify the specific inputs to manipulate (e.g., user interactions, targeting attributes) to measure their effect on the system behavior (measured outcome).
- **Confounding variables.** Identify and control variables that might correlate with both the input being manipulated and the system's behavior, potentially leading to misleading conclusions. For example, if you are testing how user interactions influence recommendations, user demographics might be a confounding variable, as they could influence both the interactions and the recommendations.
- **Baseline comparisons and noise variables.** Establish control scenarios (e.g., non-interacting profiles) to differentiate systematic behavior from random fluctuations.

- (4) **Data collection.** Define the data collection strategy, mainly defining which methodology is best to access the necessary information for the study.

- Define data to be collected based on your experimental design.
- Select data collection strategy (see §5).

- (5) **Data analysis.**

- Choose metrics that accurately capture changes in system behavior to effectively test the hypothesis.
- Use statistical tests or modeling to confirm/reject hypotheses.

4.2.2 Illustrative examples. In black-box reverse engineering studies, researchers focus on observable behavior without internal knowledge of the system. For example, Boeker et al. [10] examined how user interactions (language, location, and video engagement) influence TikTok's "For You" page recommendations (**step 1**). They hypothesized that these interactions would significantly impact the content users receive (**step 2**). To test this, they designed an experiment using twin user profiles, controlled for confounding factors (**step 3**), and collected data on recommended posts using automated accounts (**step 4**). They found that the "following" action had the strongest influence on recommendations (**step 5**). In contrast, grey-box reverse engineering involves some internal knowledge of the system. For instance, Andreou et al. [4] investigated Facebook's ad transparency focusing on the "Why am I seeing this?" feature and Ad Preference Page (**step 1**). They hypothesized that the ad targeting process involved data inference, audience selection, and user-ad matching and that some targeting attributes might be omitted in Facebook's explanations (**step 2**). They designed and conducted experiments using a browser extension and controlled ad campaigns to collect data on ad explanations and targeting attributes (**steps**

3-4) evaluating the alignment between explanations and internal ad targeting processes (**step 5**).

4.2.3 Main areas of relevance to DSA obligations. For **researchers**, these studies offer a better understanding of system operations, guiding experimental designs to evaluate risks associated with VLOPs and VLOSEs. For example, if location affects search rankings, researchers can test if this creates filter bubbles in different regions. **Third-party auditors** can use reverse engineering studies to ensure platforms meet transparency obligations. This includes verifying whether declared parameters align with actual targeting practices (Article 26 (d)), evaluating user control over ad parameters, and checking compliance with rules on targeted advertising for minors (Article 28(2)). Additionally, such studies can help assess whether the parameters in recommender systems declared by a platform actually match the true decision-making process (Articles 27(1) and 27(2)(a)) and if non-profiling options are truly non-personalized (Article 38).

4.3 Interface design study

Definition and scope. The interface design study critically analyzes and interprets digital services' User Interface (UI) and User eXperience (UX) design elements. It explores how design decisions influence user behavior and, in turn, shape the inputs fed into the platform's algorithms. By affecting algorithmic processes these user-facing-elements potentially contribute to algorithmic-driven risks.¹⁵

Who usually conducts it and why. Interface design studies are typically conducted by practitioners in the field of human-computer interaction, which builds upon psychology, design theory and field observations with users [68]. Such socio-technical perspective helps to understand the subtle cues, prompts and feedback mechanisms that can shape user behavior, often in ways that are not immediately apparent.

Main characteristics and approaches. Interface design studies analyze the outcomes of digital services' *technology affordances* [24, 41, 94]. Technology affordances are action possibilities that are enabled and facilitated by the design of the interface [41, 83]. Studying such affordances allows experts to conceptualize at a level above specific features [83] and analyse how design influences user behavior and related algorithmic-driven risks. For instance, the introduction of the angry emoji reaction by Facebook in 2016 (the *feature*) enabled a new form of engagement with posts which facilitated users' expression of anger (the *affordance*). This expanded the range of engagement metrics for which the recommender system could optimize, potentially increasing algorithm-driven polarization (the *negative outcome*) [73].

Type of evidence generated. A interface design study primarily generates qualitative evidence that focuses on the impact of design choices on user behavior and algorithmic-driven risks. This evidence includes expert assessments of the platform's UI and UX elements.

¹⁵It is important to note that certain design features, including some types of deceptive dark patterns [69], may also pose risks that are not algorithmic in nature. While the audit of such design features exceeds the scope of this article, future research could examine the best practices for interface audits in a more comprehensive manner, building on foundational work such as [70].

4.3.1 Main steps and best practices. We identified three steps integral to the interface design study:

- (1) **Scope definition.** Define the scope of the study by identifying the specific digital service and design artifact to be investigated (e.g. TikTok For You Feed, YouTube comment section).
- (2) **Affordance identification.** Explore the design artifact under study and identify the affordances enabled by its design elements. This analysis may involve examining screenshots and page segments.
- (3) **Outcome assessment.** Critically assess whether the affordances identified lead to potential negative outcomes. This can be achieved in several ways, such as through expert assessments, interviews with users and designers, or lab studies where researchers observe users interacting with an interface.

4.3.2 Illustrative examples. Munn et al. [80] examined how YouTube's comment section design (**step 1**) facilitates online discourse polarization. They identified five key affordances and interface elements that encourage leaving provocative comments: comments can be upvoted or downvoted, downvoting does not lower the number of upvotes, the most engaged-with comments appear at the top of the comment section of every video, comments can be left anonymously, and the absence of a reputation system to reward users leaving high-quality comments (**step 2**). They then evaluated that such affordances encourage the practice of leaving provocative, controversial, and generally polarizing comments, as they tend to receive more engagement and are thus showcased at the top of the comment section, creating a feedback loop that fosters toxic and polarizing discourse online (**step 3**). A similar study [64] employed user surveys to study how YouTube *Up next recommendations* feature (**step 1**), enables unplanned discovery of videos that users might find interesting (**step 2**), enabling platforms to gather more engagement data on users, and having a negative impact on users' sense of agency (**step 3**).

4.3.3 Main areas of relevance to DSA obligations. The interface design audit plays a crucial role in understanding and assessing the impact of user interface (UI) design patterns and affordances on user behavior. **Researchers** can use this type of study to identify which digital service's design elements contribute to systemic risks and inform the design of subsequent experiments that quantitatively assess such negative effect (for example using a *risk-measuring study*, §4.4). The insights gained from these studies can also help **third-party auditors** create evidence to evaluate whether digital services are compliant with obligations related to risk assessment (Article 34) and risk mitigation (Article 35). For **digital services** themselves, conducting a interface design audit can be part of their risk assessment processes, allowing them to identify and address any negative outcomes resulting from their service design (Article 34).

4.4 Risk-measuring study

Definition and scope The scope of a risk-measuring study involves identifying, measuring, and analyzing algorithmic-driven risks in digital services, such as amplification of misinformation or

algorithmic pathways to radicalization. Typically, the risk-measuring study takes place after a potential risk associated with the use of an algorithmic system has been identified, for example by a *risk-uncovering study* (§4.1), and after gaining some knowledge on the algorithmic system functionalities and processes by, for example, conducting a *reverse-engineering study* (§4.2).

Who usually conducts it and why Risk-measuring studies are typically conducted by researchers that aims to quantitatively ascertain the impact of digital service’s algorithmic systems on a specific risk.

Main characteristics and approaches The risk-measuring audit may involve conducting *experimental* and *observational* studies.

- **Experimental studies** the auditor design an experiment that involves interaction with the algorithmic system, aiming to assess its impact on systemic risks, such as testing the system’s behavior under various scenarios.
- **Observational studies** the auditor analyzes the historical data on the algorithm’s behavior, without interacting with the algorithmic system’s operation.

Type of evidence generated The risk-measuring audit study aims to provide quantitative and statistically robust evidence of the impact of VLOPs and VLOSEs’ algorithmic systems on specific risks.

4.4.1 Main steps and best practices.

- (1) **Scope determination.** Define the study’s scope by selecting the digital service and algorithmic system to investigate, focusing on a specific algorithmic-driven risk. Identify how the algorithmic-risk is related to the systemic risks outlined in the DSA.
- (2) **Research questions and hypotheses formulation.** Clearly define the research questions and hypotheses, ensuring they are grounded in existing knowledge about both the technical features of the algorithmic system and the social factors related to the risk under study. Explicitly state any assumptions that underpin your study.
- (3) **Experimental design.** Plan and design experiments to effectively measure the algorithmic-driven risk and test the formulated hypotheses, in particular identifying the following variables:
 - **Measured outcome** (aka dependent variable). Determine the most appropriate quantity to measure the risk under investigation (e.g. for algorithmic-driven exposure to harmful content one could measure the number of harmful videos recommended to users).
 - **Treatment(s)** (aka independent variable, intervention). Identify the specific element under investigation for its potential impact on the outcome (e.g. users’ time spent on video).
 - **Confounding variables.** Identify and control variables that could correlate with both the input being manipulated as well as the system’s behavior, potentially leading to misleading conclusions. For example, when testing how user time spent on video influences algorithmic-driven exposure to harmful content, users’ interest and viewing habits

might be confounding variables, as they could influence both watch time and recommendations.

- **Baseline comparison and noise variables.** Establish control scenarios (e.g. different choice of internet browser) to differentiate systematic behavior from random fluctuations.

(4) Data collection.

- Define data to be collected. The relevancy of the data depends on the experimental design (validate or reject the initial hypothesis) and its variables, as well as the strategy chosen to measure its outcome.
- Select data collection strategy (see §5).

(5) (if necessary) Ground truth labeling.

assign labels or categories to collected data based on expert knowledge or verified sources.

(6) Data analysis.

- Choose metrics that can effectively measure the impact of treatments on the identified risks.
- Use statistical tests or modeling to confirm/reject hypotheses
- Perform eventual robustness check for assumptions made

4.4.2 Illustrative examples. The study by Jiang et al. [51] presents a solid application of the risk-measuring audit study design. It investigates the potential for partisan bias in comment moderation on YouTube by focusing on whether comments on politically diverse videos are moderated differently, in line with the DSA’s concern about freedom of expression (**step 1**). The study formulates research questions and hypotheses on how political stance affects comment moderation. It tests whether moderation differs between left- and right-leaning videos and between extreme and center videos, considering both the system’s workings and relevant social factors (**step 2**). In the experimental design, they define the key variables: moderation status as *measured outcome*, political stance and magnitude as *treatments*, and video engagement and hate speech as *confounding factors* (**step 3**). Data is collected by scraping YouTube comments (**step 4**) and relies on previous research for ground truth labeling (**step 5**). The collected data is then analyzed to examine moderation patterns (**step 6**). While the study does not specifically address noise variables, it performs robustness checks to evaluate assumptions about comment removal. Another example is the study by Hussein et al. [48], which investigates the amplification of misinformation on YouTube. The authors investigate how various factors influence the amplification of conspiracy theories that could have an impact on public health and civic discourse (**step 1**). The study formulates its research questions and hypotheses on how users’ watch histories affect the recommendation of misinformative videos (**step 2**). In the experimental design, they define the treatment (watch history, age, gender...) and outcomes (number of misinformative videos recommended) and other key variables. It addresses noise variables like web browser and tracked cookies by maintaining consistency across experimental conditions (**step 3**). To collect data they deploy automated accounts to watch selected popular videos with specific stances on misinformation (**step 4**). Recommended videos were then analyzed based on human annotations following predefined heuristics (**step 5**). For data analysis the study applies non-parametric tests (**step 6**).

4.4.3 *Main areas of relevance to DSA obligations.* **Researchers** can use the risk-measuring design study to investigate in a structured way algorithmic-driven systemic risks. The identified best practices of the risk-measuring study also offer a guide to **third-party auditors** that want to test whether the digital services diligently analyzed and assessed the systemic risks stemming from the design and functioning of their algorithmic systems (Article 34), and to assess whether the platform's mitigation measures do effectively reduce such risks (Article 35). Finally, **digital services** can implement risk-measuring studies with internal data collection strategies (e.g. sampling or A/B testing) to conduct a diligent assessment of the risks stemming from the design and functioning of their proprietary algorithmic systems (Article 34 and 35).

5 STRATEGIES FOR COLLECTING DATA

The previous literature [6, 50, 97] identified the following six methods to collect data from digital services by researchers without direct access to their internal systems: *API*, *web scraping*, *crowdsourcing*, *surveys and interviews*, *sock-puppets*, and *carrier-puppets*. Each of these methods has its own strengths and limitations, and the approach taken depends on the research question, the characteristics of the digital service under analysis, and the available resources. **API** (Application Programming Interface) is a standardized method for accessing and manipulating data on remote servers or services. Historically, APIs have been a crucial tool for researchers to obtain data from digital services, but recent restrictions have limited researchers' access, citing concerns over user privacy and data misuse [8, 88, 105]. However, even when APIs are available, the digital service controls the information that is made available to researchers [65, 66, 110]. For instance, APIs may not provide access to data on user experience on a platform, or may only provide aggregated data that does not allow for nuanced analysis of systemic risks.

In contrast, other common techniques allow researchers to obtain data from digital services in a more direct and unfiltered way. For instance, **web scraping** is a method that involves extracting the data from websites using software tools or scripts to parse the website HTML page code. Such a methodology does not rely on the digital service's willingness to provide data as it gathers the publicly available information displayed on the site. However, web scraping has been historically limited by digital services' Terms of Service (ToS) [114] and require continuous monitoring for changes in the page layout, which can be time-consuming and resource-intensive. Alternatively, **crowdsourcing** and **data donation** are two related approaches to collect data directly from real-world users who consent to share their data. This can be facilitated through various tools, such as browser extensions or automated scripts, that collect information from users' devices [98]. Users can also request their personal data from online platforms and search engines, exercising their rights under the EU GDPR [33], and then voluntarily donate it [11, 12, 122]. This data collection approach offers the advantage of gathering authentic, real-world data from a potentially large user base. Nevertheless, it also presents several challenges as it may be prone to sampling bias, requires robust data protection measures to safeguard user privacy, and can be cumbersome due to the lack of

standardized user data access protocols [115]. **Surveys and interviews** provide a qualitative perspective on user experience, offering rich and detailed insights into users' perceptions, attitudes, and behaviors. However, these methods may be limited by response bias [38] (which occurs when participants' responses are influenced by their own biases, social desirability or other factors) and provide limited insight into the complex interactions between users, algorithms, and digital service's design.

Sock-puppets are virtual agents that mimic real users by interacting with a digital service's interface in a controlled and planned way. By using sock-puppets, researchers can create a controlled environment that isolates specific factors and test their influence on algorithmic behavior. This approach allows for precise testing as it can control and eliminate some of the noise associated with real-world user behavior. There are two types of sock-puppets: automated and manual. Automated sock-puppets are digital agents that mimic user behavior with varying degrees of sophistication, ranging from simple pre-programmed scripts to advanced AI-driven agents. Automated sock-puppets allows scalability, but their use is often limited by the ToS and requires significant computational resources [112]. In addition, accurately mimicking real users behavior can be difficult when using this type of sock-puppet. In contrast, manually operated sock-puppets offer more flexibility and adaptability, but are often time-consuming and labor-intensive to operate, and may still be limited by the researcher's ability to simulate user behavior. Finally, **carrier-puppets**, like sock-puppets, interact with digital services to simulate real-world users interactions. However, carrier-puppets have the potential to influence the end-user experience and may "carry" effects onto end-users [6]. This is because carrier-puppets can be used to create and deploy actual content, such as ads, that interact with the digital service's algorithmic systems and its users [4], potentially influencing the user experience. Most of the papers we analyzed in our narrative review adopt a combination of the data collection strategies described.

However, it is important to note that all these well-known data collection strategies are limited by digital services' data sharing policies and by the fact that they typically rely on indirect testing methods, such as input-output probing [15]. To overcome some of these limitations, the DSA grants a higher level of data access to researchers to conduct research contributing to the detection, identification and understanding of systemic risks in the Union [63]. More specifically, Article 40(4) of the DSA allows *vetted researchers*¹⁶ to request access to non-public data of VLOPs and VLOSEs. In addition, Article 40(12) guarantees that independent researchers have unimpeded access to publicly available data¹⁷.

The DSA also introduces two other tools to enhance transparency by providing alternative data sources. Firstly, Article 24(5) of the DSA requires online platform providers to submit "statements of reasons" for their content moderation decisions to the DSA Transparency Database¹⁸. This repository allows oversight bodies and

¹⁶Vetted researchers will be granted this status by the Digital Services Coordinator (DSC) of establishment upon meeting the conditions set out in Article 40(8) of the DSA

¹⁷At the time of writing, the Commission is drafting the *Delegated Regulation on data access provided for in the DSA* to "further specify the conditions under which sharing of data should take place and the purposes for which the data may be used and relevant procedures".

¹⁸<https://transparency.dsa.ec.europa.eu/>

researchers to track and analyze decisions in near real-time, facilitating the study of online content moderation. Secondly, Article 39 of the DSA mandates that providers of very large online platforms and search engines compile and make publicly available a repository containing information about the advertisements they present.

6 DISCUSSION AND CONCLUDING REMARKS

In this paper, we presented a narrative review of audit study designs suitable for investigating algorithmic-driven risks in digital services. We identified four main categories of methodologies: *risk-uncovering*, *reverse engineering*, *interface design*, and *risk-measuring*. We also identified the main steps and best practices associated with each approach. Furthermore, we linked each study design to its potential role in the broader risk management framework of the DSA, emphasizing how different actors in the algorithm auditing ecosystem, namely researchers, digital services, and third-party auditors, can use these methodologies. We have also provided an overview of the different data collection strategies available for researchers and third-party auditors, highlighting the limitations and challenges of traditional methods and discussing the new tools introduced by the DSA to increase transparency and oversight.

By focusing on algorithm auditing *study designs* rather than methodologies to audit specific algorithms, the proposed taxonomy, and related best practices, is applicable to a wide range of algorithmic systems used by digital services, including emerging technologies such as generative AI.

One limitation of this work is that it is based solely on a literature review of publicly available algorithm auditing examples, and therefore does not account for internal auditing conducted by the *Trust and Safety* teams of digital services. However, with the transparency provisions of the DSA, which require platforms to publish their risk assessments, this gap may be addressed in the future, potentially providing greater visibility into internal auditing practices. Finally, we acknowledge that our narrative review approach is not as comprehensive as a systematic review or meta-analysis, but we believe that our contribution provides a valuable foundation for future research on auditing algorithmic-driven risks in digital services.

REFERENCES

- [1] Philip Adler, Casey Falk, Sorelle A Friedler, Tionney Nix, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, and Suresh Venkatasubramanian. 2018. Auditing black-box models for indirect influence. *Knowledge and Information Systems* 54 (2018), 95–122.
- [2] Zeeshan Ahmad and Umut Özkaya. 2023. Machine Learning and Artificial Intelligence-based Child Abusing Tracking System for the Detection of Online Sexual Predators. In *International Conference on Trends in Advanced Research*, Vol. 1. 131–141.
- [3] John Albert. 2023. *Risky business: How do we get a grip on social media algorithms?* <https://algorithmwatch.org/en/risky-business-social-media-algorithms/>
- [4] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummedi, Patrick Loiseau, and Alan Mislove. 2018. Investigating ad transparency mechanisms in social media: A case study of Facebook’s explanations. In *NDSS 2018-Network and Distributed System Security Symposium*. 1–15.
- [5] Brooke Auxier and Monica Anderson. 2021. Social media use in 2021. (2021).
- [6] Jack Bandy. 2021. Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the acm on human-computer interaction* 5, CSCW1 (2021), 1–34.
- [7] Jack Bandy and Nicholas Diakopoulos. 2020. Auditing news curation systems: A case study examining algorithmic and editorial logic in Apple News. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 36–47.
- [8] Rebecca Bellan. 2024. Meta axed CrowdTangle, a tool for tracking disinformation. Critics claim its replacement has just 1% of the features. <https://techcrunch.com/2024/08/15/meta-shut-down-crowdtangle-a-tool-for-tracking-disinformation-heres-how-its-replacement-compares/>
- [9] Abeba Birhane, Ryan Steed, Victor Ojewale, Briana Vecchione, and Inioluwa Deborah Raji. 2024. AI auditing: The broken bus on the road to AI accountability. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 612–643.
- [10] Maximilian Boeker and Aleksandra Urman. 2022. An empirical investigation of personalization factors on tiktok. In *Proceedings of the ACM Web Conference 2022*. 2298–2309.
- [11] Laura Boeschoten, Jef Ausloos, Judith Moeller, Theo Araujo, and Daniel L Oberski. 2020. Digital trace data collection through data donation. *arXiv preprint arXiv:2011.09851* (2020).
- [12] Laura Boeschoten, Jef Ausloos, Judith E Möller, Theo Araujo, and Daniel L Oberski. 2022. A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research* 4, 2 (2022), 388–423.
- [13] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
- [14] Olivia Carville. 2023. TikTok’s Algorithm Keeps Pushing Suicide to Vulnerable Kids. *Bloomberg* (2023). <https://www.bloomberg.com/news/features/2023-04-20/tiktok-effects-on-mental-health-in-focus-after-teen-suicide>
- [15] Stephen Casper, Carson Ezell, Charlotte Siegmann, Noam Kolt, Taylor Lynn Curtis, Benjamin Bucknall, Andreas Haupt, Kevin Wei, Jérémy Scheurer, Marius Hobbahn, et al. 2024. Black-box access is insufficient for rigorous ai audits. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2254–2272.
- [16] Rumman Chowdhury. 2021. Sharing learnings about our image cropping algorithm. https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm
- [17] Rumman Chowdhury. 2023. *What’s an audit?* <https://www.get-parity.com/raiblog/whats-an-audit>
- [18] Nick Clegg. 2023. *How AI Influences What You See on Facebook and Instagram*. <https://about.fb.com/news/2023/06/how-ai-ranks-content-on-facebook-and-instagram/>
- [19] European Commission. 2015. Monitoring the digital economy & society 2016–2021. , 52 pages.
- [20] Nicholas Confessore. 2018. Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times* 4 (2018), 2018.
- [21] Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1571–1583.
- [22] Shakked Dabran-Zivan, Ayelet Baram-Tsabari, Roni Shapira, Miri Yitshaki, Daria Dvorzhitskaia, and Nir Grinberg. 2023. “Is COVID-19 a hoax?”: auditing the quality of COVID-19 conspiracy-related information and misinformation in Google search results in four languages. *Internet Research* 33, 5 (2023), 1774–1801.
- [23] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*. 271–280.
- [24] Jenny L Davis. 2020. *How artifacts afford: The power and politics of everyday things*. MIT Press.
- [25] Michael A DeVito, Jeremy Birnholtz, Jeffery T Hancock, Megan French, and Sunny Liu. 2018. How people form folk theories of social media feeds and what it means for how we study self-presentation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
- [26] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI dark patterns and where to find them: a study on mobile applications and user perception. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
- [27] Nicholas Diakopoulos. 2015. Algorithmic accountability: Journalistic investigation of computational power structures. *Digital journalism* 3, 3 (2015), 398–415.
- [28] Jacob Erickson and Bei Yan. 2024. Affective Design: The Influence of Facebook Reactions on the Emotional Expression of the 114th US Congress. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–9.
- [29] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I “like” it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2371–2382.
- [30] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. “Be careful; things can be worse than they appear”: Understanding Biased Algorithms and Users’ Behavior around Them in Rating Platforms. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. 62–71.

- [31] Eticas Foundation. 2023. Adversarial Algorithmic Auditing Guide. <https://eticasfoundation.org/wp-content/uploads/2024/04/ETICAS-Adversarial-Algorithmic-Auditing-Guide-2023.pdf>
- [32] European Commission. 2024. Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202400436
- [33] European Parliament and the Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504>
- [34] European Parliament and the Council. 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>
- [35] Eurostat. 2022. Consumption of online news rises in popularity. <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/ddn-20220824-1>
- [36] Michael Feffer, Anusha Sinha, Wesley H Deng, Zachary C Lipton, and Hoda Heidari. 2024. Red-teaming for generative ai: Silver bullet or security theater?. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 421–437.
- [37] Flavio Figueiredo, Felipe Giori, Guilherme Soares, Mariana Arantes, Jussara M Almeida, and Fabricio Benevenuto. 2020. Understanding Targeted Video-Ads in Children’s Content. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. 151–160.
- [38] Adrian Furnham. 1986. Response bias, social desirability and dissimulation. *Personality and individual differences* 7, 3 (1986), 385–400.
- [39] S Michael Gaddis. 2018. *An introduction to audit studies in the social sciences*. Springer.
- [40] Adi Gaskell. 2022. How Biased Google Search Results Affect Hiring Decisions. *Forbes* (2022).
- [41] William W Gaver. 1991. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 79–84.
- [42] Susan A Gelman and Cristine H Legare. 2011. Concepts and folk theories. *Annual review of anthropology* 40 (2011), 379–398.
- [43] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51, 5 (2018), 1–42.
- [44] Larry Hardesty. 2019. The history of Amazon’s recommendation algorithm. *Amazon Science* 22 (2019).
- [45] Jennifer A Harriger, Joshua A Evans, J Kevin Thompson, and Tracy L Tylka. 2022. The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms. *Body Image* 41 (2022), 292–297.
- [46] T Hobbs, Rob Barry, and Yoree Koh. 2021. The Corpse Bride Diet’: how TikTok inundates teens with eating-disorder videos. *The Wall Street Journal* (2021).
- [47] Jeff Horwitz and Katherine Blunt. 2023. Instagram Connects Vast Pedophile Network. *The Wall Street Journal* (2023).
- [48] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
- [49] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2024. Auditing for Racial Discrimination in the Delivery of Education Ads. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2348–2361.
- [50] Ada Lovelace Institute. 2021. *Technical methods for regulatory inspection of algorithmic systems*. <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/>
- [51] Shan Jiang, Ronald E Robertson, and Christo Wilson. 2019. Bias misperceived: The role of partisanship and misinformation in youtube comment moderation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 278–289.
- [52] Prerna Juneja, Md Momen Bhuiyan, and Tanushree Mitra. 2023. Assessing enactment of content regulation policies: A post hoc crowd-sourced audit of election misinformation on YouTube. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [53] Prerna Juneja and Tanushree Mitra. 2021. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–27.
- [54] Ioannis Kangas, Maud Schwoerer, and Lucas J Bernardi. 2021. Recommender systems for personalized user experience: lessons learned at Booking. com. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 583–586.
- [55] Levi Kaplan and Piotr Sapiezynski. 2024. Comprehensively Auditing the TikTok Mobile App. In *Companion Proceedings of the ACM on Web Conference 2024*. 1198–1201.
- [56] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic folk theories and identity: How TikTok users co-produce Knowledge of identity and engage in algorithmic resistance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–44.
- [57] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of the 2015 internet measurement conference*. 121–127.
- [58] Daniel Klug, Yiluo Qin, Morgan Evans, and Geoff Kaufman. 2021. Trick and please. A mixed-method study on user assumptions about the TikTok algorithm. In *Proceedings of the 13th ACM Web Science Conference 2021*. 84–92.
- [59] Michelle S Lam, Mitchell L Gordon, Danaë Metaxa, Jeffrey T Hancock, James A Landay, and Michael S Bernstein. 2022. End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34.
- [60] Michelle S Lam, Ayush Pandit, Colin H Kalicki, Rachit Gupta, Poonam Sahoo, and Danaë Metaxa. 2023. Sociotechnical Audits: Broadening the Algorithm Auditing Lens to Investigate Targeted Advertising. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–37.
- [61] Stephan Lewandowsky, Laura Smillie, David Garcia, Ralph Hertwig, Jim Weatherall, Stefanie Egidy, Ronald E Robertson, Cailin O’Connor, Anastasia Kozlyeva, Philipp Lorenz-Spreen, et al. 2020. Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making. (2020).
- [62] Rena Li, Sara Kingsley, Chelsea Fan, Proteeti Sinha, Nora Wai, Jaimie Lee, Hong Shen, Motahhare Eslami, and Jason Hong. 2023. Participation and Division of Labor in User-Driven Algorithm Audits: How Do Everyday Users Work together to Surface Algorithmic Harms?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [63] Anna Liesenfeld. 2024. The Legal Significance of Independent Research based on Article 40 DSA for the Management of Systemic Risks in the Digital Services Act. *European Journal of Risk Regulation* (2024), 1–13.
- [64] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Himiker. 2021. How the design of YouTube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [65] Emma Lurie. 2023. Comparing platform research API requirements. *Tech-Policy Press*. Retrieved December 7 (2023), 2023.
- [66] Emma Lurie, Dan Bateyko, and Frances Schroeder. 2023. TikTok just announced the data it’s willing to share. What’s missing?
- [67] Kim Lyons. 2021. Facebook’s ad delivery system still has gender bias, new study finds. *The Verge* (2021).
- [68] Wendy E Mackay and Anne-Laure Fayard. 1997. HCI, natural science and design: a framework for triangulation across disciplines. In *Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques*. 223–234.
- [69] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.
- [70] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.
- [71] Katerina Eva Matsa. 2022. More Americans are getting news on TikTok, bucking the trend on other social media sites. (2022).
- [72] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.
- [73] Jeremy B. Merrill and Will Oremus. 2021. Five points for anger, one for a “like”: How Facebook’s formula fostered rage and misinformation. <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>
- [74] Anna-Katharina Mefsmar and Martin Degeling. 2023. Auditing Recommender Systems—Putting the DSA into practice with a risk-scenario-based approach. *arXiv preprint arXiv:2302.04556* (2023).
- [75] Meta. 2022. *How technology helps prioritise review*. <https://transparency.fb.com/en-gb/enforcement/detecting-violations/technology-helps-prioritize-review/>
- [76] Thomas Mildner and Gian-Luca Savino. 2021. Ethical user interfaces: Exploring the effects of dark patterns on facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [77] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. 2023. “I See Me Here”: Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [78] Jakob Mökander. 2023. Auditing of AI: Legal, Ethical and Technical Approaches. *Digital Society* 2, 3 (2023), 49.

- [79] Jakob Mökander, Maria Axente, Federico Casolari, and Luciano Floridi. 2022. Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. *Minds and Machines* 32, 2 (2022), 241–268.
- [80] Luke Munn. 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Communications* 7, 1 (2020), 1–11.
- [81] Micah Musser, Andrew Lohn, James X Dempsey, Jonathan Spring, Ram Shankar Siva Kumar, Brenda Leong, Christina Liaghati, Cindy Martinez, Crystal D Grant, Daniel Rohrer, et al. 2023. Adversarial Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications. *arXiv preprint arXiv:2305.14553* (2023).
- [82] SL Myers. 2022. How Social Media Amplifies Misinformation More Than Information. *The New York Times* (2022).
- [83] Amy Orben, Adrian Meier, Tim Dalgleish, and Sarah-Jayne Blakemore. 2024. Mechanisms linking social media use to adolescent mental health vulnerability. *Nature Reviews Psychology* 3, 6 (June 2024), 407–423. <https://doi.org/10.1038/s44159-024-00307-y>
- [84] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, et al. 2023. The role of explainable AI in the context of the AI Act. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. 1139–1150.
- [85] Kostantinos Papadamou, Antonis Papasavva, Savvas Zannettou, Jeremy Blackburn, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Michael Sirivianos. 2019. Disturbed youtube for kids: Characterizing and detecting disturbing content on youtube. *arXiv preprint arXiv:1901.07046* (2019).
- [86] Olson Parmy. 2018. The Algorithm That Helped Google Translate Become Sexist. *Forbes* (2018).
- [87] Brooke Perreault, Johanna Hoonsun Lee, Ropafadzo Shava, and Eni Mustafaraj. 2024. Algorithmic Misjudgement in Google Search Results: Evidence from Auditing the US Online Electoral Information Environment. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 433–443.
- [88] Jay Peters. 2023. It's not just Apollo: other Reddit apps are shutting down, too. <https://www.theverge.com/2023/6/8/23754616/reddit-third-party-apps-api-shutdown-rif-reddplanet-sync>
- [89] Sam Power and Ben Mason. 2023. Mobilizing or chasing voters on Facebook? Analysing echo-chamber effects at the UK parliamentary General Election 2019. *Parliamentary affairs* 76, 1 (2023), 1–21.
- [90] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 33–44.
- [91] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.
- [92] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgilio AF Almeida, and Wagner Meira Jr. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 131–141.
- [93] Aja Romano. 2019. A group of YouTubers is trying to prove the site systematically demonetizes queer content. *Vox* (2019).
- [94] Alexander Ronzhyn, Ana Sofia Cardenal, and Albert Batlle Rubio. 2023. Defining affordances in social media research: A literature review. *New Media & Society* 25, 11 (2023), 3165–3188.
- [95] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* 1, 5 (2019), 206–215.
- [96] Arianna Sala, Lorenzo Porcaro, and Emilia Gómez. 2024. Social Media Use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports* 14 (2024), 100404.
- [97] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. An algorithm audit. *Data and discrimination: Collected essays* (2014), 6–10.
- [98] Leonardo Sanna, Salvatore Romano, Giulia Corona, and Claudio Agosti. 2020. YTTREX: crowdsourced analysis of YouTube's recommender system during COVID-19 pandemic. In *Annual International Conference on Information Management and Big Data*. Springer, 107–121.
- [99] Anna Semenova, Martin Degeling, and Greta Hess. 2024. Understanding TikTok's For You Feed. <https://tiktok-audit.com/blog/2024/For-You-Feed/>
- [100] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.
- [101] Craig Silverman and Jeff Kao. 2022. Infamous Russian troll farm appears to be source of anti-Ukraine propaganda. *ProPublica* (2022).
- [102] Ellen Simpson and Bryan Semaan. 2021. For You, or For" You"? Everyday LGBTQ+ Encounters with TikTok. *Proceedings of the ACM on human-computer interaction* 4, CSCW3 (2021), 1–34.
- [103] Mona Sloane, Emanuel Moss, and Rumman Chowdhury. 2022. A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns* 3, 2 (2022).
- [104] SnapChat. 2022. *Machine Learning for Snapchat Ad Ranking*. <https://eng.snap.com/machine-learning-snap-ad-ranking>
- [105] Chris Stokel-Walker. 2023. TechScape: Why Twitter ending free access to its APIs should be a 'wake-up call'. <https://www.theguardian.com/technology/2023/feb/07/techscape-elon-musk-twitter-api>
- [106] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
- [107] Javeed Sukhera. 2022. Narrative Reviews: Flexible, Rigorous, and Practical. *Journal of Graduate Medical Education* 14, 4 (08 2022), 414–417. <https://doi.org/10.4300/JGME-D-22-00480.1> arXiv:<https://meridian.allenpress.com/jgme/article-pdf/14/4/414/3251685/1949-8357-14-4-414.pdf>
- [108] The Wall Street Journal. 2021. Inside TikTok's Algorithm: A WSJ Video Investigation. *The Wall Street Journal* (2021). <https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477>
- [109] Elise Thomas. 2021. Recommended Reading: Amazon's algorithms, conspiracy theories and extremist literature. *Institute for Strategic Dialogue* (2021).
- [110] Rebekah Tromble. 2021. Where have all the data gone? A critical reflection on academic digital research in the post-API age. *Social Media+ Society* 7, 1 (2021), 2056305121988929.
- [111] Twitter. 2023. *Twitter's Recommendation Algorithm*. https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm
- [112] Roberto Ulloa, Mykola Makhortyk, and Aleksandra Urman. 2024. Scaling up search engine audits: practical insights for algorithm auditing. *Journal of information science* 50, 2 (2024), 404–419.
- [113] Aleksandra Urman, Ivan Smirnov, and Jana Lasser. 2023. The right to audit and power asymmetries in algorithm auditing. *arXiv preprint arXiv:2302.08301* (2023).
- [114] Aleksandra Urman, Ivan Smirnov, and Jana Lasser. 2024. The right to audit and power asymmetries in algorithm auditing. *EPJ Data Science* 13, 1 (2024), 19.
- [115] Patti M. Valkenburg, Amber Van der Wal, Teun Siebers, Ine Beyens, Laura Boeschoten, and Theo Araujo. 2024. From Generosity to Frustration: Research Community Reveals Obstacles in Online Platforms' Data Access. https://www.project-awesome.nl/images/PDFs/Report_data_donation_studies.pdf
- [116] Briana Vecchione, Karen Levy, and Solon Barocas. 2021. Algorithmic auditing and social justice: Lessons from the history of audit studies. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–9.
- [117] Karan Vombatkere, Sepehr Mousavi, Savvas Zannettou, Franziska Roesner, and Krishna P Gummadi. 2024. TikTok and the Art of Personalization: Investigating Exploration and Exploitation on Social Media Feeds. In *Proceedings of the ACM on Web Conference 2024*. 3789–3797.
- [118] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale commodity embedding for e-commerce recommendation in alibaba. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 839–848.
- [119] Craig E Wills and Can Tatar. 2012. Understanding what they do with what they know. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*. 13–18.
- [120] Pamela Wisniewski, Karla Badillo-Urquiola, Zahra Ashtorab, and Jessica Vitak. 2020. Happiness and fear: Using emotions as a lens to disentangle how users felt about the launch of Facebook reactions. *ACM Transactions on Social Computing* 3, 4 (2020), 1–25.
- [121] Mochen Yang, Yuqing Ren, and Gediminas Adomavicius. 2020. Engagement by design: an empirical study of the "reactions" feature on facebook business pages. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 6 (2020), 1–35.
- [122] Savvas Zannettou, Olivia Nemes-Nemeth, Oshrat Ayalon, Angelica Goetzen, Krishna P Gummadi, Elissa M Redmiles, and Franziska Roesner. 2024. Analyzing User Engagement with TikTok's Short Format Video Recommendations using Data Donations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [123] Hongying Zhao and Christian Wagner. 2022. How TikTok leads users to flow experience: investigating the effects of technology affordances with user experience level and video length as moderators. *Internet Research ahead-of-print* (2022).
- [124] Zhengwei Zhao. 2021. Analysis on the "Douyin (TikTok) Mania" phenomenon based on recommendation algorithms. In *E3S Web of Conferences*, Vol. 235. EDP Sciences, 03029.