Response to the European Commission's call for evidence on a planned Delegated Regulation on data access provided for in the Digital Services Act (DSA)

May 2023

Stiftung Neue Verantwortung

Stiftung Neue Verantwortung (SNV) is a not-for-profit think tank working on current political and societal challenges posed by new technologies. We not only invite government officials but everyone seeking information to engage with our work whether through giving us feedback on publications, participating in our events or seeking direct advice. Our experts work independently from partisan interests or political affiliations.

For questions and comments, please contact the authors Dr. Anna-Katharina Meßmer, Dr. Martin Degeling and Dr. Julian Jaursch. The responses to the questions partly draw on input from colleagues from SNV as well as other European academic and civil society experts, including AI Forensics and AlgorithmWatch, whose work and support we gratefully acknowledge. We thank the European Commission for the opportunity to participate in the call for evidence and look forward to engaging further with the Commission as well as other interested stakeholders.

Summary of key points

SNV strongly welcomes the DSA's intended push for improvements in platform-to-researcher data access. A privacy-friendly, accessible EU-wide regime is necessary to replace the current system of arbitrary and piecemeal access points provided by some platforms on a voluntary basis. This system has created unequal access for researchers, offers incomplete data and lacks key technical features, as SNV has experienced first-hand by conducting platform research and using several platforms' APIs.

While the provisions in Article 40 DSA are a good start, they do need to be spelled out in greater detail. That is why we applaud the Commission for focusing on this issue and starting work on a delegated act as provided for in Article 40(13). The delegated act can ideally lead to major improvements in how researchers can access platform data. To achieve this goal, we offer several suggestions in this response to the Commission's call for evidence.

Data access needs

- Current research APIs only provide a limited set of variables and less than is publicly available, for example, via scraping. While the current trend is to limit APIs and the available data, we suggest the opposite: All data that can already be accessed by scraping should be available through APIs.
- Research often struggles with lack of relevant meta data. That is why platforms need to properly document data schemas and content removal.
- It is necessary to clarify that scraping of public data is in accordance with Article 40(12) to remove legal uncertainties about this vital research practice.

Data formats and technical specifications

- Platforms should offer regular APIs as well as stream APIs.
- Experimentation with automated user accounts should be possible on the platform or in sandboxes to study algorithms like recommender systems.
- Data donations and exploratory research should be allowed.

Vetting process

- It should be reiterated in the delegated act that researchers both from academic and civil society organizations can apply for data access, including those from outside the EU.
- In addition to applications regarding specific research projects, it should be allowed for researchers and institutions as such to apply, which would further support exploratory and inductive research.

Governance structure

- We strongly suggest setting up an independent advisory mechanism as mentioned in Article 40(13). This body, made up of academic and civil society researchers, could facilitate the development of common vetting standards, support DSCs in the actual vetting process and serve as a forum for exchange.
- The Commission, the European Centre for Algorithmic Transparency, the DSCs as well as researchers themselves should immediately engage in capacity-building efforts to make the best use of Article 40. Member states should equip their DSCs with strong research and data science units to conduct data analyses themselves and support researchers.

1. Data access needs

a) What types of data, metadata, data governance documentation and other information about data and how it is used can be useful to DSCs for the purpose of monitoring and assessing compliance and for vetted researchers for conducting research related to systemic risks and mitigation measures?

General considerations on data access for vetted researchers

Researchers are already collecting data from very large online platforms (VLOPs) and very large online search engines (VLOSEs). Regardless of the kind of client through which data is being served to end users (e.g., web or mobile app), researchers have been obtaining information on the impacts and risks of platforms by several means.[1] Of special interest with respect to technical data access are methods such as automated audits (often called scraping), code audits (through reverse-engineering), collection of public documents and crowd-sourced data donations. These established techniques have rendered the availability of data a matter of (technical) capacity, which has in turn produced an undesirably unequal access.

Data access under the DSA could foster research on risks and mitigations by leveling the playing field with respect to the availability of data.

What types of data can be useful for vetted researchers

We propose the following guiding principle for data access: All data that is already served to end-user clients should become available for researchers (under Article 40(4) as well as under Article 40(12)). Whether or not the data in question is presented to the user should be irrelevant. This means: Regardless of the user client, all data that is produced by users or the platform and that can already be scraped, intercepted or reverse-engineered should be made readily available. This information is obtainable by technical means, although some methods often violate corporate terms of service. Access to it should not depend on the technical capacity of the researchers (see Evidence A).

[1] For a mapping of current research on platforms and risks, see SNV's overview "Auditing Recommender Systems. An Overview of Existing Audits, Risk Assessments and Studies on Potential 'VLOPs'".

Evidence A: TikTok's API offers less data than the company makes publicly available otherwise

Data available in TikTok's application programming interface (API)
TikTok offers an API to developers and researchers[2] that, when querying videos, returns 14 data points containing the most relevant information, including the date the video was created, how often it was watched, liked or shared, the hashtags, and, if available, the transcribed audio. This is far from the only data available, however, as the following points and the appendix to this document show.

Data available via the website
Using widely available open-source tools that leverage the information available on the website tiktok.com, researchers can retrieve a total of 118 data points.[3] The additional data contains information that will allow for a range of analyses. This includes:

1. Data created by the user and related to the content
   - Text elements added to the video by the user
   - Detailed information about the effects used (e.g., the name of the effect such as "GreenScreen", where the official API only provides an "id" but no way of understanding what this "id" refers to)

2. Data describing the original content such as
   - A list of "diversificationLabels" that are assigned by TikTok to around 1% of videos
   - A flag whether or not the video is an advertisement

3. Additional technical information about the content
   - Technical data about the video like height and width, but also the format and quality
   - The duration of the video

Data available via the mobile app
When retrieving the information from TikTok's mobile app, each video comes with even more data points, 657 in total. But collecting data from the app is technically more complicated and there are no open-source tools readily available to collect this data.[4] The additional data points that can be retrieved using the app include "risk_infos" and "masking labels" which trigger warnings in the user interface, how often a video was shared through specific channels (e.g., "whatsapp"), location data of the video, templates used during the video creation or whether its distribution is limited through "geofencing".

Reports have revealed that platforms process and enhance the data they receive for a variety of reasons, from automated content moderation to subtitle generation. **Some of the variables resulting from such mechanisms would be interesting to researchers and crucial to better understand platforms' inner workings to assess risks.** Over the past years, several companies have reduced the data they make available. For instance, Twitter recently closed its free API that was used for various research projects, to increase revenue.[5] YouTube removed a data point from their API in 2017 after it was used in a study[6] and recently announced it will discontinue an API that allowed researchers to find related videos, but it is unclear why[7]. This trend needs to be reversed to enable studies on various platform practices and risks.

[2] TikTok for developers, "Research API: Getting started".
[3] See also the examples SNV collected in the technical schema definitions from TikTok's website.
[4] See also the examples SNV collected in the technical schema definitions from TikTok's app.
[5] Andrew Hutchinson, "Twitter's Cancelling Free Access to Its API, Which Will Shut Down Hundreds of Apps", Social Media Today, February 2, 2023.
[6] BrodaYT, "Youtube removes 'excluded_ads' tag from the source code on videos in two days after NerdCity's video exposes it!", Reddit Post, December 1, 2017.
[7] Bernhard Rieder, "YouTube Will Close the RelatedToVideoId API Parameter…", Twitter, May 5, 2023.

As just one example for this, journalists uncovered that TikTok has implemented a curation mechanism called "heating" that is used by employees to push videos to a broader audience.[8] **Researchers** studying how content distribution works on social networks like TikTok **would benefit from a data point explaining whether a video was manually "heated", which should be a simple "flag" in the video attribute.**

Similarly, **information available in the app data about content sharing on other platforms**, such as the number of shares on WhatsApp, would help researchers understand whether the virality of disinformation is based solely on user interactions or is also promoted by the platform. Moreover, the use of templates for the video editing tool CapCut that is closely integrated with TikTok is also only disclosed in the data retrieved through the app. Templates can be reused by other users and therefore simplify the distribution and creation of similar content. Analyzing the use of templates could help understand how these might or might not be employed to reproduce disinformation.

Another report suggests that TikTok makes use of **categorizations for videos and users**.[9] These categorizations are not shown to users but would be **necessary to understand and evaluate content moderation decisions**. Therefore, it is crucial that additional data which users provide (e.g., text written on "stickers" that are added to videos) and platforms add (e.g., automated classification like the "diversificationLabels" or whether a posting was "heated") are made available alongside the content.

On some platforms, internal content moderation or content classification processes lead to additional information that is shown to users alongside specific content. For example, some platforms show warnings about potential disinformation, trigger warnings or adult content warnings alongside specific content. Labels highlighting content from "authoritative sources" also fall in this category. For some research questions, access to such information is vital (see Evidence B). TikTok discloses risk labels as well as masking labels only in the data retrieved through the app.

Evidence B: Lack of data to test usefulness of content labels

A SNV study about digital news and information literacy[10] shows that people often do not see, use or understand information cues like labels about disinformation, fact checks or further information about a channel. At the same time, platforms mention in their transparency reports that labels on disinformation can decrease the share of links and therefore the spread of disinformation.

If researchers want to empirically validate or falsify such statements, they need access to the following data:

1. **Label IDs and description** how these labels are presented and what the actual information is
2. **Possibilities to compare the spread of content with and without labels** (especially timestamped data of content before and after a label was added)
3. **Access to experiment data** of A/B tests of such labels

What types of metadata can be useful for vetted researchers

Most data is served to end-user applications in a format that implies an underlying structure, sometimes referred to in technical terms as schema. This means a set of variables of specific types, sizes, possible values and other characteristic features. **These schemas should be made available in addition to a codebook/dictionary that**

[8] Emily Baker-White, "TikTok's Secret 'Heating' Button Can Make Anyone Go Viral", Forbes, January 20, 2023.
[9] TikTok internally created lists of users that watch "gay" content, see Georgia Wells and Byron Tau, "TikTok Tracked Users Who Watched Gay Content, Prompting Employee Complaints", The Wall Street Journal, May 5, 2023.
[10] Anna-Katharina Meßmer, Alexander Sängerlaub and Leonie Schulz, "Source: the internet – testing digital news and information literacy in Germany", Berlin: Stiftung Neue Verantwortung, March 22, 2021.

describes the information implied by every variable.[11] The variables should also be described in detail, which means explaining the possible values (see Evidence C).

Evidence C: Lack of descriptions of metadata

The TikTok Research API returns a list of "effect_ids" for each video, without offering a codebook or separate API explaining what "effect_ids" are available to understand what effects were applied to a video.

TikTok also discloses a "view_count" in its current APIs. Since the app allows users to rapidly dismiss videos it would be necessary to understand what TikTok counts as a "view". **Is a video already counted as "viewed" when it loads, even if users close it immediately, or does it have to play for a certain number of seconds or a certain percentage of its duration to be considered "viewed"?**

Scraping tools based on the web version of TikTok ("unofficial APIs") give researchers access to additional data TikTok holds about each video, but the meaning of some data points remains unclear. For example, the web API (widely used for scraping) contains two values, "hit_shark" and "hit_dolphin", for each video, which are not explained anywhere. The app version contains a value "bodydancescore" that would require a description by TikTok to understand it.

Changes to these schemas should be documented in detail, so that any vetted researcher can refer to the schema version their data belongs to.

Availability and timeliness of data

Data should be available to researchers as soon as it has been published by users/creators and made available to users. This means that only reasonable delays should be expected in terms of the "freshness" of the data.

Data governance documentation

Records on data governance documentation, broadly understood as keeping track of any decisions concerning the handling of data and metadata, should be accessible to researchers. Some research formats require retrospective access to data and metadata. Therefore, tracking data governance decisions such as deletion or changes to access privileges is fundamental to the feasibility of studies on content moderation practices. For example, in certain research scenarios data and metadata on specific pieces of content is required. If, for any reason, this content has been deleted, there should be – among other information regarding the data governance decisions – a record stating that this was the case, when it happened and why. Another relevant scenario for this record-keeping is the research reproducibility, which is in the best interest of all stakeholders (see Evidence D).

[11] One example, not necessarily meant as a good practice, is TikTok's codebook.

Evidence D: Missing information on content removal and incomplete data

As part of the DataSkop project conducted by the not-for-profit research and advocacy organization AlgorithmWatch[12], SNV obtained a dataset consisting of user-donated records of the videos they access on TikTok. To study certain characteristics of this VLOP's content recommendation algorithm, reproducing specific user behavior was desirable. In attempting to do so, we found that several of the videos that users had engaged with were marked as unavailable, without any details of what this meant. Are these videos now unavailable because they were removed by the user? Was the removal a platform decision? Answers to such questions are crucial to achieving the objectives of our study.

If the content in question has been removed by the platform, this governance decision should be documented in the content metadata or some other record. This would allow studies on risks of recommender systems to make correct inferences, for instance.

Twitter, for example, does provide a "Compliance" API[13] that distinguishes between different statuses like "deleted" (meaning that the tweet or user account has been deleted), "deactivated" (meaning that the tweet or user account has been deactivated), "scrub_geo" (meaning that the geo-information associated with the tweet or user has been removed), "protected" (meaning that the account from which the tweet originated has become private) and "suspended" (meaning the account from which the tweet originated has been suspended).

During the data donation project with AlgorithmWatch and in discussions with researchers at AI Forensics, we also found that the data TikTok provides to users who request a copy of data through the app or website is limited and sometimes even incomplete. Users have to go through the more complex process of writing an e-mail referring to their rights under Article 15(3) of the EU's General Data Protection Regulation (GDPR) to request a more complete copy of the data TikTok has about them. **Platforms should offer users a simple interface to request their data and provide a complete data set.**

b) What sort of analysis and research might DSCs and vetted researchers conduct for the purposes of monitoring and assessing compliance and conducting research related to systemic risks and mitigation measures?

General considerations on conducting research related to systemic risks and mitigation measures

The list of systemic risks mentioned in the DSA presents a challenge to researchers as it covers an enormous range of risks that vary in how well they are specified or can be defined. Some of these risks can be operationalized quite easily based on existing legislation, for example, when it comes to "illegal activities", such as the sale of products or services prohibited by EU or national law.

Other risks, however, are completely undefined, quite abstract and vague or have multiple, sometimes contradictory definitions in different disciplines. This is especially true for all risks that are described as having an impact on individuals (e.g., psychological well-being), groups of individuals (nondiscrimination) or societies (civil society discourse and election processes). These risks depend on a complex socio-technical interplay between platform design, different algorithms, users and society.

Therefore, research on systemic risks must consider the specifics of each platform, its products (e.g., "For You feed", search bar, marketplace) and its elements (e.g., the user interface or the different stages of a recommender system). Additionally, it will be necessary for any study of risks to differentiate between those risks that can be observed on a platform, for example, the spread of content about depression and the impact these risks have on certain people, communities or society. Proving the spread of specific content does not necessarily have a causal relation to developments outside of the platform, for example, the mental well-being of individuals.

[12] AlgorithmWatch, "DataSkop", 2023.
[13] Twitter Developer Platform, "Batch compliance".

Therefore, the level at which these risks can be observed varies widely. What is needed here is, first, more concrete descriptions of specific risk scenarios and, second, accompanying sociological and/or psychological research to capture the effects that take place beyond the platforms or in interaction between platforms and users.

That is why **SNV has proposed a risk-scenario-based approach to conduct research related to systemic risks and mitigation measures and also to monitor and assess compliance of VLOPs and VLOSEs with the DSA**.[14] Focusing on systemic risks in concrete scenarios will help to **clarify questions like the following: What is the definition of "hate speech" or "mental well-being"? Where should this risk be observed and how? How does it manifest on a specific platform or via a specific product of this platform? What is the impact on individuals or society?**

Because high-level risks are impossible to be assessed as a whole, they need to be broken down into smaller pieces considering the potential harms they entail. Each harm can be further described by a (potentially large) number of scenarios. To standardize the way scenarios are described, we propose including at least the following information:

1. Who is **the affected party?**
2. What **characterizes** this party?
3. What **harm** does this actor, person or group experience?
4. How is **the platform** involved in this experience?
5. What are the **macro impacts** that go beyond the individual?

Answering these questions should allow for a thorough characterization of the risk scenario. The following descriptions should clarify what is required of each component:

1. **Affected party:** Every risk scenario starts with the definition of the affected party. Who do you think of as being affected by a specific harm? Individuals or groups? Recipients or creators? Private persons or advertisers? All platforms have different user groups that show different usage behaviors because there is no "general user". There is often a fluid transition between someone consuming content for various reasons and those using a platform for professional and business purposes or creators with different levels of expertise and professionalism.
2. **Characteristics:** The party (potentially) experiencing harm by being on the platform might have characteristics that are relevant to this process: For example, users from marginalized groups, creators with a specific agenda or organizations engaged in a policy debate might be more likely to experience a specific harm. The relevant characteristics should be specified as clearly and thoroughly as possible.
3. **Harm:** This component of the scenario should describe what the affected party experiences that would be considered harmful. Harms can be quite individual and encompass even psychological effects, such as "feeling depressed", in which case harm is connected to mental health risks. If the affected party is a group, for example, a community posting empowering videos on living with mental health issues, harms can have an effect on individual, collective and macro levels. For example, the tightrope of moderating mental health content could lead to over-blocking and shadow-banning content about mental illnesses, which is meant to empower people. Demoting or blocking that content is a risk that can influence the personal well-being of a user experiencing content about their mental health issues being blocked after sharing it, but this is also an example of discrimination and an issue related to freedom of expression.
4. **Platform involvement:** Several elements and products of a platform can be involved in promoting a specific risk or harm. This part of the scenario should describe or hypothesize what elements of the platform are influencing the recommendation system of a specific product and, therefore, directly or indirectly fostering the harm, such as content moderation (e.g., shadow banning), the user interface

[14] Anna-Katharina Meßmer and Martin Degeling, "Auditing Recommender Systems. Putting the DSA into Practice with a Risk-Scenario-Based Approach", Berlin: Stiftung Neue Verantwortung, January 23, 2023.

(e.g., angry emojis) or user experience (e.g., frictionless endless feeds). This involvement mainly shapes the need for specific data access requirements and what kind of data is needed.

5. **Macro impact:** The abstract risks mentioned in the DSA might concern negative impacts on individuals and groups as well as on society as a whole. To differentiate between these levels, every risk scenario not only defines a specific harm, but it can also define the larger impact this harm can have on society, for example, civil discourse. These platform effects might be harder to assess and need additional sociological and/or psychological research such as questionnaires.

Potential research data use cases to study the spread of disinformation

To highlight just one example of potential areas of research based on the approach using risk-based scenarios described above, the spread of disinformation could be considered a systemic risk. It is related to potential negative effects on fundamental rights and to civic discourse and might have different repercussions for various groups, depending on historical, regional and language contexts. For this field of research, **scientists have suggested a multitude of research avenues to better understand disinformation online, if they had appropriate access to data.**[15] For instance, with extended APIs, researchers could run better studies on the general usage of these platforms and how they might or might not foster the spread of disinformation. With more detailed data on how users interact with content, they could study the difference between posts that are actually read versus those which are "just" shared. With access to randomized control trials, it would be easier to figure out what mitigations would actually limit the spread of disinformation. This might include studying accuracy ratings, labels or nudges/prompts that create some friction. With a better understanding of people's demographic and network data, studies on motivation to share various types of content are possible.

As a matter of course, any of such studies – not just on disinformation but on all systemic risks – should comply with the GDPR and generally safeguard users' privacy. **When directly accessing user data, for instance, through enhanced browser plug-ins enabled via Article 40, informed user consent is essential**.

It has to be noted that not all studies researchers can imagine at the moment can be realized using Article 40. **That is why it is important to regularly revisit the delegated act and also the DSA itself to determine whether adaptations and additions are necessary.**

2. Data access application and procedure

a) Digital Services Coordinators (DSCs) in the Member States will play a key role in assessing researchers' applications and they will act as intermediaries with the platforms. How should the application process be designed in practice? How can the vetting process ensure efficient exchanges between researchers and platform providers?

b) Article 40(8) exhaustively defines criteria for vetting researchers. How can a consistent assessment across DSCs be ensured, while still taking into consideration the specificities of each request?

Questions 2a) and 2b) are considered together to combine suggestions for the vetting process and its criteria.

Access for researchers from civil society

Very basically, the application process should be electronic, contain standardized elements relating to the criteria laid out in Article 40(8) and be easily available and accessible for the variety of researchers and organizations

[15] Irene V. Pasquetto et al., "Tackling Misinformation: What Researchers Could Do with Social Media Data", Harvard Kennedy School Misinformation Review, December 9, 2020.

eligible to request data access. There is a great diversity regarding disciplines, methods and types of organizations that the vetting process should accommodate.

Crucially, **Article 40(8) allows researchers from civil society organizations and not only from academic institutions such as universities to request data access. DSCs should not discriminate in any way between the applicants** and base their vetting decisions solely on compliance with the criteria laid out in Article 40(8). Proving their own focus on "conduct[ing] scientific research" or "carry[ing] out educational activities involving also the conduct of scientific research" in accordance with Article 2(1) of Directive (EU) 2019/790 should neither become a burden for civil society groups nor a pretense to exclude organizations from applications. The number of peer-reviewed and published scientific articles should not be the only criterion to establish an organization's research focus. Instead, the delegated act should make clear that other forms of verification are possible, too: DSCs should accept, among other things, proof that organizations have produced or have it as their mission to produce quantitative or qualitative analyses using diverse methodologies such as data analysis, expert interviews and stakeholder consultations. Past research publications and projects, within a certain time frame, should count as evidence that organizations fulfill the necessary criteria. However, to not exclude new organizations without publications, DSCs should also take into account the organization's research mission and the individual applicants' expertise.

Access for researchers from outside the EU

Article 40(8) does not prevent researchers from outside the EU to apply for data access, but the language on this should be clarified in the delegated act. As long as the researchers focus on systemic risks according to Article 34(1) and fulfill the other criteria from Article 40(8), DSCs should also consider applications from non-EU-based researchers.

Enabling exploratory research

Research into VLOPs and VLOSEs almost always requires a discovery and exploration stage. Because of this, it is often not possible to specify a set of hypotheses or other aspects of a proper research design, let alone the specific data required to carry it out. For this reason, DSCs should not only consider applications regarding individual research projects, but additionally allow researchers and their institutions as such to undergo the vetting process. This is similar to what was already implemented by some VLOPs (e.g., Twitter), where only a general description of the researcher's interest is required.

Common processes for communication between DSCs and for vetting standards

In addition, there should be more detailed standards on the communication process between DSCs, as the corresponding rules in Article 40(9) remain vague. **Especially in cases where DSCs disagree on an application, it is necessary to have a working system to exchange views**. For instance, a Slovakian researcher team might turn in an application to the Slovakian DSC for data from Ireland-based TikTok. The Slovakian DSC's initial assessment is positive but the Irish DSC's – which has the final say – is not. What happens now? Is there any way for the Slovakian DSC (or researchers) to engage with the Irish DSC and potentially amend the application? The delegated act could establish a specific time frame and set out the conditions under which the communication between researchers, the DSCs and, if necessary, the platform works.

Common standards for assessing researchers' "independence from commercial interests" would be helpful as well. It should at least be established that applicants cannot currently or within a set time frame before the application (say, six months) be paid by any VLOP. While it is highly useful to include expertise from corporate researchers to study systemic risks, these researchers do not need the Article 40 access provisions. Instead, Article 40 should be used to check and replicate internal platform research. This exclusion might also alleviate platforms' concerns that the data access provisions could be used by their competitors.

It is also necessary to address the issue of potential rejections or delays in data access by platforms. A system and time frame for amending data access requests is laid out in Article 40(4), (5) and (6), in case platforms cannot immediately comply with the request. It should be clarified that this process also includes DSC requests coming from vetted researchers. **If a data access request is rejected, there should always be clear feedback as to the reasons for denial and an appeal mechanism should be available.**

Relatedly, to ensure consistent assessment of data access requests, it would be useful to **not only have a database of approved and terminated vetted researchers as suggested in Article 40(11) but also *rejected* requests.** It could collect rejections as well as information on why platforms requested amendments. This would allow both DSCs and researchers to better understand the interpretation of the Article 40(8) criteria.

Any considerations of streamlining the vetting process should take into account, and not duplicate, the corresponding discussions from the Code of Practice on Disinformation. Specifically, if efforts on the Code's commitment 27 (data access for studying disinformation) have yielded tangible results before the Board and DSCs are in place, the latter should further develop these processes. This way, voluntary commitments from a multistakeholder forum could become part of the legally binding DSA.

Strong potential for an independent advisory mechanism for the vetting process

Overall, the vetting process should be oriented along the practical needs of the research community. To that end, **it is vital that DSCs establish both internal expertise on platform research and build networks to outside experts**. These endeavors would be greatly aided by establishing an independent advisory mechanism (see the response to question 2e), which would allow DSCs and/or the Board[16] to work in close cooperation with researchers and ideally develop EU-wide guidelines on the vetting process. Such guidelines could be a framework based on the criteria from Article 40(8) but should provide more detail on how DSCs make judgments on the individual applications and deal with the issues highlighted in the response to this question. Annex 1a in the report by the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access (EDMO report) might serve as an initial inspiration for such guidelines.[17]

c) What additional provisions or specifications could be useful to help balance the new data access rights and the protection of users' and business' rights, e.g., related to data protection, confidential information, including trade secrets, and security?

In order to protect user rights, researchers need to implement safeguards as described in section 6 of the EDMO report. This includes pseudonymization and data minimization. It is worth noting that the responsibility to implement safeguards like pseudonymization or data minimization is mainly with the researchers, not the platforms, as **pseudonymization techniques need to take the purpose of the data processing and the whole data set into account. Platforms should therefore refrain from altering the data they provide through APIs**.

Considerations of trade secrets are valid and rightfully included in Article 40(5)b. However, it is important to reiterate the points from recital 97, which states that **commercial interests should not override Article 40 obligations**. Trade secrets might lead to data access requests being amended, but not rejected. The rules laid out for these amendments in Article 40(4), (5) and (6) should apply to access requests coming straight from DSCs as

[16] Even though Article 40 does not directly mention the Board, supporting the vetting process could be deducted from Article 63(1)c (issuing recommendations to DSCs) and Article 63(1)e (supporting the development of guidelines). Specifically, the Board could use the findings from this Commission consultation and engage with the independent advisory mechanism mentioned in Article 40(13) (and in line with Article 62(6) regarding Board invitations to external experts) to draw up "Guidelines on researcher vetting for public-interest platform data research".

[17] Rebekah Tromble et al., "Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access", Brussels: European Digital Media Observatory, May 31, 2022.

well as from DSCs via researchers. The trade secrets exemption should not become an easy way for platforms to get around their legally binding data sharing obligations.

d) What kind of safeguards can be put in place to assure that data gathered under Article 40 is used for the purposes envisaged and to minimize the risk of abuses?

A data sharing agreement rooted in the GDPR should be a part of the data access regime. The point of Article 40 is to enhance understanding of platform business models and systemic risks. The EDMO report outlines safeguards for researchers in Part II, section 6 that could also be applied to the DSA. Such safeguards should be chosen in relation to the sensitivity of the data and should apply to both DSCs and vetted researchers. To minimize this risk of abuse, in a data sharing agreement, it should be reiterated that researchers cannot share data with law enforcement agencies unless legally obligated to do so. **Article 40 should not be used, directly or indirectly, to enhance the surveillance of citizens.**

The safeguards required for research for publicly available data according to Art. 40(12) should be adequate and not excessive when compared to industry standards, for example, when the same data is used for marketing purposes where simple contractual clauses regulate a wide ranged data access.

e) Article 40(13) introduces the possibility of an independent advisory mechanisms (sic!) to support the management of data access requests and vetting of researchers. What would be the added value of such a mechanism?

We see considerable need and potential for DSCs to seek independent advice on data access requests and vetting researchers. More generally, we deem it vital to the functioning of the DSCs to both build up their own expertise on platform risks and to foster networks of outside experts. This is crucial, for instance, when it comes to overseeing non-VLOPs in the national setting, working as the central complaints body for citizens and taking part in Board deliberations.[18]

Both needs – for internal expertise and external expert networks – are especially crucial for the Article 40 provisions. While DSCs should develop in-house knowledge on platform research, they can **greatly benefit from also establishing an independent advisory mechanism to assist** them.[19] This could fulfill various functions: It would support the development of in-house expertise, it would allow for an institutionalized exchange with scientists and it would create a forum in which to discuss common principles for guidelines. Overall, **it would alleviate some of the burdens that DSCs face when having to formulate the data access regime on their own from scratch** (especially considering they will have to deal with various other coordination and oversight tasks right away). DSCs with limited resources would particularly benefit from such a set-up.

In practice, an "independent advisory mechanism" could take the form of an advisory body comprised of scientists from academia and civil society organizations as well as journalists. Since the ultimate vetting decisions will and should remain with DSCs, the advisory body could help with everything leading up to this decision: reviewing proposals, communicating with researchers to amend their requests and offering initial opinions to the DSC. **This way, a type of peer review system would be established. This is desirable to underscore researchers' independence and allow for a vetting process not dictated by regulatory or commercial interests.** Strengthening researchers' independence is also crucial when considering financing such a body. Monetary support from governments and/or platforms could be an option if risks of governmental and corporate capture can be minimized.

---

[18] Cf. Julian Jaursch, "Platform Oversight: Here Is What a Strong Digital Services Coordinator Should Look Like", Verfassungsblog, October 31, 2022.

[19] The EDMO report lays out the argument and ideas for an "independent intermediary body", which could greatly aid the discussions on an "independent advisory mechanism" in the context of Article 40, see Tromble et al., "Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access".

Ideally, such an advisory body would emerge at the EU level and not at each individual national DSC. This could reduce the risk of uneven interpretations of the Article 40 provisions across the EU. Over the long term, having joint guidelines and practices for vetting researchers would contribute to widening the access to the Article 40 provisions. **Since not every university and civil society group might have the necessary resources or expertise to dive into the data access request process, having such guidelines from an independent body could help reduce the burden for such institutions.**

3. Data access formats and involvement of researchers

a) What technical specifications could be considered for data access interfaces, which takes into account security, data protection, ease of use, accessibility, and responsiveness (e.g., APIs, data vaults and other machine-readable data exchange formats)?

Stateless API access / stream APIs

Many studies on platforms focus on the spread of mis- and disinformation for which stateless API access is feasible. Recommender systems and other dynamic components of VLOPs and VLOSEs that the DSA now specifically wants to be scrutinized require additional different kinds of automated information request systems for them to be investigated.

Static queries –known as stateless queries in technical terms – provide information about what has happened on the platforms irrespective of what user interactions have taken place. For example, when looking for information on a specific topic, researchers can request a list of videos that come up using a specific search term multiple times a day, to understand changes in the information ecosystem on that topic.

To better study the distribution of disinformation in real-time platforms could also provide **stream-based APIs**, as they were provided by Twitter until recently. These types of APIs **would notify researchers about new content related to their searches** instead of requiring them to request updates regularly.

User simulations

While there is value in static data and a number of research questions can be answered with this type of access, other impactful and relevant aspects of platforms and the risks they may carry relate to the way user input determines platform output, for instance, in recommender systems (see Evidence E). **Researchers need to be able to automatically simulate users and user inputs to be able to obtain data on what actual users experience by interacting with the platform**. This means that one crucial research affordance is the simulation of user sessions, providing user information (e.g., date of birth or languages spoken), actively created interaction data (e.g., liking or following users or their content) and implicit behavioral data (e.g., what videos are watched for how long, what elements of a platform are used). This type of access can be either granted within sandboxes or as part of the regular systems, with rules developed to limit any negative impact on the regular functioning of the platform.

It is important that researchers have access to multiple methods to probe platforms, for instance, scraping *and* user simulation in sandboxes. This is vital to verify and validate results, as previous research has shown that studying personalization through APIs and user simulations can lead to different results.[20]

[20] Bernhard Rieder, Ariadna Matamoros-Fernández and Òscar Coromina, 2018, "From Ranking Algorithms to 'Ranking Cultures': Investigating the Modulation of Visibility in YouTube Search Results", Convergence: The International Journal of Research into New Media Technologies 24 (1): 50-68.

Evidence E: Restrictions to data access by technical means

Previous research has shown that users on TikTok can end up in "rabbit holes" of depressive content. In an automated study, journalists created bots that started to watch videos that are supposed to present "deep thought" and sad quotes about life. After a while, the recommender systems showed videos of more depressive nature that was considered as suggesting self-harm.[21]

We attempted to reproduce this research using the web and app versions of TikTok and could generally verify that watching videos with specific hashtags will lead to more videos with similar hashtags to be shown to the (automated) user. **But we also found that TikTok limits this kind of research by restricting the ways an automated bot can interact with the platform**. For example, TikTok limits the search function if there is no account. Even with an account, additional interactions such as liking and following are restricted by the platform.

Data donations

In addition to automated user simulation and API data, other means of data collection for risk assessments are equally important. Accordingly, data donation mechanisms, where real users can export and contribute data about their own platform usage, should be shareable with researchers when users give informed consent.

Exploratory research

Research into VLOPs and VLOSEs often requires a discovery and exploration stage. As outlined in our risk-scenario-based audit process (see the response to question 1b), it is necessary for researchers to first get a good understanding of the platform, its elements and affordances. Thus, there is a need to conduct explorative and inductive research in addition to deductive research. It is often not possible to specify a set of hypotheses or other aspects of a proper research design, let alone the specific data required to carry it out. While this exploration might be done with public data as specified in Article 40(12), conducting exploratory research phases must also be possible on other data as also outlined in the EDMO report (see also the response to questions 2a/b).

b) What capacity building measures could be considered for the research community to take advantage of the opportunities provided by Article 40?

To support the research community on Article 40, all involved actors – regulators, platforms and researchers themselves – can and should take immediate action to facilitate the use of the data access provisions. Without the capacity and motivation among the wider research community to use Article 40, its provisions will be useless.

Platforms should support the research community by opening up the study of platform risks. This means that apart from technical data access interfaces, **platforms should develop their own interfaces for common research tasks such as monitoring elections**. Those interfaces should consider different levels of technical complexity with respect to accessibility.

Researchers themselves can and should already prepare well for data access requests, especially compliance measures regarding the DSA and GDPR, and ensure their universities'/organizations' internal standards are met. Yet, these requirements – especially those laid out in Article 40(8)d, such as "specific data security and confidentiality requirements" – are often complex, expensive and hard to handle for smaller research and civil society organizations**. That is why common guidelines on the Article 40 regime are necessary to support**

[21] The Wall Street Journal, "Inside TikTok's Algorithm", July 21, 2021.

researchers. Regulators and/or an independent advisory body (see the response to question 2e) could develop such instructions or guidelines.

Moreover, DSCs as well as the Commission and especially the European Centre for Algorithmic Transparency (ECAT) can play important roles in helping researchers strengthen their networks and build knowledge on DSA issues. **Researchers have already begun to foster cooperative networks to continue building trust amongst each other, share knowledge and experiences on data access requests and other research matters and make connections across national borders.** For instance, in Germany, a collective of scientists from various disciplines has come together to discuss joint interests in Article 40.[22]

Yet, despite such efforts, it is often hard for academic and civil society researchers to develop and maintain communities, for instance, due to budgetary constraints or lack of institutional support. The German initiative requires volunteer coordination that is added on to participating scientists' regular teaching and research duties. This is a stark contrast to VLOPs and VLOSEs, who have vastly more financial resources and personnel for their internal data analysis, in addition to the already existing information asymmetries compared to external observers. Against this backdrop, regulators and the ECAT should continue to strive to build a network of outside experts (e.g., via consultations, an expert database, regular conferences and events and/or a permanent advisory council) and engage in public and science communication. This should include not only outreach on data access according to Article 40, but **also promote the use of other databases and reports established via the DSA** such as the ad repository (Article 39) and the database for content moderation decisions (Article 24(5)).

Monetary support from governments and/or platforms to engage in capacity building can be considered but requires striking a delicate balance to **maintain researchers' independence**. The corresponding experiences from the Code of Practice on Disinformation, in particular commitments 27 and 28, should be taken into account. Among other things mentioned in the Code, this specifically means that platforms should dedicate adequate resources to teams that work on data access requests and that communicate with researchers.

DSCs themselves need staff and resources dedicated to their own data access requests as well as researchers' applications. They should aim to establish strong and well-equipped data and research units.[23] This includes hiring a diverse set of experts (i.e., from various academic disciplines as well as practitioners from different fields such as content moderation, "trust and safety" or user experience design) and engaging in the type of capacity building highlighted above.

c) Would it be desirable and feasible to establish a common and precise language for DSCs, vetted researchers, VLOPs and VLOSEs to use when communicating about data access, e.g., by formulating a standard data dictionary and/or business glossary? How might this be implemented?

For the research questions addressed here, an overall "data dictionary" may not make sense as long as there is no technical interoperability. Platforms will change and adapt, which will make the maintenance of a general data dictionary challenging. Researchers would benefit from the establishment of general guidelines for data dictionaries, codebooks or schemas. A common understanding of what documentation should be made available is necessary. If platforms would describe their data structures in a way that makes comparisons possible, researchers could use this metadata to develop cross-platform studies (also see the response to question 1a).

[22] Disclaimer: SNV is part of this group, which has developed a joint response to this call for evidence as well.
[23] Julian Jaursch, "Here Is Why Digital Services Coordinators Should Establish Strong Research and Data Units", DSA Observatory, March 10, 2023.

4. Access to publicly available data

a) Not only vetted researchers will have greater opportunities for accessing data, all researchers meeting the conditions set out in Article 40(12) will be able to get direct access to publicly available data. What processes and mechanisms could be put in place to facilitate this access in your view?

We argue for a broad definition of public data that is comparable to what researchers with the technical capacities can already access today (see the response to question 1a).

No matter what process or mechanism is chosen with regards to access to public data, it should not duplicate but build on the efforts undertaken in the context of the Code of Practice on Disinformation. Any workable solutions found under the Code's commitment 26 (real-time access to public data) could be used to develop the Article 40(12) public data access.

Appendix

Data and metadata on TikTok videos can be obtained for research purposes through at least two different means: the official TikTok research API and directly scraping clients (e.g., the iPad app). This appendix documents the video data and metadata that SNV has obtained by scraping TikTok's website and Android clients.

To document the data that is available through scraping, we have produced two codebooks describing what information has been collected. The codebooks contain two layers of data variables collected and one with descriptions that attempt to infer what these variables contain. The names of the variables are directly attributable to the platform, but the descriptions have been created by our team and are limited to cases where we could identify the meaning of a variable, either by its name or its use on the website or in the app. The list of descriptions is a work in progress.

The codebooks have been produced considering subsets of the records gathered from the two clients mentioned above. These records were collected during different periods (see Table 1).

Table 1: Number of variables available per source and the period during which data was collected

Source	Variable count	Data collection
TikTok Research API	14	May 15, 2023
Web	118	June 2022 to May 2023
Android (multiple APIs)	657	January to May 2023

These codebooks are not meant as reference guides but shared only for illustrative purposes. Some variables in them may contain further granularity, meaning some nested properties are not being documented. Also note that in both cases, the metadata used to produce the codebooks may have evolved over time. **This means that the codebook is not a snapshot of the metadata available at one particular time but a summary of all the metadata available during the data collection period.** This also means that some fields may be unavailable today, may have different names or – despite our best efforts – may even be duplicated. The findings underscore the need for proper data governance documentation on behalf of platforms, in this case, on the evolution of the schema of the data being served to the clients.

The web codebook corresponds to a database containing records of video data and metadata scraped from the web client, and the schema used to produce it consisted of 100,000 random records. The app codebook data corresponds to the data gathered from over 200,000 videos. Both codebooks contain TikTok video data and metadata obtained through different APIs and their variable counts represent a conservative estimate, as some variables have not been investigated up to their deepest possible layer. For comparison, we have also included TikTok's official codebook for the "video object" returned by the official research API. Note that this API is not particular to a specific client.

The table on the following pages provides a detailed breakdown of what data points are available per source (TikTok API, web, app).

TikTok Research API			Web			App		
variable	subvariable	description	variable	subvariable	description	variable	subvariable	description
id	-none-	Unique identifier for the TikTok video. Also called “item_id” or “video_id”	adAuthorization	-none-	-unknown-	ad_aweme_source	-none-	-unknown-
create_time	-none-	UTC Unix epoch (in seconds) of when the TikTok video was posted. (Inherited field from TNS research API)	adLabelVersion	-none-	-unknown-	added_sound_music_info	album	-unknown-
username	-none-	The video’s author’s username	author	avatarLarger	Profile picture larger size	added_sound_music_info	artists	-unknown-
region_code	-none-	A two-digit code for the country the video was posted in	author	avatarMedium	Profile picture medium size	added_sound_music_info	audition_duration	-unknown-
video_description	-none-	The description of the video, also known as the title	author	avatarThumb	Profile picture thumbnail	added_sound_music_info	author	-unknown-
music_id	-none-	The music_id used in the video	author	commentSetting	-unknown-	added_sound_music_info	author_deleted	-unknown-
like_count	-none-	The number of likes the video has received.	author	duetSetting	-unknown-	added_sound_music_info	author_position	-unknown-
comment_count	-none-	The number of comments the video has received.	author	ftc	-unknown-	added_sound_music_info	avatar_medium	-unknown-
share_count	-none-	The number of shares the video has received.	author	isADVirtual	-unknown-	added_sound_music_info	avatar_thumb	-unknown-
view_count	-none-	The number of video views the video has received.	author	nickname	User display name	added_sound_music_info	binded_challenge_id	-unknown-
effect_ids	-none-	The list of effect ids applied on the video	author	openFavorite	-unknown-	added_sound_music_info	can_not_reuse	-unknown-
hashtag_names	-none-	The list of hashtag names that the video participates in	author	privateAccount	-unknown-	added_sound_music_info	collect_stat	-unknown-
playlist_id	-none-	The ID of playlist that the video belongs to	author	relation	-unknown-	added_sound_music_info	commercial_right_type	-unknown-
voice_to_text	-none-	Voice to text and subtitles (for videos that have voice to text features on, show the texts already generated)	author	secret	-unknown-	added_sound_music_info	cover_large	-unknown-

TikTok Research API			Web			App		
			author	secUid	*-unknown-*	added_sound_music_info	cover_medium	*-unknown-*
			author	signature	User bio	added_sound_music_info	cover_thumb	*-unknown-*
			author	stitchSetting	*-unknown-*	added_sound_music_info	dmv_auto_show	*-unknown-*
			author	ttSeller	*-unknown-*	added_sound_music_info	duration	*-unknown-*
			author	uniqueId	Unique user id	added_sound_music_info	duration_high_precision	*-unknown-*
			author	verified	Wether the user is verified or not	added_sound_music_info	external_song_info	*-unknown-*
			authorId	*-none-*	Unique author id	added_sound_music_info	extra	*-unknown-*
			authorSecId	*-none-*	*-unknown-*	added_sound_music_info	id	*-unknown-*
			authorStats	diggCount	Number of videos or comments the user has liked	added_sound_music_info	id_str	*-unknown-*
			authorStats	followerCount	Number of accounts that follow user	added_sound_music_info	is_audio_url_with_cookie	*-unknown-*
			authorStats	followingCount	Number of accounts that the user follows	added_sound_music_info	is_author_artist	*-unknown-*
			authorStats	heart	*-unknown-*	added_sound_music_info	is_commerce_music	*-unknown-*
			authorStats	heartCount	Number of likes received by the user	added_sound_music_info	is_matched_metadata	*-unknown-*
			authorStats	videoCount	Number of videos posted by user	added_sound_music_info	is_original	*-unknown-*
			avatarThumb	*-none-*	Profile picture thumbnail	added_sound_music_info	is_original_sound	*-unknown-*
			challenges	coverLarger	*-unknown-*	added_sound_music_info	is_pgc	*-unknown-*
			challenges	coverMedium	*-unknown-*	added_sound_music_info	is_play_music	*-unknown-*
			challenges	coverThumb	*-unknown-*	added_sound_music_info	is_shooting_allow	*-unknown-*
			challenges	desc	Description of the challenge	added_sound_music_info	lyric_short_position	*-unknown-*

TikTok Research API			Web			App		
			challenges	id	TikTok ID for the challenge	added_sound_music_info	matched_pgc_sound	-unknown-
			challenges	isCommerce	Whether the challenge is part of an ad campaign	added_sound_music_info	matched_song	-unknown-
			challenges	-none-	-unknown-	added_sound_music_info	mid	-unknown-
			challenges	profileLarger	-unknown-	added_sound_music_info	multi_bit_rate_play_info	-unknown-
			challenges	profileMedium	-unknown-	added_sound_music_info	music_release_info	-unknown-
			challenges	profileThumb	-unknown-	added_sound_music_info	mute_share	-unknown-
			challenges	stats.videoCount	Number of videos “using” this challenge?	added_sound_music_info	offline_desc	-unknown-
			challenges	stats.viewCount	Number of videos “using” this challenge that have been viewed?	added_sound_music_info	owner_handle	-unknown-
			challenges	title	Name of the challenge	added_sound_music_info	owner_id	-unknown-
			comments	-none-	Comments on the video	added_sound_music_info	owner_nickname	-unknown-
			createTime	-none-	Timestamp of when the video was posted	added_sound_music_info	play_url	-unknown-
			desc	-none-	-unknown-	added_sound_music_info	position	-unknown-
			digged	-none-	-unknown-	added_sound_music_info	prevent_download	-unknown-
			diversificationLabels	-none-	TikTok “For You” feed hashtags	added_sound_music_info	preview_end_time	-unknown-
			duetDisplay	-none-	-unknown-	added_sound_music_info	preview_start_time	-unknown-
			duetEnabled	-none-	-unknown-	added_sound_music_info	recommend_status	-unknown-
			duetInfo	duetFromId	-unknown-	added_sound_music_info	search_highlight	-unknown-

TikTok Research API			Web			App		
			effectStickers	-none-	Metadata on the effect used in a video	added_sound_music_info	sec_uid	-unknown-
			forFriend	-none-	-unknown-	added_sound_music_info	shoot_duration	-unknown-
			id	-none-	-unknown-	added_sound_music_info	source_platform	-unknown-
			indexEnabled	-none-	-unknown-	added_sound_music_info	status	-unknown-
			isAd	-none-	-unknown-	added_sound_music_info	strong_beat_url	-unknown-
			itemCommentStatus	-none-	-unknown-	added_sound_music_info	tag_list	-unknown-
			itemMute	-none-	-unknown-	added_sound_music_info	title	-unknown-
			music	album	-unknown-	added_sound_music_info	tt_to_dsp_song_infos	-unknown-
			music	authorName	-unknown-	added_sound_music_info	user_count	-unknown-
			music	coverLarge	-unknown-	added_sound_music_info	video_duration	-unknown-
			music	coverMedium	-unknown-	allow_gift	-none-	-unknown-
			music	coverThumb	-unknown-	anchors	-none-	-unknown-
			music	duration	Length in seconds	anchors_extras	-none-	-unknown-
			music	id	Unique ID of the video's music	animated_image_info	effect	-unknown-
			music	original	Whether music originated from the user	animated_image_info	type	-unknown-
			music	playUrl	Link to mp3 file of the video's music	author	accept_private_policy	Agreement with privacy policy
			music	scheduleSearchTime	-unknown-	author	account_labels	-unknown-
			music	title	-unknown-	author	account_region	Region where the account is registered
			nickname	-none-	Profile display name	author	ad_cover_url	-unknown-
			officalItem	-none-	-unknown-	author	advance_feature_item_order	-unknown-

TikTok Research API			Web			App		
			originalItem	-none-	-unknown-	author	advanced_feature_info	-unknown-
			privateItem	-none-	-unknown-	author	apple_account	Apple ID
			scheduleTime	-none-	-unknown-	author	authority_status	-unknown-
			secret	-none-	-unknown-	author	avatar_168x168	Profile picture of specific size
			shareEnabled	-none-	-unknown-	author	avatar_300x300	Profile picture of specific size
			showNotPass	-none-	-unknown-	author	avatar_larger	Profile picture larger size
			stats	commentCount	Number of comments video had at time of scraping	author	avatar_medium	Profile picture medium size
			stats	diggCount	Number of likes at time of scraping	author	avatar_thumb	Profile picture thumbnail
			stats	playCount	Number of times video had been played at time of scraping	author	avatar_uri	-unknown-
			stats	shareCount	Number of times video was shared at time of scraping	author	aweme_count	-unknown-
			stickersOnItem	stickerText	Text displayed on video sticker	author	bind_phone	-unknown-
			stickersOnItem	stickerType	-unknown-	author	bold_fields	-unknown-
			stitchDisplay	-none-	-unknown-	author	can_message_follow_status_list	-unknown-
			stitchEnabled	-none-	Wether stitching is enabled or not	author	can_set_geofencing	-unknown-
			takeDown	-none-	-unknown-	author	cha_list	-unknown-
			textExtra	awemeId	-unknown-	author	comment_filter_status	-unknown-

TikTok Research API			Web			App		
			textExtra	end	Hashtag text position end	author	comment_setting	-unknown-
			textExtra	hashtagId	Unique hashtag id	author	commerce_user_level	-unknown-
			textExtra	hashtagName	Hashtag text	author	cover_url	-unknown-
			textExtra	isCommerce	Whether the hashtag is part of an ad campaign?	author	create_time	-unknown-
			textExtra	secUid	-unknown-	author	custom_verify	-unknown-
			textExtra	start	Hashtag text position start	author	cv_level	-unknown-
			textExtra	subType	-unknown-	author	download_prompt_ts	-unknown-
			textExtra	type	-unknown-	author	download_setting	-unknown-
			textExtra	userId	-unknown-	author	duet_setting	-unknown-
			textExtra	userUniqueId	-unknown-	author	duetSetting	-unknown-
			video	bitrate	Bitrate	author	enterprise_verify_reason	-unknown-
			video	codecType	Compression standard	author	events	-unknown-
			video	cover	Still image that available in video feed	author	favoriting_count	-unknown-
			video	definition	Resolution	author	fb_expire_time	-unknown-
			video	downloadAddr	-unknown-	author	follow_status	-unknown-
			video	duration	Duration in seconds	author	follower_count	Number of accounts that follow user
			video	dynamicCover	Short clip of video	author	follower_status	-unknown-
			video	encodedType	Encoding label	author	followers_detail	-unknown-
			video	encodeUserTag	-unknown-	author	following_count	Number of accounts that the user follows

TikTok Research API			Web			App		
			video	format	Resolution	author	friends_status	-unknown-
			video	height	Vertical size in pixels	author	geofencing	-unknown-
			video	id	Unique video id	author	google_account	Google account username
			video	originCover	First frame of the video	author	has_email	-unknown-
			video	playAddr	-unknown-	author	has_facebook_token	-unknown-
			video	ratio	Resolution	author	has_insights	-unknown-
			video	reflowCover	-unknown-	author	has_orders	-unknown-
			video	shareCover	-unknown-	author	has_twitter_token	-unknown-
			video	subtitleInfos	Subtitle metadata	author	has_youtube_token	-unknown-
			video	subtitles	Subtitle content	author	hide_search	-unknown-
			video	videoQuality	Quality label	author	homepage_bottom_toast	-unknown-
			video	width	Horizontal size in pixels	author	id	-unknown-
			warnInfo	-none-	-unknown-	author	ins_id	-unknown-
						author	is_ad_fake	-unknown-
						author	is_block	-unknown-
						author	is_discipline_member	-unknown-
						author	is_phone_binded	-unknown-
						author	is_star	-unknown-
						author	item_list	-unknown-
						author	language	-unknown-
						author	live_agreement	-unknown-
						author	live_commerce	-unknown-
						author	live_verify	-unknown-
						author	mention_status	-unknown-
						author	mutual_relation_avatars	-unknown-
						author	need_points	-unknown-
						author	need_recommend	-unknown-
						author	nickname	User display name

TikTok Research API			Web			App		
						author	platform_sync_info	-unknown-
						author	prevent_download	-unknown-
						author	qa_status	-unknown-
						author	react_setting	-unknown-
						author	region	-unknown-
						author	relative_users	-unknown-
						author	room_id	-unknown-
						author	search_highlight	-unknown-
						author	sec_uid	-unknown-
						author	secret	-unknown-
						author	share_info	-unknown-
						author	share_qrcode_uri	-unknown-
						author	shield_comment_notice	-unknown-
						author	shield_digg_notice	-unknown-
						author	shield_edit_field_info	-unknown-
						author	shield_follow_notice	-unknown-
						author	short_id	-unknown-
						author	show_image_bubble	-unknown-
						author	signature	User bio
						author	special_account	-unknown-
						author	special_lock	-unknown-
						author	status	-unknown-
						author	sth2	-unknown-
						author	stitch_setting	-unknown-
						author	story_status	-unknown-
						author	total_favorited	-unknown-
						author	tw_expire_time	-unknown-
						author	twitter_id	Twitter ID number
						author	twitter_name	Twitter display name
						author	type_label	-unknown-
						author	uid	-unknown-
						author	unique_id	Unique user ID

TikTok Research API			Web			App		
						author	unique_id_modify_time	Last user ID modification timestamp
						author	user_canceled	-unknown-
						author	user_mode	-unknown-
						author	user_now_pack_info	-unknown-
						author	user_period	-unknown-
						author	user_profile_guide	-unknown-
						author	user_rate	-unknown-
						author	user_tags	-unknown-
						author	verification_type	-unknown-
						author	verify_info	-unknown-
						author	video_icon	-unknown-
						author	white_cover_url	-unknown-
						author	with_commerce_entry	-unknown-
						author	with_shop_entry	-unknown-
						author	youtube_channel_id	-unknown-
						author	youtube_channel_title	-unknown-
						author	youtube_expire_time	-unknown-
						author_user_id	-none-	Unique user id
						aweme_acl	download_general	-unknown-
						aweme_acl	download_mask_panel	-unknown-
						aweme_acl	platform_list	-unknown-
						aweme_acl	share_general	-unknown-
						aweme_acl	share_list_status	-unknown-
						aweme_acl	share_third_platform	-unknown-
						aweme_detail	id	-unknown-
						aweme_detail	log_pb	-unknown-
						aweme_detail	status_code	-unknown-
						aweme_detail	status_msg	-unknown-

TikTok Research API			Web			App		
						aweme_id	-none-	-unknown-
						aweme_info	anchors	-unknown-
						aweme_info	anchors_extras	-unknown-
						aweme_info	author	-unknown-
						aweme_info	author_user_id	-unknown-
						aweme_info	aweme_acl	-unknown-
						aweme_info	aweme_id	-unknown-
						aweme_info	aweme_type	-unknown-
						aweme_info	bodydance_score	-unknown-
						aweme_info	branded_content_accounts	-unknown-
						aweme_info	cc_template_info	-unknown-
						aweme_info	cha_list	-unknown-
						aweme_info	challenge_position	-unknown-
						aweme_info	cmt_swt	-unknown-
						aweme_info	collect_stat	-unknown-
						aweme_info	commerce_config_data	-unknown-
						aweme_info	commerce_info	-unknown-
						aweme_info	content_desc	-unknown-
						aweme_info	content_desc_extra	-unknown-
						aweme_info	cover_labels	-unknown-
						aweme_info	create_time	-unknown-
						aweme_info	desc	-unknown-
						aweme_info	desc_language	-unknown-
						aweme_info	disable_search_trending_bar	-unknown-
						aweme_info	distance	-unknown-
						aweme_info	distribute_type	-unknown-
						aweme_info	follow_up_item_id_groups	-unknown-
						aweme_info	follow_up_publish_from_id	-unknown-
						aweme_info	geofencing	-unknown-
						aweme_info	geofencing_regions	-unknown-

TikTok Research API			Web			App		
						aweme_info	green_screen_materials	-unknown-
						aweme_info	group_id	-unknown-
						aweme_info	group_id_list	-unknown-
						aweme_info	has_vs_entry	-unknown-
						aweme_info	have_dashboard	-unknown-
						aweme_info	hybrid_label	-unknown-
						aweme_info	image_infos	-unknown-
						aweme_info	interact_permission	-unknown-
						aweme_info	interaction_stickers	-unknown-
						aweme_info	is_ads	-unknown-
						aweme_info	is_description_translatable	-unknown-
						aweme_info	is_hash_tag	-unknown-
						aweme_info	is_pgcshow	-unknown-
						aweme_info	is_preview	-unknown-
						aweme_info	is_relieve	-unknown-
						aweme_info	is_text_sticker_translatable	-unknown-
						aweme_info	is_top	-unknown-
						aweme_info	is_vr	-unknown-
						aweme_info	item_comment_settings	-unknown-
						aweme_info	item_duet	-unknown-
						aweme_info	item_react	-unknown-
						aweme_info	item_stitch	-unknown-
						aweme_info	label_top	-unknown-
						aweme_info	label_top_text	-unknown-
						aweme_info	long_video	-unknown-
						aweme_info	mask_infos	-unknown-
						aweme_info	misc_info	-unknown-
						aweme_info	music	-unknown-
						aweme_info	music_begin_time_in_ms	-unknown-
						aweme_info	music_selected_from	-unknown-

TikTok Research API			Web			App		
						aweme_info	music_title_style	-unknown-
						aweme_info	need_trim_step	-unknown-
						aweme_info	need_vs_entry	-unknown-
						aweme_info	nickname_position	-unknown-
						aweme_info	no_selected_music	-unknown-
						aweme_info	origin_comment_ids	-unknown-
						aweme_info	playlist_blocked	-unknown-
						aweme_info	poi_re_tag_signal	-unknown-
						aweme_info	position	-unknown-
						aweme_info	prevent_download	-unknown-
						aweme_info	products_info	-unknown-
						aweme_info	question_list	-unknown-
						aweme_info	rate	-unknown-
						aweme_info	reference_tts_voice_ids	-unknown-
						aweme_info	reference_voice_filter_ids	-unknown-
						aweme_info	region	-unknown-
						aweme_info	risk_infos	-unknown-
						aweme_info	search_desc	-unknown-
						aweme_info	search_highlight	-unknown-
						aweme_info	share_info	-unknown-
						aweme_info	share_url	-unknown-
						aweme_info	sort_label	-unknown-
						aweme_info	statistics	-unknown-
						aweme_info	status	-unknown-
						aweme_info	text_extra	-unknown-
						aweme_info	text_sticker_major_lang	-unknown-
						aweme_info	tts_voice_ids	-unknown-
						aweme_info	uniqid_position	-unknown-
						aweme_info	upvote_preload	-unknown-
						aweme_info	user_digged	-unknown-
						aweme_info	video	-unknown-
						aweme_info	video_control	-unknown-

TikTok Research API			Web			App		
						aweme_info	video_labels	-unknown-
						aweme_info	video_text	-unknown-
						aweme_info	voice_filter_ids	-unknown-
						aweme_info	with_promotional_music	-unknown-
						aweme_info	without_watermark	-unknown-
						aweme_type	-none-	-unknown-
						behind_the_song_music_ids	-none-	-unknown-
						behind_the_song_video_music_ids	-none-	-unknown-
						bodydance_score	-none-	-unknown-
						branded_content_accounts	-none-	-unknown-
						cc_template_info	author_name	CapCut template author
						cc_template_info	clip_count	-unknown-
						cc_template_info	desc	CapCut template description
						cc_template_info	duration_milliseconds	-unknown-
						cc_template_info	related_music_id	-unknown-
						cc_template_info	template_id	CapCut template id
						cha_list	author	-unknown-
						cha_list	banner_list	-unknown-
						cha_list	cha_attrs	-unknown-
						cha_list	cha_name	-unknown-
						cha_list	cid	-unknown-
						cha_list	collect_stat	-unknown-
						cha_list	connect_music	-unknown-
						cha_list	desc	-unknown-
						cha_list	extra_attr	-unknown-
						cha_list	hashtag_profile	-unknown-
						cha_list	is_challenge	-unknown-
						cha_list	is_commerce	-unknown-
						cha_list	is_pgcshow	-unknown-
						cha_list	schema	-unknown-
						cha_list	search_highlight	-unknown-
						cha_list	share_info	-unknown-

TikTok Research API			Web			App		
						cha_list	show_items	-unknown-
						cha_list	sub_type	-unknown-
						cha_list	type	-unknown-
						cha_list	use_count	-unknown-
						cha_list	user_count	-unknown-
						cha_list	view_count	-unknown-
						challenge_position	-none-	-unknown-
						challenges	desc	String with the challenge description
						challenges	id	Challenge ID
						challenges	isCommerce	Whether the challenge is part of a commercial campaign
						challenges	title	Challenge hashtag
						cmt_swt	-none-	-unknown-
						collect_stat	-none-	-unknown-
						commerce_config_data	-none-	-unknown-
						commerce_info	ad_source	-unknown-
						commerce_info	adv_promotable	-unknown-
						commerce_info	auction_ad_invited	-unknown-
						commerce_info	bc_label_test_text	-unknown-
						commerce_info	dark_post_source	-unknown-
						commerce_info	dark_post_status	-unknown-
						commerce_info	mission_id	-unknown-
						commerce_info	mission_item_status	-unknown-
						commerce_info	prevent_share	-unknown-
						commerce_info	study_id	-unknown-
						commerce_info	with_comment_filter_words	-unknown-
						content_desc	-none-	-unknown-
						content_desc_extra	-none-	-unknown-
						content_original_type	-none-	-unknown-
						cover_labels	-none-	-unknown-
						create_time	-none-	Timestamp of when the video was posted

TikTok Research API			Web			App		
						desc	-none-	-unknown-
						desc_language	-none-	-unknown-
						disable_search_trending_bar	-none-	-unknown-
						distance	-none-	-unknown-
						distribute_type	-none-	-unknown-
						doc_id	-none-	-unknown-
						doc_type	-none-	-unknown-
						duet_info	original_item_id	-unknown-
						feedback_type	-none-	-unknown-
						follow_up_publish_from_id	-none-	-unknown-
						game_info	game_score	-unknown-
						game_info	game_type	-unknown-
						geofencing	-none-	-unknown-
						geofencing_regions	-none-	-unknown-
						green_screen_materials	-none-	-unknown-
						group_id	-none-	-unknown-
						group_id_list	GroupdIdList0	-unknown-
						group_id_list	GroupdIdList1	-unknown-
						has_danmaku	-none-	Contains danmaku-style comments
						has_promote_entry	-none-	-unknown-
						has_vs_entry	-none-	-unknown-
						have_dashboard	-none-	-unknown-
						hybrid_label	-none-	-unknown-
						id	-none-	-unknown-
						image_infos	-none-	-unknown-
						image_post_info	image_post_cover	-unknown-
						image_post_info	images	-unknown-
						image_post_info	music_volume	-unknown-
						image_post_info	title	-unknown-
						interact_permission	allow_adding_to_story	-unknown-
						interact_permission	allow_create_sticker	-unknown-
						interact_permission	duet	-unknown-

TikTok Research API			Web			App		
						interact_permission	duet_privacy_setting	-unknown-
						interact_permission	stitch	-unknown-
						interact_permission	stitch_privacy_setting	-unknown-
						interact_permission	upvote	-unknown-
						interaction_stickers	-none-	-unknown-
						interaction_tag_info	interest_level	-unknown-
						interaction_tag_info	tagged_users	-unknown-
						interaction_tag_info	video_label_text	-unknown-
						is_ads	-none-	-unknown-
						is_description_translatable	-none-	-unknown-
						is_hash_tag	-none-	-unknown-
						is_nff_or_nr	-none-	-unknown-
						is_on_this_day	-none-	-unknown-
						is_pgcshow	-none-	-unknown-
						is_preview	-none-	-unknown-
						is_relieve	-none-	-unknown-
						is_text_sticker_translatable	-none-	-unknown-
						is_top	-none-	-unknown-
						is_vr	-none-	-unknown-
						item_comment_settings	-none-	-unknown-
						item_duet	-none-	-unknown-
						item_react	-none-	-unknown-
						item_stitch	-none-	-unknown-
						label_top	height	-unknown-
						label_top	uri	-unknown-
						label_top	url_list	-unknown-
						label_top	width	-unknown-
						label_top_text	-none-	-unknown-
						log_pb	impr_id	-unknown-
						long_video	-none-	-unknown-
						mask_infos	-none-	-unknown-
						misc_info	-none-	-unknown-
						muf_comment_info_v2	-none-	-unknown-

TikTok Research API			Web			App		
						music	album	Music album
						music	artists	Music artists
						music	audition_duration	-unknown-
						music	author	-unknown-
						music	author_deleted	-unknown-
						music	author_position	-unknown-
						music	avatar_medium	-unknown-
						music	avatar_thumb	-unknown-
						music	binded_challenge_id	-unknown-
						music	can_not_reuse	-unknown-
						music	collect_stat	-unknown-
						music	commercial_right_type	-unknown-
						music	cover_large	-unknown-
						music	cover_medium	-unknown-
						music	cover_thumb	-unknown-
						music	coverLarge	-unknown-
						music	coverMedium	-unknown-
						music	coverThumb	-unknown-
						music	dmv_auto_show	-unknown-
						music	duration	Length in seconds
						music	duration_high_precision	-unknown-
						music	external_song_info	-unknown-
						music	extra	-unknown-
						music	id	Music unique ID
						music	id_str	-unknown-
						music	is_audio_url_with_cookie	-unknown-
						music	is_author_artist	-unknown-
						music	is_commerce_music	-unknown-
						music	is_matched_metadata	-unknown-
						music	is_original	-unknown-

TikTok Research API			Web			App		
						music	is_original_sound	-unknown-
						music	is_pgc	-unknown-
						music	is_play_music	-unknown-
						music	is_shooting_allow	-unknown-
						music	lyric_short_position	-unknown-
						music	matched_pgc_sound	-unknown-
						music	matched_song	-unknown-
						music	mid	-unknown-
						music	multi_bit_rate_play_info	-unknown-
						music	music_release_info	-unknown-
						music	mute_share	-unknown-
						music	offline_desc	-unknown-
						music	owner_handle	-unknown-
						music	owner_id	-unknown-
						music	owner_nickname	-unknown-
						music	play_url	-unknown-
						music	playUrl	-unknown-
						music	position	-unknown-
						music	prevent_download	-unknown-
						music	preview_end_time	-unknown-
						music	preview_start_time	-unknown-
						music	recommend_status	-unknown-
						music	search_highlight	-unknown-
						music	sec_uid	-unknown-
						music	shoot_duration	-unknown-
						music	source_platform	-unknown-
						music	status	-unknown-
						music	strong_beat_url	-unknown-
						music	tag_list	-unknown-
						music	title	-unknown-
						music	tt_to_dsp_song_infos	-unknown-
						music	user_count	Number of users using music

TikTok Research API			Web			App		
						music	video_duration	-unknown-
						music_begin_time_in_ms	-none-	-unknown-
						music_end_time_in_ms	-none-	-unknown-
						music_selected_from	-none-	-unknown-
						music_title_style	-none-	-unknown-
						music_volume	-none-	-unknown-
						need_trim_step	-none-	-unknown-
						need_vs_entry	-none-	-unknown-
						nickname_position	-none-	-unknown-
						no_selected_music	-none-	-unknown-
						origin_comment_ids	-none-	-unknown-
						origin_volume	-none-	-unknown-
						original_client_text	markup_text	-unknown-
						original_client_text	text_extra	-unknown-
						playlist_blocked	-none-	-unknown-
						playlist_info	index	-unknown-
						playlist_info	is_default_name	-unknown-
						playlist_info	item_total	-unknown-
						playlist_info	mix_id	-unknown-
						playlist_info	name	-unknown-
						poi_data	address_info	-unknown-
						poi_data	collect_info	-unknown-
						poi_data	comment_anchor	-unknown-
						poi_data	info_source	-unknown-
						poi_data	poi_id	-unknown-
						poi_data	poi_mapkit_collect	-unknown-
						poi_data	poi_name	-unknown-
						poi_data	poi_type	-unknown-
						poi_data	video_anchor	-unknown-
						poi_data	video_count	-unknown-
						poi_re_tag_signal	-none-	-unknown-
						position	-none-	-unknown-
						prevent_download	-none-	-unknown-
						products_info	-none-	-unknown-
						promote_toast	-none-	-unknown-

TikTok Research API			Web			App		
						promote_toast_key	-none-	-unknown-
						provider_doc_id_str	-none-	-unknown-
						question_list	-none-	-unknown-
						rate	-none-	-unknown-
						recReason	reasonID	-unknown-
						recReasons	-none-	-unknown-
						reference_tts_voice_ids	-none-	-unknown-
						reference_voice_filter_ids	-none-	-unknown-
						region	-none-	-unknown-
						retry_type	-none-	-unknown-
						risk_infos	content	-unknown-
						risk_infos	notice	-unknown-
						risk_infos	risk_sink	-unknown-
						risk_infos	type	-unknown-
						risk_infos	url	-unknown-
						risk_infos	vote	-unknown-
						risk_infos	warn	-unknown-
						search_highlight	-none-	-unknown-
						searchsuggestion	searchstring	-unknown-
						share_info	bool_persist	-unknown-
						share_info	now_invitation_card_image_urls	-unknown-
						share_info	share_desc	-unknown-
						share_info	share_desc_info	-unknown-
						share_info	share_link_desc	-unknown-
						share_info	share_quote	-unknown-
						share_info	share_signature_desc	-unknown-
						share_info	share_signature_url	-unknown-
						share_info	share_title	-unknown-
						share_info	share_title_myself	-unknown-
						share_info	share_title_other	-unknown-
						share_info	share_url	-unknown-
						share_info	whatsapp_desc	Template text when sharing video with WhatsApp

TikTok Research API			Web			App		
						share_url	-none-	-unknown-
						should_add_creator_tts_watermark_when_downloading	-none-	-unknown-
						sort_label	-none-	-unknown-
						source	-none-	-unknown-
						special_mode	-none-	-unknown-
						statistics	aweme_id	-unknown-
						statistics	collect_count	-unknown-
						statistics	comment_count	-unknown-
						statistics	digg_count	-unknown-
						statistics	download_count	Number of times the video has been downloaded
						statistics	forward_count	Number of times the video has been forwarded
						statistics	lose_comment_count	-unknown-
						statistics	lose_count	-unknown-
						statistics	play_count	Number of times the video has been played
						statistics	share_count	Number of times the video has been shared overall
						statistics	whatsapp_share_count	Number of times video has been shared via WhatsApp
						status	allow_comment	-unknown-
						status	allow_share	-unknown-
						status	aweme_id	-unknown-
						status	download_status	-unknown-
						status	in_reviewing	-unknown-
						status	is_delete	Whether video has been deleted
						status	is_prohibited	-unknown-
						status	private_status	-unknown-

TikTok Research API			Web			App		
						status	review_result	-unknown-
						status	reviewed	-unknown-
						status	self_see	-unknown-
						status	video_mute	-unknown-
						status_code	-none-	-unknown-
						status_msg	-none-	-unknown-
						sticker_detail	attributions	-unknown-
						sticker_detail	children	-unknown-
						sticker_detail	icon_url	-unknown-
						sticker_detail	id	-unknown-
						sticker_detail	linked_anchors	-unknown-
						sticker_detail	name	-unknown-
						sticker_detail	owner_id	-unknown-
						sticker_detail	sec_uid	-unknown-
						sticker_detail	tags	-unknown-
						stickers	-none-	-unknown-
						suggest_words	suggest_words	-unknown-
						support_danmaku	-none-	Supports danmaku-style comments
						text_extra	awemeId	-unknown-
						text_extra	end	Hashtag text position end
						text_extra	hashtag_id	Unique hashtag id
						text_extra	hashtag_name	Hashtag text
						text_extra	is_commerce	Whether the hashtag is part of an ad campaign
						text_extra	sec_iud	-unknown-
						text_extra	start	Hashtag text position start
						text_extra	subType	-unknown-
						text_extra	type	-unknown-
						text_extra	user_id	-unknown-
						text_sticker_major_lang	-none-	-unknown-
						tts_voice_ids	-none-	-unknown-
						type	-none-	-unknown-
						uniqid_position	-none-	-unknown-

TikTok Research API			Web			App		
						upvote_preload	need_pull_upvote_info	-unknown-
						user_digged	-none-	-unknown-
						video	ai_dynamic_cover	-unknown-
						video	ai_dynamic_cover_bak	-unknown-
						video	animated_cover	-unknown-
						video	big_thumbs	-unknown-
						video	bit_rate	Bitrate
						video	cdn_url_expired	-unknown-
						video	cla_info	-unknown-
						video	cover	-unknown-
						video	cover_is_custom	-unknown-
						video	CoverTsp	-unknown-
						video	download_addr	-unknown-
						video	downloadAddr	-unknown-
						video	duration	Duration in seconds
						video	dynamic_cover	-unknown-
						video	has_watermark	-unknown-
						video	height	Vertical size in pixels
						video	is_bytevc1	Codec is ByteVC1
						video	is_callback	-unknown-
						video	is_long_video	-unknown-
						video	meta	-unknown-
						video	misc_download_addrs	-unknown-
						video	need_set_token	-unknown-
						video	origin_cover	-unknown-
						video	play_addr	Video playback address
						video	play_addr_bytevc1	Video playback address for ByteVC1 codec
						video	play_addr_h264	Video playback address for H264 codec
						video	playAddr	-unknown-
						video	ratio	-unknown-
						video	source_HDR_type	-unknown-

TikTok Research API			Web			App		
						video	subtitles	-unknown-
						video	tags	-unknown-
						video	video_model	-unknown-
						video	width	Horizontal size in pixels
						video	zoomCover	-unknown-
						video_control	allow_download	Download permission
						video_control	allow_duet	Duet permission
						video_control	allow_dynamic_wallpaper	Dynamic wallpaper permission
						video_control	allow_music	-unknown-
						video_control	allow_react	-unknown-
						video_control	allow_stitch	Stitch permission
						video_control	draft_progress_bar	-unknown-
						video_control	prevent_download_type	-unknown-
						video_control	share_type	-unknown-
						video_control	show_progress_bar	-unknown-
						video_control	timer_status	-unknown-
						video_labels	-none-	-unknown-
						video_reply_info	alias_comment_id	-unknown-
						video_reply_info	aweme_id	-unknown-
						video_reply_info	collect_stat	-unknown-
						video_reply_info	comment_id	-unknown-
						video_reply_info	comment_msg	-unknown-
						video_reply_info	comment_user_id	-unknown-
						video_reply_info	user_avatar	-unknown-
						video_reply_info	user_name	-unknown-
						video_text	-none-	-unknown-
						voice_filter_ids	-none-	-unknown-
						with_promotional_music	-none-	-unknown-
						without_watermark	-none-	-unknown-