Ref. Ares(2025)933487 - 06/02/2025

Meeting with [redacted] Youtube

Online, 25 October 2023, 17:30

Member: Wojtek Talko

Steering brief

Scene setter

You will meet [redacted] YouTube [redacted] He has worked in the company since 2015 where he managed a number of its services [redacted] He worked for [redacted]

YouTube, which is owned by Google, is one of the largest online video sharing and social media platforms in the world. It has around 400 million monthly active users in the EU and was designated as a very large online platform (VLOP) under the Digital Services Act (DSA) on 25 April 2023 by the European Commission.

On 28 August, YouTube submitted its first risk assessment report under the DSA, where it identifies, analyses and assesses systemic risks stemming from its services and systems. The Commission is currently assessing this report.

On 13 October 2023, Commissioner Breton addressed a letter to YouTube, requesting the platform to vigilantly ensure that Hamas-connected disinformation and illegal content does not spread on YouTube (letter in the annex).

YouTube's [redacted] replied to the letter on 18 October 2023.

The Commission has urgently adopted a Recommendation encouraging Member States to respond in a coordinated and consistent manner to incidents, in particular arising from the dissemination of illegal content, posing a clear risk of intimidating groups of population and destabilising political and social structures in the Union or parts thereof. The Recommendation sets out mechanisms of preparedness, cooperation and coordination between the Commission and the Member States ahead of the full application of the DSA on 17 February 2024.

The fact-checking and disinformation community expresses regular concerns about the fact that YouTube does not take fact-checking and countering disinformation seriously enough (the least active among signatories of the

Code). Apparently no European fact-checking association has an agreement with YouTube (which is one of the Code's commitments). The reported funding for fact-checking organizations mainly goes to training and media literacy campaigns, but no or very little funding goes to invest in fact-checking on YouTube. The organisations also expressed concerns about access to data.

What do we want:

- To understand better how YouTube sees the impact of the Israel – Hamas conflict on its platform (illegal and violent content, disinformation, anti-Semitism).

- To ask about the situation with the pro-Kremlin disinformation or if there are any recent trends.

- To call on them to step work with fact-checkers and fulfill their obligations under the Code.

- To push YouTube towards strong engagement with DSA/DMA obligations and the Code's development of a risk assessment framework and rapid response system for crisis situations.

- Call on Youtube to work with us ahead of EP elections on safe information space.

- Ask about the situation of Russian independent media / journalists on Youtube.

What do we expect to hear:

- That YouTube will try to convince the Commission that YouTube is doing everything it can to mitigate the harm of misinformation and disinformation surrounding the Israel-Hamas conflict and complying with DSA/DMA conflicts.

(YouTube indicated they will not ask about political ads, AI Act or the Media Freedom Act. The information below about AI is therefore more interesting than needed for the call.)

With regards to artificial intelligence, YouTube makes various uses of AI. This includes algorithms for recommending new videos based on the history of watched videos (these algorithms are mainly covered in regulations such as the DSA package). YouTube has vast experience in multimodal (text comments, audio, video) content moderation, including policies and practices regarding spam, deceptive practices, scams, and copyright issues.

More recently YouTube stared to use generative AI for a range of purposes, including AI for subtitle and voice-over generation and translation, video and audio editing capabilities (beat matching, voice changing) and AI-informed guidance to increase audience and reach. Those tools are in part provided by their parent company Google. Planned uses include AI for content moderation.

YouTube has become a prime target platform for AI generated content. It has hence been struggling to harmonise its users and its own ambitions on using generative AI with the industries interest in safeguarding copyrights. YouTube is dependent on the music industry for e.g. blanket licences as it would be otherwise much more difficult to operate for them.

In response, in August 2023, YouTube announced a collaboration with the music industry to establish rules for generative AI content, promising to balance rights protection with creator-led innovation, including figuring out ways to monetise AI-generated content under the inclusion of the original artists. These efforts seem to be of high priority to [redacted]. YouTube furthermore plans to use AI tools to flag content in violation of its policies.

In September 2023, Google announced the integration of its chat bot Bard (their competitor to ChatGPT, based on generative AI) into YouTube.

Many of these technologies (such as generative AI tools, including chatbots) would have to conform to some provisions of the AI act (e.g. transparency requirements).

Key messages

Israel, Hamas

- The onslaught of mis- and disinformation around the Hamas' terrorist attacks against Israel and the ensuing spiral of violence shows once again the extreme importance of all major online platforms having fast and effective mitigation measures in place.
- Under the DSA you have to take measures on removal of illegal content. This is highly relevant in this situation. As you know, this includes reporting on risk assessment and risk mitigation measures, providing vetted researchers access to data, and crisis management obligations.

- In addition, if content moderation tools or recommender systems are found to amplify disinformation, those very large online platforms will need to take appropriate mitigation measures to address such risks.
- The DSA is also complemented by the Code of Practice provides a flexible framework with specific commitments and measures to fulfil such obligation and counter disinformation related risks in crisis situations.
- Reducing the visibility of false information, promoting authoritative information, preventing the spread of disinformation via advertising, giving users access to fact-checks provided by independent fact-checking organisations – these are just some of the commitments that YouTube has taken under the Code, and that we now expect to see strongly applied in this crisis.

- I would like to facilitate better and faster system of flagging such content (illegal in the first place but also disinformation) and would like to convene those who monitor the space (incl. Antisemitism envoys, EDMO) and platforms. I hope that I count on your engagement in the next two weeks.

Disinformation during electoral periods

- The DSA requires very large online platforms such as YouTube to take measures to protect the integrity of elections and freedom of expression. These platforms are important for public discourse around elections.
- As stated in the European Democracy Action Plan, the rapid growth of online campaigning and online platforms has opened new vulnerabilities for democracies and made it more difficult to protect the democratic process from disinformation and other manipulation.
- The DSA offers instruments to contribute to fight the spread of disinformation online.
- In particular, platforms such as YouTube have to conduct risk assessments in relation to potential negative effects on electoral processes and civic discourse, apply election related terms and conditions consistently and provide more detailed information and access to data on their platforms.
- In addition, the Working Group on Elections under the Code of Practice on Disinformation will serve as a wider forum of discussion, including multiple stakeholders and will continue to run in parallel to identify best practices.

Defensives + background

Defensives

Is the Commission going to publish an assessment of YouTube's reports?

- The Commission has published a preliminary overview of the baseline reports in February.

- The baseline reports marked an important first step in establishing the monitoring and reporting under the new, 2022 Code of Practice.

- Our first impression regarding this new reporting round is that today's reports attest of the signatories' efforts to provide further insight into their actions to fight disinformation, with more stable data covering a full 6-month reporting period.

- In the reports published today, platforms generally make improvements in providing more granular and insightful data, closing some data gaps identified in their baseline reports (from February 2023). At the same time, further efforts are needed to provide more targeted, complete and meaningful data.

- This is of course only a very first impression. The Commission services will further analyse the reports submitted. [If and in what form a potential assessment of the reports will be published will be decided in due course. Of note that the reports are overall around 900 pages so Commission services will need more time to digest and analyse them in detail]

What are the next steps for the Code?

- In view of the important risks posed by disinformation to our democracies and societies the Commission expects signatories to continue their work and increase their efforts in implementing the Code. Particular attention should be paid to fight disinformation on Ukraine and around elections.

- The next set of reports are due in early 2024 (with information and data covering the second half of 2023) and should provide further insights on the Code's implementation. Those reports will also contain information on how signatories are preparing and putting in place measures to reduce the spread of disinformation ahead of the 2024 EU elections, while continuing to report on their efforts in the context of the war in Ukraine.

How would General Purpose AI systems/ foundation models/ generative AI like ChatGPT be regulated by the EU AI Act?

- According to the Commission proposal, the requirements and obligations are applied to systems that are considered high-risk.

- There are transparency requirements for systems intended to interact with humans (e.g. chatbots) and systems that generate certain synthetic content (e.g. audio, video, images), to disclose that the content has been artificially generated.

- Furthermore, there seems to be a consensus between the EU co-legislators that targeted rules are needed for these very capable foundation models, considering that these models are expected to be integrated or used in a very large number of AI based products.

- This could include certain obligations for foundation models and an obligation for the upstream provider to provide information to the downstream provider and possibly safeguards against illegal content, and watermarking.

Designation process should not lead to unequal treatment of competitors – and you have now designated some but, potentially, not all VLOPs/VLOSEs

- We can assure you that we will do utmost to ensure equal treatment when it comes to designation.

- However, it is not for fears of unequal treatment that we will give blank checks for not implementing the DSA. We adopted a first batch of 19 designation decisions which cover largest, systemic platforms.

- You will appreciate that not all cases are the same and some designation may be more complex than others. This may mean in certain cases we would need to engage in more in-depth investigations of reported user numbers, before being able to designate companies concerned.

- If there are other platforms whose numbers suggest they may be VLOPs, too, we will proceed with designation swiftly. In any event, such borderline cases do not reach the level of systemic risk of the clear-cut cases where we already proceeded with the designation.

- Companies' delaying tactics will not pay off. If we see that there has been bad faith and an attempt to escape our rules, this may indicate a lack of compliance with the DSA obligation to publish user numbers in a clear and accurate manner, and may be sanctioned.

Why has the Commission met with VLOPs/VLOSEs ahead of national elections?

- Protection of electoral processes is directly in the scope of the DSA via risk assessment provisions in Art. 34. Applicable to VLOPS such as YouTube, these are part of a systemic oversight cycle where the platforms need to (1) first assess the risks stemming from the design or functioning of their service and its related systems and then (2) propose and employ measures to mitigate those risks (Art. 35).The Commission is in a continuous exchange with the platforms, civil society, and governments to monitor the application of the DSA.

Is there a risk for these exchanges to interfere with national processes?

- The DSA aims to ensure that platforms – and in particular the very large ones such as YouTube – are more accountable and assume their responsibility for the actions they take and the systemic risks they pose, including on disinformation and manipulation of electoral processes. To this end, it is important that the Commission talks to all relevant stakeholders to establish facts and gather evidence, and to understand the specific situation in the Member States. These exchanges are part of the enforcement tasks that has been entrusted to the Commission by the Member States during the negotiation of the legal text. What type of content is allowed or not on the platforms is outside the remit of the Regulation and it is determined by national rules. The Commission's role is to ensure that platforms have sufficient resources and safeguards in place to walk the walk and comply with their risk management obligations.

Background

YouTube's actions concerning the Israel-Hamas conflict

YouTube's spokesperson has stated that the company's teams "have removed thousands of harmful videos", are "working around the clock to monitor for harmful footage and remain vigilant to take action quickly across YouTube, including videos, Shorts and livestreams"[1].

YouTube has been a major venue of advertisements placed by the state of Israel to raise awareness about the terrorist attacks the country suffered, and to denounce Hamas as a terrorist organisation similar to the Islamic State[2].

The Digital Services Act and the Israel/Hamas Conflict

Content circulating online that can be associated to Hamas qualifies as terrorist content, is illegal, and needs to be removed under and Terrorist Content Online Regulation – as the organisation is listed on the EU's terrorism list (list of persons, groups and entities involved in terrorist acts and subject to restrictive measures). In relation to content that cannot be clearly associated to Hamas, the rules apply where the content is inciting or glorifying terrorist offences.

In addition, the DSA requires effective action against illegal content by putting in place the necessary content moderation tools and by engaging in the risk management cycle. Also, on violent content that is not terrorist content, very large online platforms and search engines have a responsibility under the DSA to mitigate risks stemming from content that is illegal or harmful in a proportionate manner, and in consideration of the fundamental right to freedom of expression.

Commissioner Breton sent a letter to YouTube on 13 October to remind that the Digital Services Act sets very precise obligations regarding content moderation:

- First, platforms need to be very transparent and clear on what content is permitted under their terms and consistently and diligently enforce their own policies. This is particularly relevant when it comes to violent and terrorist content.

[1] https://edition.cnn.com/2023/10/13/tech/google-youtube-hamas-israel-disinformation/index.html

[2] https://www.businessinsider.com/israel-youtube-x-advertising-pr-offensive-gaza-hamas-bombardment-2023-10?op=1&r=US&IR=T

- Second, when receiving notices of illegal content in the EU, platforms must act timely, diligently and objectively in taking action and removing the content when warranted.
- Third, platforms need to have in place proportionate and effective mitigation measures to tackle the risks to public security and civic discourse stemming from disinformation.

YouTube's [redacted] replied to the letter on 18 October 2023.

The Commission will enforce the DSA and urges online platforms to diligently comply with EU rules.

YouTube's work under the Code of Practice on Disinformation

YouTube is an active Member of the Code of Practice Permanent Task-force, including the working group on elections.

Google provided the working group with insights on the preparation of its services – including YouTube – for the Slovak elections ahead of the voting

YouTube is also engaging with the process to devise a rapid response system and risk assessment framework for special situations such as elections or crises, but has expressed strong concerns and reluctance on documents that define any actions that signatories are to take in such a situation

YouTube's position is that any protocols developed under the Code – which may also become voluntary crisis protocols under the Digital Services Act – should outline the procedural steps that signatories will take during elections or crises, but not the detailed actions that each signatory will take. A representative of Google is currently preparing new discussion paper on a rapid response system for elections and crisis

Measures in the Code related to electoral processes

The Code's task-force has established a dedicated working stream on elections. The role of the group is to draft a rapid response system and risk assessment methodology for what it concerns electoral disinformation risks.

In addition, the Code also contains **specific commitments and deliverables related to elections:**

- to cooperate and coordinate their work during elections (M37.2)
- to establish a risk assessment methodology and rapid response system to be used during elections (M37.2) (in practice this means applying the Code's commitments in the context of elections as detailed below).

- to report within the regular monitoring on the actions related to elections (C42).
- Moreover, by a dedicated study Structural Indicators could give quantitative insights on disinformation spread on signatories' platforms during electoral periods.

Commitments and actions to be taken by VLOPs/VLOSEs signatories under the Code of Practice in the context of elections

Demonetisation:

- Ensure by targeted policies and their efficient implementation that the placement of advertising does not provide financial incentives for the dissemination of disinformation around the elections

Transparency of Political Advertising:

- Label in a clear and visible way all political ads;
- Make sure that the labels applied remain in place when shared by users on the same platform;
- Conduct research to understand if the labelling has been properly designed to help users to understand the nature of a political advertisement;
- Provide users with information about the political ads they see (identity of the sponsor, display period, ad spend…) and why they are seeing such an ad;
- Ensure that the identity of political sponsors is appropriately verified;
- Maintain and make public complete and updated repositories of political and issue ads.

Integrity of Services:

- Strengthen actions and teams to take effective measures against manipulative behaviour applied on their services (e.g. fake accounts, bot-driven amplification) related to the dissemination of disinformation around elections;
- Establish or strengthen policies regarding manipulative practices generated by AI around elections (e.g. proactively detecting and warning users about deep fakes, AI generated pictures and other manipulated media);

- Cooperate between signatories to counter cross-platform influence operations and prevent that malicious actors, banned by one platform migrate to another platform to continue their disinformation related activities.

User Empowerment:

- Put in place dedicated media literacy campaigns related to elections,)
- Provide users with tools to access authoritative sources (e.g., national electoral commissions) when it comes to electoral information (e.g., how and where to vote, information about the electoral system, process and official results).

Fact-checking:

- Step up and prioritise the integration of fact-checks related to elections into their services (e.g. by labelling, demotion etc.).
- Strengthen cooperation with fact-checkers during elections and focus fact-checking on narratives related to the integrity of elections, alleged voter fraud etc.;

Research:

- Provide adequate tools to allow research on disinformation campaigns during elections – i.e. swift access to specific data sets, relevant APIs, access to complete and updated repositories of political and issue based ads and APIs.
- Engage with the research community ahead of the elections to better understand and support their needs.

Platforms' actions on disinformation related to Russian war against Ukraine

- Since March 2022, regular meetings are taking place with the main platform signatories of the Code of Practice on Disinformation (Google, Meta, Twitter, TikTok, Microsoft).
- In general, since March 2022, online platforms are stepping up their efforts to de-prioritise and/or remove disinformation content related to the war.
- Social media are labelling content/accounts and/or down-ranking/de-amplifying Russian state affiliated media globally.

- Several platforms have taken action regarding Russian government affiliated accounts to limit their ability to share disinformation related to the war. (Twitter and Instagram use de-amplification, while Meta adopted a new policy to stop accounts to share content that denies the use of the force against other states.)
- Online platforms have increased cooperation with fact-checkers in and around the areas of the conflict and boosted the capacity of content moderation teams.
- All major platforms have demonetised players spreading disinformation related to the war and limited the possibility of monetising content related to the war. Moreover, several platforms have stopped their ads services in Russia (e.g. Meta and Google).

New AI tools and risks for disinformation

- The latest advances in **Generative AI, utilizing Foundation models**, have led to systems capable of producing innovative and original content, including images, text, and music. These models are developed by training a Foundation model on vast amounts of data and refining it for specific tasks, resulting in more precise and efficient outcomes. Generative AI's ability to create personalized and customized content is expected to **revolutionize many fields, such as the entertainment industry, journalism, and marketing**.
- Although most Generative AI models have focused on text generation, additional modalities beyond text presents a fascinating avenue for extending the capabilities of AI systems. By incorporating diverse modalities such as visual media, robotics, virtual worlds, and multi-modal experiences, innovative AI applications can emerge.
- However, as the marginal cost of generating content approaches zero, there is **a growing risk** that these models **may be used for malicious purposes**. Such uses include social engineering and **election manipulation campaigns** that exploit automated bots on social media platforms, as well as **the creation of fake news and web content**.

- When working with Generative AI models, one issue is their **propensity to 'hallucinate' and provide nonfactual or inaccurate information**. Furthermore, some models have exhibited inappropriate behaviour in interactions with users. Biased outputs present another concern, as these models can reflect biases in the training data. The use of racially and gender-biased language and other offensive content is especially problematic in real-world applications.

Latest AI products by Google (Alphabet) and YouTube

- Alphabet is very active in AI research and development across domains. In addition to Google and its AI division, several other subsidiaries of Alphabet also work on AI, for example DeepMind or the autonomous driving company Waymo.
- In 2017, Google published the paper "Attention Is All You Need", introducing the transformer model, which is at the heart of the latest wave of generative AI. Despite leading in research, Alphabet was caught off-guard by the success of ChatGPT used by 100 million users by January 2023. ChatGPT was developed by OpenAI, which is funded by Microsoft which runs Bing. Bing competes with Alphabet's Google online search business that generated €283 billion in 2022 for the group.
- More recently YouTube stared to use generative AI for a range of purposes, including AI for subtitle and voice-over generation and translation, video and audio editing capabilities (beat matching, voice changing) and AI-informed guidance to increase audience and reach. Those tools are in part provided by their parent company Google. Planned uses include AI for content moderation.
- In May 2023, Google announced PaLM 2 at the annual Google I/O keynote. The AI model, which outperforms other leading systems in some tasks, is expected to be incorporated into 25 new products and features. As a general-purpose AI, it can power chatbots, facilitate translation between languages, write computer code, and analyse and respond to images.
- In September 2023, Google announced the integration of its PaLM2-powered chat bot Bard (their competitor to ChatGPT, based on generative AI) into YouTube.

CV of [redacted] of Youtube

Born [redacted] United States

Education:

- [redacted]

Professional Career:

- [redacted]