EUROPEAN DEMOCRACY ACTION PLAN
GUIDANCE & INDICATORS FOR THE CODE OF PRACTICE 2.0

INITIAL CONSIDERATIONS

This document outlines ACT's initial views regarding the Digital Services Act (DSA) Regulatory backstop and the six objectives outlined under point 4.2 of the European Democracy Action Plan (EDAP) to inform policy-makers thinking in relation to the review of the Code of Practice on Disinformation (CoP). From the outset, ACT would like to stress that the review of the CoP as an information gathering exercise to inform the Commission's future work on how to ensure comprehensive, legally binding, obligations and accountability of online platforms through regulation.

Disinformation, certainly harmful but not necessarily illegal, has been identified by international, EU and national bodies as one of the most detrimental externalities of online uploaded content, notably for the sustainability of democratic discourse and in the context of the current public health crisis. It is essential to ensure that business models do not take advantage of disinformation, particularly when it is done at the expense of civil and democratic discourse, undermining reliable sources of information in the process.

At present, the intensive algorithmic data processing underlying the delivery of messages on online platforms is often opaque[1]. When such activity involves disinformation this can have severe impacts on the democratic process. To prevent this we welcome the Commission's call for a stronger enforcement of GDPR in respect of online platforms. In particular, Article 5(1)(b) GDPR which states that personal data shall be collected for specified, explicit and legitimate purposes should be more stringently applied.

Commercial broadcasters are professional media companies that focus on quality news and reliability. This underpins viewer trust and with it loyalty. Viewer trust is a core driver of our activities and a key asset. Each broadcaster has rules governing what it publishes and meets the responsibilities expected of them according to the laws in place that are there to ensure a functioning democracy.

ACT, along with other members of the Media, Academic, Research and Consumer communities, has long called for ambitious proposals. We are therefore pleased by the EDAP and its aim to significantly reinforce the Code of Practice on Disinformation (CoP) as a first step towards a binding instrument.

ACT supports a regulatory approach. This echoes the learnings from the short-comings of the self-assessment reports by signatories, as well as improvements suggested in independent reviews from researchers to regulators across the EU. As the European Regulators Group for Audiovisual Media Services (ERGA) stressed in its report on the implementation of the Code of Practice on Disinformation, the Code has "significant weaknesses" which "justify a shift from the current flexible self-regulatory

1 https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_202008_onthetargetingofsocialmediausers_en.pdf

approach to a co-regulatory one". This paper outlines ACT's views on the necessary transparency and oversight mechanisms, including key performance indicators, to carry out this shift.

TOWARDS A COP 2.0 - INITIAL ASSESSMENT FOR OBJECTIVES TO BE MET

The DSA backstop and the CoP go hand in hand. The DSA backstop is welcome. Yet alterations need to be taken onboard if it is to serve its purpose and prevent sudden and massive systemic shocks induced by disinformation campaigns. We argue that the CoP dialogues is and should be an opportunity to discuss the core attributes of (co-)regulation rather than its merits, which no longer need justification. The regulatory backstop needs strong CoP outcomes – indicators & oversight – to ensure effectiveness. The present EDAP discussion is an opportunity to set required measurements, transparency and access obligations and formats, oversight metrics and audit requirements, financial flows and sanctions / fines.

Ensure multi-stakeholder across the piece to guarantee openness and legitimacy of outcomes. All parties involved have an important input in this discussion and are directly impacted. Therefore all dialogues, whether this be on monetisation or research APIs, should be open to those taking part. This also includes thinking about the scope of expertise involved. The link to the DSA backstop and audits suggests that the dialogues would benefit from the participation of data analyst and auditors.

Smart measurement & real transparency. Comprehensive, measurable and verifiable Key Performance Indicators (KPIs) are essential. A major shortcoming of the CoP is the lack of verifiability of the data shared by co-signatories. This has been compounded by an erosion of trust with social networks and other predominant platforms. For this we call on securing strong algorithmic ex ante transparency requirements as regards KPIs that help all parties affected address, prevent and respond to systemic threats. More precisely, moderation and recommendation algorithms deployed by signatories of the CoP should meet the highest level of transparency to limit the spread of disinformation. There is no reason for public security and health to not be afforded the same protections as anti-trust measures seeking to keep markets fair and contestable. For the CoP 2.0 to deliver actual results, platforms will need to make real commitments with indicators and commitments that pave the way for a true enforcement framework. This transparency may remain elusive and as such the Guidance could benefit from data that better reflects impact and helps appreciate the commitment level of the signatory. As such, more precise indicators but also indicators that reflect wider commitment (eg financial/staffing) and progress consistently across time and key dimensions would be welcome.

A broad approach. The revised CoP should not adopt an overly strict approach and define the problem narrowly. It should cover both intentional coordinated disinformation and misinformation. Similarly, political and issue-based advertising need to be addressed on a similar level in the Code.

Aligned with EU co-regulation principles. The CoP should be fully aligned with the EU's the "Principles for better self- and co-regulation"[2], in particular the principles of openness, good faith, monitoring and evaluation.

[2] https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/CoP%20-%20Principles%20for%20better%20self-%20and%20co-regulation.pdf

Spelling out the standards and routinely review/assess. As researchers and regulators need a common and standard way to access and interpret data, key standards are going to have to be set. This can be reviewed periodically but the goal should be to ensure that researchers can depend on a list of access and data transfer that is reliable and backed up by the EU.

Independent research should be simplified/encouraged/protected. The proposal would benefit in supporting researchers wishing to publish data without fear of reprisal by platforms. A strong commitment by signatories in this field is required as a show of good faith.

Key KPIs and commitments:

Objective 1: Monitoring

Commitment:

- An effective **transparency and monitoring framework** allowing to assess the scope of the progress
- **Involvement of regulators** in the monitoring and enforcement of the CoP via ERGA
- **Independent** audits and monitoring

KPIs:

- Disclosure of figures on revenues tied to disinformation vs. expenditures
- Granular information on actions taken by online platforms with regards to posts, accounts and websites spreading dis/misinformation (suspension, demotion, demonetisation)

Objective 2: Visibility of reliable information and ranking

Commitments

- **Transparency of moderation and recommendation algorithms:** extension and reinforcement of DSA principles
- **Measures to boost reliable content** to rely on a non-discriminatory basis commitment for reliable media of which licensed media should be one of the criteria – basis of existing regulation (AVMSD) and editorial responsibility as a primary axis.
- No measures that leading to a **double screening** beyond editorial control and/or compliance with license rules (eg. AVMSD)

KPIS:

- Figures on investments in technological means to prioritise and demote information
- Information on actual uptake of tools that help consumers understand why they are seeing particular advertisements
- Number of complaints submitted by users about disinformation content and ratio of response by platforms

Objective 3: Monetization, political and issue-based ads

Commitments:

- **Adequate advertising transparency, accurate repositories and measures ensuring traceability**. Extension of DSA principles on advertising transparency to allow researchers, regulators and policy-makers to have a view on flows and their evolution.
- Ambitious commitments on the **demonetisation and refunding** of ads tied to disinformation
- Revenues platforms made from **political and issue-based ads** overall and share of that revenue attributed to content taken down

KPIs:

- Revenues from legitimate advertising sources displayed next to content then identified as disinformation / misinformation and impressions showed to accounts identified as bots / fake accounts
- Amounts refunded due to ads tied to disinformation, including both ads appearing next to disinformation and ad impressions to inauthentic accounts
- Figures on platform reaction times following advertiser reports / complaints for their ads appearing next to content covered by the code
- Revenues made from political and issue-based ads overall and share of that revenue attributed to content taken down

Objective 4: Fact-checking

Commitments:

- Real cooperation and investments with the **fact-checking community**
- **Empower fact-checkers** to do their work and act upon their notification

KPIs:

- Amounts allocated to fact-checking and to supporting the fact-checking community
- Figures on direct and retroactive labelling of fact-checked content

Objective 5: Integrity of services

Commitments

- Measures to tackle artificial and endemic **amplification and virality** on online platforms
- Clear, effective, enforced and transparent **policies** on the use of bots, fake-accounts and proxy services

KPIs:

- Investments (financial and staff) directly devoted to the identification of fake accounts and bots
- Figures on the number of bots and fake account identified, actions undertaken, information on their reach (e.g. posts, likes, comments, shares) before actions and use of technical tools like proxies

Objective 6: Access to data

Commitments:

- An effective data access framework for researchers, fact-checkers and regulators

KPIs:

- Figures on the use by researchers, fact-checkers and regulators of the data access framework
- Ratio of number of total request to accepted researcher requests received by signatories
- Figures on donations made to NGOs, civil society, and others to support initiatives combating disinformation (eg. media literacy campaigns) against total turnover

REFERENCES

Relevant ACT stand-alone and joint statements

- Broadcasters call for Media/Democracy Action Plans to ensure investment in content, media pluralism & trustworthy news
- Commercial Broadcasters welcome Council Conclusions support for a level playing field, territorial licensing and greater responsibility of Platforms
- ACT Perspectives on the Digital Services Act
- ACT Feedback on Roadmap on European Democracy Action Plan
- ACT Response to the European Commission's Inception Impact Assessment on political advertising
- Sounding Board Members react to EC Communication on disinformation and call for stronger measures in light of "infodemic"
- ACT welcomes European Commission's Communication on Shaping Europe's digital future

Resources consulted on KPIs

- Sounding Board recommendations on KPIs (30/8/2018 - *Feedback from Sounding Board on the latest Draft Code of Practice presented by the Working Group*) made to the European Commission in the framework of the discussions leading to the set-up of the CoP.
- Report from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the implementation of the Communication "Tackling online disinformation: a European Approach"
- "Principles for Co-Regulation of Systemic Online Platforms" by Trevor Davis, Deputy Director (Research), George Washington University
- ERGA Report on disinformation: Assessment of the implementation of the Code of Practice
- Study for the "Assessment of the implementation of the Code of Practice on Disinformation" Final Report by VVA and EU Disinfo Lab

We would particularly like to thank Professor Trevor Davis for his invaluable contribution outlined in "Principles for Co-Regulation of Systemic Online Platforms".

CODE OF PRACTICE: DSA BACKSTOP LEAVES GAPS TO ADDRESS IN CoP PROCESS

DSA BACKSTOP

The DSA will establish a co-regulatory backstop for the measures which would be included in a revised and strengthened Code of practice on disinformation.

Tackling potential gaps in the DSA's co-regulatory backstop approach

As outlined in ACT's response to the European Democracy Action Plan (EDAP), the only way to move forward on tackling disinformation is to establish a co-regulatory framework backed by a regulatory framework on content moderation, data access and transparency with an adequate oversight mechanism.

The co-regulatory backstop in the DSA starts an important discussion on establishing a regulatory framework for content moderation, data access and transparency. While the principles are adequately set out, the question is whether the backstop can respond to a real stress test – as we have witnessed with the US Capitol and/or pandemic situations. The CoP cannot address all elements, but Guidance can be supportive and prepare the ground for strong incentives to drive compliance.

The key lesson from the CoP is how to focus on indicators that incentivise compliance. With good metrics that shed more light on progress, and effective measures tackle reach and spread. This means KPIs, tools, funds and manpower for researchers and regulators, investments by platforms, tools and processes for fact-checkers, mechanisms to avoid limitations on/boost reliable news sources.

Further key areas of the DSA may change in the course of the legislative process. Some identified limitations:

- **Timing**. The DSA will come into play by 2023/4 by some estimates. The CoP will effectively be the backstop until then for the hundreds of elections expected to take place in the interim. Just in 2020 there were +30 local, regional elections and referendums according to the Council of Europe database.
- **Systemic risks**. The ability to be responsive to threats, starting with the definitions to be clarified ("foreseeable harms"), potentially lengthy procedures, limitation to algorithmic transparency are concern as regards the applicability and responsiveness of the envisioned regime to tackle (sometimes) immediate threats.
- **Robust horizontal rules.** Transparency of content moderation and recommendation algorithms, advertising and access to data for researchers and regulators are all an essential part of establishing a robust co-regulatory backstop but need to be reinforced across the board.
- **Effective (co-)regulatory regime**. Guidance on mitigation measures for adequate/coordinated approach, starting with relevant transparency KPIs, investment criteria, cross-platform reporting to users and sanctions regime to ensure compliance are complementary tools.
- **Incorporating the DSA backstop into the CoP**. As outlined above, the DSA will not come into force for a number of years. In the meantime, the CoP will have to do. It is therefore absolutely essential that the Commission incorporates the relevant elements of the DSA Proposal (eg. around recommendation tools and content moderation, transparency and access to data) in its guidance and support their application through co-regulation until the DSA takes over.

CODE OF PRACTICE: OBJECTIVES & IMPLICATIONS FOR KPIs & GUIDANCE

OBJECTIVE 1

Monitor the impact of disinformation and the effectiveness of platforms' policies, on the basis of a new methodological framework which includes principles for defining Key Performance Indicators (KPIs). In this context, timely information on platforms' policies and access to relevant data, needs to be available to allow, amongst others, measuring progress against the KPIs;

Perspective > Impact for CoP framing & topics to address

Effectiveness remains an open question without mechanisms for verification and oversight. Transparency is key. As part of DSA backstop, the verification of KPIs is problematic without regulator access to data and/or where platforms oppose business secret.

As this transparency may remain elusive, a rounded approach is required to complement the picture with KPIs that are wider in scope and information to users that is meaningful. These KPIs go hand in hand with timely information to measure progress but ideally to ensure immediate threats are notified, tracked and acted upon. This in turn requires input from audit and data experts to assess some of the core technical parts of the KPIs proposed.

ACT supports a regulatory approach. We therefore believe that regulators should be involved in the monitoring and enforcement of the code. The European Regulators Group for Audiovisual Media Services (ERGA) and its members at the national level, given their expertise on the matter and the fact that they are increasingly responsible for these issues at national level, should be given formal tasks and powers in this process.

KPIs will need to be cross-platform to be able to measure consistently the scale of the problem and the impact of actions on the whole online ecosystem, but also to make the CoP future-proof and facilitate on-boarding of future signatories.

KPIs in the spotlight – Transparency on suspended accounts (more details in Annex I on integrity of service / fake accounts, bots and account suspensions)

- Number and type of promoted pages, channels or posts related suspended
- Breakdown by Member State of suspended accounts
- Number of suspended accounts reported to law enforcement
- Average duration for which the suspended accounts operated
- Numbers on the total number of posts/tweets/videos/comments/shares suspended accounts
- Numbers on the total number of engagements, content created or shared by the suspended account
- Numbers on the total number of impressions that content created by the suspended account received
- Numbers on the use of proxies by the suspended accounts
- Numbers on the nature of suspension: Permanent, conditional or temporary

- Number of accounts removed for violations of platform advertising policies (eg. policies against misrepresentation)
- Numbers and type of tools available per Member State, number of full time employees whose primary focus is the issue; external contractors should be clearly identified as such
- Direct investment spent on identifying/deactivating disinformation content as a percentage of turnover and per user as a percentage of total turnover per user
- Total of foregone (lost) revenue due to certain accounts being closed due to purveyors of disinformation

Additional consideration for the CoP discussions on Guidance:

- **A scoreboard** as part of the deliverable of the methodological framework, as the evaluations in themselves are hard to read, summarise and show evolution
- **Access to data and reporting criteria that are useable** and create reliable indicators for users within and across platforms – specific reporting standards and standards/incentives for timely access
- **KPIs included in methodological framework to better assess commitment level of signatory –** eg direct staffing to tackle issue and political based ads, evolution, weight in proportion to user base
- **Advertising revenue money flow** tied to disinformation – money flow analysis to inform view on incentive structure, approach to fines and potential reinvestment
- **Direct – in country – access for regulators**. There is no reasons for platforms to use business secret to prevent regulators from accessing directly platforms' systems. On-site visits in dedicated centers are only transparency decoys.
- **Controls and thresholds triggering notification to regulators** – virality of content and rapid reaction mechanisms
- **Dialogues on Guidance to include data and privacy auditors to give insight on KPIs –** the DSA audits and the criteria used for systemic risk covering these aspects. View from independent auditors with experience in data audits should be included in the discussion to fact check data use, verifiability, aggregation and cross platform.

OBJECTIVE 2

Support adequate visibility of reliable information of public interest and maintain a plurality of views: by developing accountability standards (co-created benchmarks) for recommender and content ranking systems and providing users with access to indicators of the trustworthiness of sources;

Perspective > Impact for CoP framing & topics to address

Plurality of views is key to ensure the approach to public interest is not interpreted narrowly and undermines support from the wider media ecosystem. Measures that would lead to a double screening beyond editorial control and/or compliance with license rules would not meet the criteria of support.

Boost content that is produced by reliable sources and ensure this mechanism is not used to limit access to legitimate services. In the case of licensed and regulated broadcasters this is paramount, as this would

make the regulatory asymmetry with these platforms even greater. The support of the media community rests on the ability for the guidance to incentivise the platforms to provide transparency and the tools and commitments needed for them to be effective and accountable. But not in a manner that puts legitimate AV, print and other sectors content with an additional hurdle to access consumers. The aim is to avoid community standards that apply public interest criteria to content that is already regulated under other provisions, notably AVMSD.

The Code should include commitments implementing ahead of time and reinforcing obligations related to the transparency of recommender systems foreseen in the Digital Services Act.

KPIs in the spotlight – Data indicators on recommender systems and user empowerment

- Information on investments in technological means to prioritise relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels
- Information on investments in features and tools that make it easier for people to find diverse perspectives about topics of public interest or other progress towards this commitment
- Information related to partnerships with civil society, governments, educational institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy
- Information on actual uptake of tools that help consumers understand why they are seeing particular advertisements
- Number of complaints submitted by users about disinformation content and ratio of response by platforms
- Number of sites (eg. click-bait) and posts demoted to measure the effectiveness of policies put in place to demote sites or accounts that distribute disinformation or inauthentic information
- Number of demonetized content (eg. posts) to demonstrate the effectiveness of policies and processes aimed at disrupting advertising and monetization incentives for relevant disingenuous behaviors

Additional consideration for the CoP discussions on Guidance:

- **Measures to boost reliable content** to rely on a non-discriminatory basis commitment for reliable media of which licensed media should be one of the criteria – basis of existing regulation (AVMSD) and editorial responsibility as a primary axis.
- **Tie in with AVMSD measures on video sharing platforms** – How will Guidance and AVMSD provisions on VSPs articulate given some pieces of content may be covered at the same time by AVMSD application and measures aimed at tackling harmful content?
- **Adequacy of visibility or prominence decisions to be left at Member State level with guidance to focus.** Priority access to, and prominence on, platforms raises issues of definition and of equal treatment. A more equitable, transparent and future-proof approach would be to leave this to the discretion of Member States since the regimes for the delivery of public value, for broadcasters by way of example, vary significantly across the EU according to cultural and historical traditions.

OBJECTIVE 3

Reduce the monetization of disinformation linked to sponsored content: in cooperation with advertisers, limiting the false or misleading issue-based advertisement on online platforms or on third-party websites as well as the placement of ads on websites that are purveyors of disinformation;

Perspective > Impact for CoP framing & topics to address

Monetisation is a key incentive driver. The EC rightly outlines the need to act on two fronts with regards to false or misleading ads placement or legitimate ads on sites geared to purvey disinformation.

Reducing flows is important, but redirecting and identifying amounts also presents regulators and policy-makers with input on status, evolution and guidance. Putting a cost on disinformation is the next step to driving change. Accounting for and redirecting flows to tackling disinformation and promoting trusted sources ties in with the overall objective of identifying the cash volumes generated by purveyors of disinformation. This gives policy-makers and regulators reliable oversight and numbers to identify, track and gauge sanctions.

It is essential to create market incentives for platforms to tackle the spread of disinformation. To do so, we must ensure that there is near perfect information on the advertising market. Participants must clearly understand where advertisements are placed and for what reasons. This commitment was already spelt out in the previous code, yet the problem subsists. The new commitments need to be a step up and deliver on tools within and across platforms to ensure standards emerge and complaints can be tracked and independently reported to ensure actual progress. This would ensure bad activity can be rapidly identified but also allow recourse for websites/accounts that may be unduly flagged.

This information should be available to all market participants so as to ensure they can also understand how disinformation affects the online advertising markets as a whole. The ultimate decision of where their ads appear should rest with the advertiser, but to make this decision they need appropriate transparency tools based on data he can trust and easily understand at a granular level whilst market participants will need information at an aggregated level to understand how this is affecting market dynamics.

The commitments should also ensure advertisers are able to benefit from the work done of the disinformation community (fact checkers, academics, regulators and others) and the flagging of harmful websites or accounts on which their ads land rather than rely solely on the assessment provided by the Platform.

Finally, the code needs broad and ambitious commitments covering both political and issue-based advertising, regardless of the entities paying for them. Transparency and labelling have a key role to play here. Given the fact that the latter are not disclosed, issue-based ads often contribute more to the outcome of elections as actual political advertising placed by political parties or candidates. As such, they have a comparable or larger impact on the democratic discourse and should therefore be treated in the same way as traditional political advertising. Issue-based ads can not only be dishonest and misleading, but they can also be very harmful, negatively impacting democratic processes. Focusing commitments purely on political advertising, by official political parties, would not only be ineffective, it would also give a false sense of progress.

However, we do not see the Code of Practice as a replacement for the upcoming regulation on political advertising. Moreover, we welcome that the European Commission has launched a public consultation on transparency in political advertising, as a follow-up to the commitments made in the European Democracy Action Plan and in order to prepare a legislative proposal on the transparency of sponsored political content as well as support measures and guidance. This initiative should address all actors involved in financing, preparing, placing and disseminating of political advertising and complement the rules set out in the proposed Digital Services Act.

KPIs in the spotlight – Demonetisation and Political & Issue-based advertising – placement and removal

- Total advertising revenues from legitimate advertising sources displayed next to content then identified as disinformation / misinformation and impressions showed to accounts identified as bots / fake accounts
- Total amounts refunded to advertisers due to ads being tied to disinformation / misinformation, this should include both ads appearing next to disinformation / misinformation and ad impressions to accounts identified as bots / fake accounts
- Figures on platform reaction times following advertiser reports / complaints for their ads appearing next to content covered by the code
- Percentage of contracts between advertisers and ad network operators with brand safety stipulations against placement of ads on disinformation websites
- Revenues platforms made from political and issue-based ads overall and share of that revenue attributed to content taken down
- Number of websites blocked for duplicating or "scraping" content produced by other websites
- Total and average aggregated spend/revenue by misleading advertising/website purveyors
- Total advertising revenue from top websites identified as purveyors of disinformation
- Number of political or issue-based ads taken down
- Information on amounts received from individuals, companies, political parties, candidates, campaigns and foundations for political or issue-based advertising
- Report on proper labelling of all political and issue-based ads as readily recognisable as paid-for communication across platforms, incl. number of mislabelled and relabelled ads

Additional consideration for the CoP discussions on Guidance:

- **Cooperation with advertisers should be extended to media, civil society, fact-checkers and academia which play a part in this discussion**. Follow the multi-stakeholder approach which is spelt out in the Communication as concerns the approach to sensitive issues.
- **Full transparency on content moderation and recommendation policies**, in particular for issue-based advertising and articulation with trusted sources. Principles for algorithmic transparency drivers.
- **Adequate advertising transparency, accurate repositories and KPI targets on traceability –** extension of DSA principles on advertising transparency to allow researchers, regulators and policy-makers to have a view on flows and their evolution. This includes information on the person

on whose behalf advertising us placed, the period during which it was displayed, the criterion used to target one or more particular groups and the total advertisement reached.
- Existence of policies to verify the identity of political or issue ads providers and publicly disclose them
- Existence of tools to enable users to understand why they have been targeted by a given advertisement
- Existence of registries and processes to disclose on "issue-based advertising", including quarterly reporting on their use and measures / processes in place to encourage their usage by consumers
- **Auditing capacities to ensure these criteria can be fully integrated in the DSA exercise**

OBJECTIVE 4

Step up fact-checking, by establishing transparent standards and procedures for open and non-discriminatory collaboration between fact-checkers and platforms and foster cooperation;

Perspective > Impact for CoP framing & topics to address

This is an important step to ensure the gap in fact checking does not continue to widen. We note that is partly due to the growth in volumes of disinformation and the difficulty for fact-checkers in their cooperation with platforms on this issue which has been documented.

Targeted and retroactive fact-checking labels to those who have uploaded, viewed, shared or otherwise engaged with disinformation content is one way to tackle disinformation and it should be included in the code.

We would also consider that – as the Communication rightly identifies – Platforms face a responsibility to act, and this should be seen through the eye of their own investments in technology and people to fact-check. However, the mere fact of investing in fact-checking should not shield platforms from their responsibility. An obligation of result, not of means, should be the rule. At the very least, the measures should contain some indicators that give a reasonable assessment as to the direct investments that the platforms do, whether such effort is outsourced, or internalised, and how this amounts to numbers on actual people reviewing the content.

How fact-checkers agree and engage with platforms is also important. It should however not obscure the needs of data requests by fact-checkers. Or the legitimate use of publicly obtainable information towards fact-checking studies on tracking the source and spread of online disinformation. How these interactions are managed and the commitments for the code will be determinant. Signatories of the code give fact-checkers access to a larger toolbox to be able to detect and analyse disinformation. Such a toolbox should take the following elements into account:

- EXIF data - When pictures or videos are uploaded on online platforms the Exchangeable Image File (EXIF) data is deleted or hidden and not accessible anymore. However, this EXIF data is essential for fact-checkers as it facilitates access to information like recording location.
- Access to the sender - Users of online platforms are not obliged to provide a masthead so accounts with millions of viewers cannot be accessed without proper contact information.

- Access to data - Only online platforms have information about who has seen which content, how content (and disinformation) has spread and who was reached but this data would be useful for fact-checkers. They should therefore also benefit from an empowering data access framework.

Data access criteria in the spotlight – Commitments, policies & standards on data sharing (more details in Annex II on a data access framework for scholars and public authorities)

- The minimal standard for data delivery shall be direct download
- Summaries and systems which require interaction with graphical user interfaces are not acceptable substitutes
- Under no circumstances should a compression format that requires a commercial license be used
- Values must be typed consistently
- Primitive data types native to C, C++, Java or ECMAScript are acceptable
- Encoding shall be consistent and according to UTF and UCS standards
- Complete documentation for fields, objects shall adhere to an open (non-commercial) standard for machine and human comprehension
- An API shall not be a substitute for full, downloadable archives. APIs must adhere to an open (non-commercial) standard for machine and human comprehension

Additional consideration for the CoP discussions on Guidance:

- Reporting indicators for CoP signatory investments in internal/external fact-checkers
- Reporting indicators for investments in technical driven solutions (AI, APIs, Take down)
- Guidance of commitments for platforms in relation with fact-checkers with whom they cooperate and for those seeking data access for legitimate purposes
- Common and adapted guidance as regards standards for data access, compilation and analysis
- Figures on direct and retroactive labelling of fact-checked content

OBJECTIVE 5

Strengthen the integrity of services offered by online platforms by developing appropriate measures to limit the artificial amplification of disinformation campaigns;

Perspective > Impact for CoP framing & topics to address

Rules for recommendation tools and content moderation in relation to artificial amplification should be a key element of the new Code of Practice.

Getting a better assessment of fake accounts, bots and other means used to amplify disinformation through metrics of the situation. Another important element related to coordinated behavior is the use of proxy services[3], used by malicious actors to coordinate numerous accounts from across the world. If this is a mainstream tool used to develop coordinated behavior the commitments could narrow in on such tools and at the very least should include indicators on the extent of their use.

[3] For instance services such as https://smartproxy.com/proxies/facebook-proxies and https://luminati.io/

However, the Commission should be careful not to focus only on artificial, or rather intentional / inauthentic, amplification. Amplification and virality on online platforms are issues in themselves. It is well known that controversial content creates more user interactions, which platforms reward in higher ranking and visibility. Algorithmic transparency and proper oversight are an essential part of the answer.

KPIs in the spotlight – Commitments, policies & standards on integrity of their services (more details in annex I on integrity of service / fake accounts, bots and account suspensions)

- Number of bots disabled for malicious activities violating the platforms' policies
- Number of posts, images, videos or comments acted against for violation of platform policies on the misuse of automated bots
- Figures on proxies, the extent of their use on the platform in general and specifically in relation of posts, images, videos or comments tied to coordinated behaviour / disinformation
- Average engagement touch points and recurrence (e.g. posts, likes, comments, shares) between inauthentic accounts/users and genuine users before being detected
- Information on policies about the use of proxies
- Information on policies about the misuse of bots, including information about such bot-driven interactions
- Number of staff directly devoted to the identification of fake accounts and bots, including time and monetary investment made by Platforms to address this issue, including investment in third-party fact checking organisations

Additional consideration for the CoP discussions on Guidance:

- **Analyse flows to determine benchmark and reporting KPIs that serve as early alert warnings/identifiers** to regulators, fact checkers, academics and fellow signatories – with common criteria – on identifying purveyors, or specific flagged content, evolution in and across platforms.
- Election rules that create a level playing field

OBJECTIVE 6

Ensure an effective data disclosure for research on disinformation, by developing a framework in line with applicable regulatory requirements and based on the involvement of all relevant stakeholders (and independent from political influence). The European Digital Media Observatory (EDMO) can facilitate the development of such a framework. The Commission notes that the GDPR does not a priori and across the board prohibit the sharing of personal data by platforms with researchers.

Perspective > Impact for CoP framing & topics to address

As noted previously, the issue of transparency is key given the need for verifiable progress and meaningful benchmarks. Data disclosure is a big part of this and requires a more streamlined approach to ensure

better outcomes. Making available data to scholars and public authorities to allow them to verify the veracity of platforms' reports and allow them to meaningfully investigate the spread of disinformation online.

Data availability in the spotlight – principles for developing a data access framework (more details in Annex II on a data access framework for scholars and public authorities)

- Data access for researchers limited to public content - content that is or was available to all users of a particular platform and not restricted to a group or individual by a user setting
- Data available for download or API consumption on computers and systems external to the platform. Researchers retain ability to publish conclusions with supporting evidence.
- Data must include original identifying fields from the schematic of the provider. This includes non-obfuscated UUID.

Additional consideration for the CoP discussions on Guidance:

- Ratio of number of academics / research organisations that enter into relevant arrangements with / are able to access APIs / download the data they seek from platforms against number of requests received
- Policies ensuring not to prohibit or discourage good faith research into Disinformation and political advertising on their platforms (eg. no non-disclosure agreements)
- Policies to encourage research into disinformation and political advertising
- Ratio of number of total request to accepted Researcher requests received by signatories
- Ration of total donations made to NGOs, civil society, and others to support initiatives combating disinformation (eg. media literacy campaigns) against total turnover

ANNEX I - INTEGRITY OF SERVICE / FAKE ACCOUNTS, BOTS AND ACCOUNT SUSPENSIONS

Platforms should report on activities carried out to ensure the integrity on their services on the following:

- Numbers demonstrating progress on labelling, identification and closure / disabling of fake accounts and bots
- Number of posts, images, videos or comments acted against for violation of platform policies on the misuse of automated bots
- Number of bots disabled for malicious activities violating the platforms' policies
- A breakdown of actions/account suspensions undertaken by platform on the following:
  - Number of suspended accounts that used deceptive automation
  - Number of promoted page, channel or post related to elections or matter of political controversy
  - Number of suspended accounts which self-reported location inside the European Union but operated elsewhere
  - Breakdown by country of the location of suspended accounts
  - Number of suspended accounts engaged in activities designed to compromise the security of the platform
  - Numbers of suspended accounts using proxies
  - Number of suspended accounts that impersonated a public figure
  - Number of suspended accounts identified as being operated by or in association with a nation-state
  - Number of suspended accounts reported to law enforcement
  - Number of suspended accounts counted in platforms statements to shareholders, advertisers or regulators in written estimates of Monthly Active
  - Average duration for which the suspended accounts operated. This should include a breakdown (eg. X% suspended in under 24 hour since the creation of the account, X% suspended after one week, one month, over a year)
  - Breakdown in percentage of channels by which users were identified as violative
  - Evolution over time of suspensions.
  - Numbers on the total number of posts/tweets/videos/comments/shares suspended accounts
  - Numbers on the total number of engagements content created or shared by the suspended account
  - Numbers on the total number of impressions that content created by the suspended account received
  - Numbers on the monetisation of the suspended accounts
  - Numbers on the nature of suspension: Permanent, conditional (for example, on removing a particular piece of content or verifying identity), temporary (limited for a duration of time)
- Ratio of all engagement (e.g. posts, likes, comments, shares) inauthentic accounts/users have had with genuine users before being detected and deactivated due to a breach of platform policies against all engagement inauthentic accounts/users have had with other inauthentic accounts/users before being detected and deactivated due to a breach of platform policies
- Information on policies about the misuse of bots, including information about such bot-driven interactions
- Number of staff directly devoted to the identification of fake accounts and bots, including time and monetary investment made by Platforms to address this issue, including investment in third-party fact checking organisations

- Investment in staff and resources to ensure the monitoring of the "policies on what constitutes impermissible use of automated systems"

ANNEX II – COMMITTING TO A DATA ACCESS FRAMEWORK FOR SCHOLARS AND PUBLIC AUTHORITIES

Excerpts from "Principles for Co-Regulation of Systemic Online Platforms" by Trevor Davis, Deputy Director (Research), George Washington University

Data Access for Scholars and Public Authorities

In addition to their reporting obligations, platforms must make available data to scholars and public authorities to allow them to verify the veracity of platforms' reports and allow them to meaningfully investigate the spread of disinformation online. The following quotes detail how to establish a data access framework.

"Data access for researchers shall be limited to public content - content that is or was available to all users of a particular platform and not restricted to a group or individual by a user setting.

As all systemic platforms currently provide commercial partners with public data in some form, as described in well documented APIs. Marginal impact on individual security and privacy is negligible.

Under no circumstances should a researcher be required to sign a Non-Disclosure

Agreement that impedes or could impede their ability to publish their conclusions with supporting evidence.

Thus the data must be available for download or API consumption on computers and systems external to the platform.

Data must include original identifying fields from the schematic of the provider. This includes non-obfuscated UUID."

Type of data platforms should make available to scholars and public authorities

"Systemic platforms must report on make available content, engagement metrics and historical impression counts from public-facing accounts (both officially sanctioned and "fan" pages/accounts, such as pages like "Bikers for Afd 2) that:

1. Represent itself as acting in support or opposition to any European political party, politician or causes of current controversy. (eg. Climate Change.)
2. Post on or about political parties, politicians or causes of current controversy.

Content should be included on the basis of the account/channel/page that produced it, not the nature of the specific video/post/tweet. This data must be available even if the content itself has been removed by the platform or the administrator. Redactions and obfuscation should only be employed if the content is related to a specific and ongoing legal proceeding or other legal obligation. This should be noted in the record to the extent permitted by the court."

"Platforms shall provide the following data [to researchers and public authorities]:

In order to reduce the burden on the platforms and to guard against the resolution of actioned accounts to individual identities, certain data may be produced in obfuscated form. However, unique identifiers shall be consistently hashed rather than uniquely hashed. In other words, it should be possible to see that a particular user was suspended and then un-suspended without being able to determine their identity.

All accounts suspended by the platform with UUID or hashed UUID with the following data:

1) The reason the user, page, group, channel or account were suspended.
    a) This shall reference the platform policy violation that was the basis for the suspension.
2) Additional flags if the account met any of the following criteria:
    a) Used deceptive automation - for example, operated via a headless browser.
    b) Promoted any page, channel or post related to elections or matter of political controversy.
    c) Self-reported location inside the European Union but was operated elsewhere. (Include locations.)
    d) Engaged in an activity designed to compromise the security of the platform.
    e) Impersonated a public figure.
    f) Identified as being operated by or in association with a nation-state.
    g) Were reported to law enforcement.
    h) Were counted in platforms statements to shareholders, advertisers or regulators in written estimates of Monthly Active
    i) Users or Daily Active Users.
        i) Complete list of months in which the user was included in those calculations.
3) The channel by which the user was identified as violative.
    a) Eg. User report, third party moderator (subcontracted service), platform employee, algorithmic/automated detection, law enforcement request or court order.
4) The timestamp of the suspension.
5) The total months' user was active on the platform. If the user was active less than one month but more than one day, the total days the user was active on the platform.
6) The total number of posts/tweets/videos/comments/shares the user produced.
7) The total number of engagements (see platform-specific notes) content created or shared by the user received.
8) The total number of impressions that content created by the user received.
9) If the user directly monetised their activity via the platform, how much money was paid to this user to the nearest thousand?
10) Nature of suspension: Permanent, conditional (for example, on removing a particular piece of content or verifying identity), temporary (limited for a duration of time).

Redactions and obfuscation should only be employed if the content is related to a specific and ongoing legal proceeding or other legal obligation. This should be noted in the record to the extent permitted by

the court. For example, content that is removed by the platform for infringing copyright should reference the original work, the dates it was available and all other engagement and interaction records.

Should content be added at a time other than the original creation timestamp, platforms should note the timestamp of the addition and the process that led to its retroactive inclusion. For example, external reports (academic, government or account holder) or change in platform procedures (updating classifier algorithm, internal policy change).

All content and account unique identifiers should be in the original and persistent form commonly used by the platform. For example, on Facebook, pages and posts are referenced by 64-bit integers. YouTube identifies channels with a string. Platforms should not provide resource identifiers that are referentially consistent only within their transparency data resource."

Reporting standards

"Data should reflect the native formats of the producers rather than ad-hoc aggregations.

The minimal standard for data delivery shall be direct download. Summaries and systems which require interaction with graphical user interfaces are not acceptable substitutes.

Formats must be one of the following:

1. JSON
2. Avro
3. CSV

If data source natively serves JSON in its commercial APIs or server to server communication, it shall be the preferred format.

Acceptable compression formats shall be bzip, gzip or tar. Under no circumstances should a compression format that requires a commercial license be used.

Values must be typed consistently as one of the following:

1. string
2. enum
3. signed integer
4. unsigned integer
5. binary
6. blob (binary large object for media files)
7. double
8. hex
9. boolean
10. Datetime (EN ISO 8601)
11. Unix Epoch in milliseconds.

Additionally, primitive data types native to C, C++, Java or ECMAScript are acceptable.

Encoding shall be consistent and according to UTF and UCS standards.

Complete documentation for fields, objects shall adhere to an open (non-commercial) standard for machine and human comprehension.

An API shall not be a substitute for full, downloadable archives in one of the aforementioned formats. APIs must adhere to an open (non-commercial) standard for machine and human comprehension."

ANNEX III – ACT INTERVENTION ON DEMONETISATION PANEL OF THE STAKEHOLDER DIALOGUE

Thank you for giving us the floor Kristina. I will get straight into the matter and then close my intervention with general remarks

Demonetization of advertising online

For the purposes of the presentation I will use the terms bad ads or bad websites/accounts as short hand to designate advertisement carrying content considered as disinformation or disguised political ads or websites/accounts that promote disinformation.

Let's kick off with some positive outlook

Demonetising the disinformation online ecosystem will be one of the most effective tools to stop the spread of this type of harmful content at its source.

Compared to other areas of disinformation, the ambition level for the commitments and KPIs can be greater for two reasons:

On the one hand, as we are dealing here with flows on a market place regarding paid for content issues related to privacy, censorship or double filtering are less central to the discussion.

On the other hand, this is really an area where the work of regulators, fact checkers, NGOs and others can be leveraged to achieve greater information to market participants.

So what is the issue?

Disinformation pays and pays well. It also has a cost.

It comes at the detriment of the brand integrity of advertisers, of the accounts/websites on which bad ads appear, the general public that are exposed to them and other market participants in the online ecosystem.

We understand that a number of services and safeguards exist or are being put in place.

But if these services are so effective why do we continue to see a stream of reports by the press, academics, NGOs and fact checking organisations revealing that the problem is still very much prevalent and the economy of disinformation is still going strong.

So what is wrong?

Our view is that there is inherently a market failure at play here which results from a broken set of incentives. The Commission's Guidance needs therefore to ensure strong commitments from signatories to align incentives. Ensuring in the process that disinformation is seen and addressed as truly bad for business.

To do so we encourage the Commission to look at market instruments to rectify this failure. And notably ensuring that the first requirement for a functioning market is met, notably achieving near perfect information for market participants.

The reality is that market participants remain largely unable to independently verify whether the solutions in place are effective, what volumes of financial flows are involved and have independently strong KPIs to measure whether there is progress within and across platforms on this issue.

…

How (and which kind of) commitments and KPIs are useful in doing so.

Lets first look at the aspect of legitimate ads ending up on illegitimate sites or accounts

The first question for the commitments is to ensure better information as to the size of the issue. We cannot understand or address what we cannot measure. We need commitments to report and independently verify the size of the disinformation economy, where this money goes, whether advertisers aware, refunded and or otherwise compensated.

This means in turn KPIs that measure the amounts of money paid out, from which legitimate advertising sources, and for what amount of time. This will help understand the amounts we are dealing with, the reaction times to curb the phenomenon and the principle victims and how they are being compensated for this.

The second question is whether the commitments can incorporate strong measures to prevent connection methods that are known to be used by persons creating multiple accounts or websites containing disinformation. The use of proxy services here is one example, used to coordinate several accounts. If this is a mainstream tool to develop coordinated behavior the commitments could narrow in on such tools and at the very least give indicators such as the extent of their use.

The third area, is whether market participants clearly understand where their advertisements are placed and for what reasons. This commitment is spelt out in the Code of Practice, yet the problem subsists – the commitments need to step up here in terms of ensuring tools are provided within and across platforms to ensure standards emerge and complaints can be tracked and independently reported to ensure actual progress. The commitments should also ensure advertisers are able to benefit from the work done by the community of what could be called independent flaggers regarding the harmful websites or accounts on which their ads land rather than rely solely on the assessment provided by the Platform.

In sum are we providing the near perfect information required to rebalance the incentive structure and in so doing are we leveraging all of the good work done by fact checkers, academics, regulators and other parties.

And is this information, available to all market participants so as to ensure they can also understand how disinformation affects the online advertising markets as a whole. This is to ensure bad activity can be rapidly identified but also allow recourse for websites/accounts that may be unduly flagged.

I want to be clear that again we are dealing with commitments regarding the transparency of ad placements, not a tool for platforms to arbitrarily decide, beyond existing legal requirements, criteria that could filter out legitimate news or other sites.

The ultimate decision should rest with the advertiser, but in order to do so he/she will need appropriate transparency tools based on data he can trust and easily understand at a granular level whilst market participants will need information at an aggregated level to understand how this is affecting market dynamics.

Now I turn to the bad advertisers and how we can deal with them

I will be shorter on this point given time.

At the source the main issue is about proper vetting of advertisers and their content. There needs to be a commitment towards knowing and vetting your business customer and the advertisements being placed.

This carries a number of KPIs focusing on review and bad ads removed despite passing the vetting process.

As concerns ads of a political nature, it is essential to take a broad view encompassing both official political advertising and the more pernicious issue-based political advertising

There also need to be commitments to inform legitimate websites/accounts on whether such ads have appeared on their space and ensure they have the tools to ensure this does not happen again. The same should be expected for legitimate ads that ended up on the same page as the bad ads.

This includes information on the person on whose behalf advertising is placed, the period during which it was displayed, the criterion used to target one or more particular groups and the total reach.

Users that have been affected by this should also benefit from commitments whether this resulted in a financial scam or other financial transactions using the platforms' tools

These are some of the commitments that we believe may ensure a realignment of incentives at market level to ensure removing disinformation from the advertising space is achieved for the bettering of an absolute business priority.

General remarks

We welcome this dialogue as an opportunity to rebuild public trust in the online environment and in so doing protect our societies and democracies. We see it as a first step towards far more detailed proposals on commitments and KPIs.

The expectations are high – not only from stakeholders present in these dialogues but members of the general public, users of these platforms and governments across the world.

To rebuild trust we need to be clear about why the Code has failed to deliver and put in place the necessary commitments and KPIs that can help rebuild public trust.

To achieve this, will not mean marginal improvements to the Code but a step change in terms of transparency, investment, and internal/external policies. The requests are not new and had already been proposed by the Sounding Board in the establishment of the Code as a key prerequisite to achieve tangible and measurable outcomes.

There is much to say about each and every of the fundamental changes required. A key starting point in terms of commitments for the Guidance will necessarily be that all relevant data sets for the reporting are independently verified. The General public (and their respective Governments) cannot be expected to support a reporting system that does not have this fundamental element of transparency at its core.

As Commissioner Breton said in the Special Committee of the European Parliament just last Monday:

The code of practice needs to evolve to become a real co-regulation framework".

"Become a real co-regulation framework". The Commission's own principles for better self and coregulation are clear on what this means and notably that monitoring be done in a manner that is "sufficiently open and autonomous to command respect from all interested parties"

So this is the standard we should look to achieve as part of this process.

Again this requires a step change in terms of the principles for the guidance and the KPIs required to satisfy the criteria of such schemes – and far more detailed technical discussions involving the data experts. We look forward to an ongoing discussion of the proposals that will be put forward to ensure all interested parties can support the Code 2.0 when it is enacted.