

## THE RESONANT PROJECT and the FOCUS GROUPS

### 1. The RESONANT project

FIMI (Foreign Information Manipulation and Interference) is one of the main obstacles societies are confronted with in the last years. Due to ongoing technological developments, it became easy to circulate misinformation and carry out information suppression. This has consequences for individuals as well as for societies as if it is difficult to recognise and detect such activities. In December 2023 the multi-disciplinary research project [RESONANT](https://resonantproject.eu) was launched. Its primary objective is to analyse and document the Tactics, Techniques, and Procedures (TTPs) employed by both state and non-state actors in the realm of Foreign Information Manipulation and Interference (FIMI). The project is funded by the Horizon Europe programme and will run for three years. (<https://resonantproject.eu>)

### 2. The purpose of focus groups activity

RESONANT Work Package 2 (WP2) investigates Tactics, Techniques and Procedures (TTPs) that are part of Foreign Information Manipulation and Interference (FIMI) operations targeting diaspora communities. As RESONANT D1.2 *Baseline Report* made clear, FIMI find in online platforms, especially social media, a fertile ground to breed.<sup>1</sup> The aim of WP2 is to advance knowledge about the use of TTPs to spread disinformation online. As part of this endeavour, WP2 also seeks to provide insights into the tools and measures that are or can be used to stem FIMI's diffusion or to mitigate its negative effects on online platforms. In this framework, task 2.2 and the present one, task 2.3, are both concerned with the detection and the evaluation of online FIMI operations, but they adopt a different approach.

On the one hand, Task 2.2, led by KEMEA, traces and annotates the development and the spreading of known events of FIMI that were observed in the past. By doing so, Task 2.2 adopts a reactive police investigative approach, i.e., based on tools such as investigative research and handwriting annotation. Such a reactive approach (to detection and evaluation of FIMI) is well aligned with the legal approach to speech regulation and protection as enshrined under the international human rights' legal framework (art.19 ICCPR see below 'The discussion topic').<sup>2</sup>

On the other hand, this task, Task 2.3, led by VUB, deals with measures and tools that seek to proactively detect and, leveraging on modern technologies, such as Artificial Intelligence (AI), automatically evaluate content or material at the point it is uploaded and before it is disseminated to the public. Instead of the reactive approach taken by Task 2.2, Task 2.3 seeks to advance knowledge of the methods and tools for proactive identification and automated evaluation and, more specifically, about the legal, technological, and social

---

<sup>1</sup> REACTION, D1.2 Baseline Report.

<sup>2</sup>Emma J. Llansó, 2020. "No amount of "AI" in content moderation will solve filtering's prior-restraint problem." *Big Data & Society* 7:1; Barrie Sander. 2019. "Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation." *Fordham Int'l LJ* 43: 939. Judit Bayer et al. "Disinformation and propaganda—impact on the functioning of the rule of law in the EU and its Member States." *European Parliament, LIBE Committee, Policy Department for Citizens' Rights and Constitutional Affairs* (2019).

barriers, requirements, and conditions that these methods and tools should or have to consider.

### 3. The discussion topic

#### **Technologies for proactive detection and automated evaluation of online content**

In the context of the efforts to stem information manipulation, the EU and certain member states, such as Germany and France<sup>3</sup>, have adopted policies or laws that encourage or require online platforms, under certain circumstances, to implement technological measures to proactively detect and automatically evaluate illegal or harmful content. Powered by Artificial Intelligence (AI), tools for proactive detection and automated evaluation of online content have the capacity to flag, block, or remove content.<sup>4</sup> They include Natural Language Processing (NLP), Optical Character Recognition (OCR) and Digital hash technologies that recognizes text, image and voice, translates into data that computer can analyse enabling automated evaluation.

Directive 2000/31 on e-commerce and the recent Digital Services Act (DSA), which came into force in August 2023, prevent member states from imposing a general obligation on the hosting platforms to monitor the material hosted (the so-called liability exemption), except for imposing monitoring obligations in specific cases and following orders emanating from national authorities in accordance with national legislation.<sup>5</sup> Whilst retaining the principle of the liability exemption, the DSA introduces an innovation. It puts providers of very large online platforms (VLOPs) under the obligation to “diligently identify, analyse and assess any systemic risks” (article 34), where ‘systemic risks’ include generic risks such as “actual or foreseeable negative effects for the exercise of fundamental rights”, “any actual or foreseeable negative effects on civic discourse and electoral processes, and

---

<sup>3</sup> Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) , in short DSA. In Germany, the Network Enforcement Act (NetzDG) adopted in June 2017 to improve the enforcement of existing criminal provisions on the Internet and, more specifically, on social networks. The NetzDG is available at: <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html>. In France, two related laws on information manipulation were adopted in December 2018 and a law on online hate speech, the so-called Avia law, was adopted in May 2020. LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037847559/> and LOI n° 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042031970>

<sup>4</sup> For instance, Natural Language Processing (NLP) understands text and spoken words to humans and then translates them into data that a computer can analyze. Optical Character Recognition (OCR) recognizes text within an image and converts it into text, allowing for automated flagging. Digital hash technology translates images and videos into strings of text and numbers called hashes that are then matched with pre-existing databases of classified hashes, enabling identification.

<sup>5</sup> European Parliament and Council (2000) On certain legal aspects of information society services, in particular electronic commerce, in the Internal Market. Directive 2000/31/EC, Recital 47 and article 15.

public security", etc.<sup>6</sup> Given the large volume of material uploaded every day on online platforms, and the vague terms in which the systemic risks are defined, proactive detection and automated filtering offer an attractive option, in terms of cost and speed, as compared with fact-checking by human operators.<sup>7</sup>

The development and use of these technologies (for proactive detection and automated evaluation of online content) is a recent phenomenon whose impacts and implications remain to be appraised. One of the major sources of concern and disagreement among lawmakers and academics revolves around the impacts that AI powered technologies of this sort have on the regulation and protection of the fundamental right to free speech (art. 19 ICCPR; art. 10 ECHR).<sup>8</sup> The DSA is to be interpreted and applied, states Recital 153, "in accordance with [those] fundamental rights, including the freedom of expression and of information, as well as the freedom and pluralism of the media."<sup>9</sup>

### **Collisions with the human right to freedom of expression**

The human right to freedom of expression includes the right "to seek, receive and impart information and ideas through any media and regardless of frontiers."<sup>10</sup> This human right is recognised in the constitutions of modern western states as a key defence of an open and pluralistic democratic society.<sup>11</sup>

Under international human rights law (IHRL), states are prevented from requiring media, like newspapers, Tv, or online platforms, to take down content or material, unless specific conditions are met. Under IHRL, states can indeed require platforms to remove, label or restrict circulation of material on the basis that to do so is provided for in the law, meet a legitimate aim and is necessary and proportionate to attain one of the purposes in art.19(3) ICCPR, such as protection of public health or national security, or if speech incites the commission of crimes. These are derogations that must be applied narrowly and reasonably. States cannot require platforms to remove, label or restrict circulation of material on other grounds, such as that the content is shocking, or disturbing, offensive, or

---

<sup>6</sup> "...any actual or foreseeable negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being." Ibid.

<sup>7</sup> Back in 2013, Twitter reported that it saw, on average, 500 million posts per day, excluding re-tweets. In 2023, X reported it saw 100 million to 200 million posts per day, excluding re-tweets. Social Media Today, "Elon Musk Says X Users Are Posting Fewer Posts Per Day Than People Had Been Tweeting", by Andrew Hutchinson, 19 September 2023. <https://www.socialmediatoday.com/news/elon-musk-says-x-users-posting-fewer-posts-per-day-people-been-tweeting/694043/>

<sup>8</sup> The concern is acknowledged in Regulation 2021/784 on addressing the dissemination of terrorist content online according to which "effective online measures to address terrorist content online and the protection of freedom of expression and information are not conflicting but complementary and mutually reinforcing goals. Competent authorities and hosting service providers should thus only adopt measures which are necessary, appropriate and proportionate within a democratic society, taking into account the particular importance accorded to the freedom of expression and information and the freedom and pluralism of the media." Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 (2021) On addressing the dissemination of terrorist content online. (OJ L 172, 17.5.2021, pp. 79-109) (Recital 10)

<sup>9</sup> DSA, Recital 153.

<sup>10</sup> Article 19(2), 1966 International Covenant on Civil and Political Rights (ICCPR)

<sup>11</sup> Thomas Irwin Emerson .1970. *The System of Free Expression*. New York: Vintage Books. Jacob Mchangama .2022. *Free speech: A global history from Socrates to social media*. UK:Hachette.

that it provides an inaccurate or biased portrayal of reality, even downright false. The reason for this rigidity is that if these other, broad restrictions on speech were condoned, they could easily be – and historically have been – bent towards authoritarian ends.<sup>12</sup> For this reason, under IHRL, any restrictions on speech must be provided for in the law and be narrowly compelled. This includes States' interventions designed to stem FIMI by imposing filtering obligations on online platforms. These obligations-setting interventions must be proven necessary, justified by an actual risk to fundamental rights, and the measures adopted to stem it must be proportional to the severity and likelihood of occurrence of the risk.<sup>13</sup>

### **The regulation of tools for proactive detection and automated evaluation of online content**

Given the hefty toll that information manipulation online levies on the fabric of democracy, requiring a response, considering the pressure from governments, and the huge volume of content uploaded every day, recourse to proactive detection and automatic evaluation tools by online platforms is likely to remain or increase. It is plausible that the conflict between IHRL and the laws encouraging or dictating the use of proactive detection and automatic evaluation tools will lead to compromises so that efforts against online manipulation can withstand freedom of expression challenges or meet the requirements of art.19(3).<sup>14</sup> It is equally conceivable that such reconciliation (between responses to information manipulation with the fundamental right to free speech) will not be based on top-down, command and control-type legislation only. The dynamic nature of regulation in cyberspace suggests that, to be effective, any compromise will include the voices of other actors and stakeholders.<sup>15</sup> In addition to lawmakers and online platforms, law enforcements agencies (LEAs) and civil society organisations (CSOs), this includes journalists, media professionals, media councils, as well as new actors such as fact checkers, researchers of false narratives, and developers of social media content moderation programs.

Against this backdrop, Task 2.3 contributes insights about some of the legal, social and technological conditions, barriers, requirements, that proactive detection and automatic evaluation tools in online platforms should or will have to consider addressing FIMI operations, while respecting and protecting freedom of speech.

---

<sup>12</sup> Jacob Mchangama .2022. *Free speech: A global history from Socrates to social media*. UK:Hachette.

<sup>13</sup> Kate Jones. "Protecting political discourse from online manipulation: The international human rights law framework." *European Human Rights Law Review* 1 (2021): 68-79; J. Bayer et al., "Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States" (European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, 2019). For the latter, freedom of expression imposes obligations on states in respect of disinformation, when the state is responsible for distorting or allowing others to distort completely the whole information environment.

<sup>14</sup> See for instance the decision of the French Constitutional Court on hate speech law. Décision n° 2020-801 DC du 18 juin 2020 Loi visant à lutter contre les contenus haineux sur internet <https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm>

<sup>15</sup> Andrew Murray, 2007. *The regulation of cyberspace: control in the online environment*. London: Routledge-Cavendish.

## 4. Three Focus Groups

To advance knowledge about the regulation – i.e., the legal, technological, and social conditions – that tools and measures for proactive identification and automated evaluation must take into account, task 2.3 organises three focus groups.

The three focus groups leverage on the disciplinary expertise of each hosting partner - URJC in the social and political sciences, VUB in law and ethics, BayHfoD in law enforcement and are named the city where the partner is located:

- Project partner and task leader VUB organises the Brussels FG on legal and ethical issues.
- Project partner URJC oversees the Madrid FG on social issues.
- Project partner BayHfoD oversees the Munich FG on technological issues.

The FGs elicit the views and perspectives of experts coming from government agencies, online platforms, academia, journalism, CSOs.

### 4.1 The Brussels Focus Group : *The responsibility of states to protect democratic processes from online Foreign Information Manipulation and Interference (FIMI)*

Recent years have shown that campaigns of online manipulation and interference, originating at the behest of domestic or foreign agents, be they state and non-state actors, can threaten the fabric of democratic processes. They leverage on modern technologies, such as AI, data analytics, sentiment analysis etc. and have the effect of stymieing the political debate, polarising views, directing or deterring people from voting in elections.

This focus group discusses the accountability of states under IHRL for foreign manipulation and interference operations. FIMI – Foreign Information Manipulation and Interference – begs the unsettled problem of distinctions such as between legitimate political debate and illegal manipulation or interference. Distinctively, the qualification of “foreign” calls to the fore the rules and principles governing the relations and conduct of sovereign states, individuals, and multinational corporations, such as online platforms.

The Focus Group explores which international human rights law references could and should be mobilised to frame the challenges and responses to online FIMI operations. As starting point, the Brussels Focus Group adopts three distinctions regarding accountability for foreign interference operations under IHRL (Jones, 2021):

- International accountability*: State A's obligations (under IHRL) to individuals in State B ("international"), to the extent State A has "extraterritorial jurisdiction" in respect of individuals in State B;
- Domestic regulation and accountability*: State B's obligations (under IHRL) to individuals (and occasionally groups) within its jurisdiction.
- Corporate responsibilities*: the responsibilities of platforms, wherever located, to respect the human rights of individuals in State B.<sup>16</sup>

---

<sup>16</sup> Kate Jones. "Protecting political discourse from online manipulation: The international human rights law framework." *European Human Rights Law Review* 1 (2021): 68-79. p.3

Starting from this point, the focus group will seek to advance knowledge about the accountability of State-sponsored information interference & manipulation under IHRL, about the impacts of FIMI operations and responses on individual human rights, and about the negative and positive obligations of states towards their citizens.

The discussion will be articulated around a series of discussion points:

- **State-sponsored information interference & manipulation under international human rights law (IHRL)**
  - The jurisdictional limits of the state's duty to respect (not manipulate, not interfere) human rights;
  - What is the threshold beyond which state responsibility is engaged? The "physical power or control over the individual" threshold and the case of surveillance undertaken extraterritorially;
  - Does the right of peoples to "freely determine their political status" (collective self-determination) provide a practicable legal basis for seizing FIMI?
- **The impacts of FIMI operations and responses on individual human rights**
  - Right to participate in public affairs and to vote and the limits of manipulation and interference;
  - Right to personal data protection and the targeting of messages without receipt's awareness;
  - Right to privacy and the protection of a free space of deliberation. The chilling effect of interferences;
  - Right to free speech and the narrow scope of restrictions provided in national law.
- **Domestic regulation and accountability**
  - Negative and positive obligations owned by states in tackling information manipulation and interference domestically;
  - State law and online platforms: giving guidance and ensuring accountability, problems of jurisdiction and conflict of laws;
  - Obligations of online platforms as corporate business under IHRL.

**Potential Participants:**

- **Legal Experts:** Specializing in digital rights, privacy law, data protection, cybersecurity, and EU regulations (e.g., GDPR, DSA).
- **Public Authorities:** Representatives from law enforcement agencies, data protection authorities, and national regulators involved in digital investigations, EU institutions
- **Academics:** Researchers specializing in criminal law, digital rights, human rights and international law, data protection and privacy law.
- **Digital Service Providers:** Legal and policy advisors from major online platforms (e.g., social media companies, search engines).
- **Technology Experts:** Specialists in cybersecurity, data encryption, and platform design to discuss technical safeguards for data access and security.

## 4.2 The Madrid Focus Group: *media professionals, online platforms, and checking content*

Online platforms possess the technological know-how and are already using advanced tools for generating user-engagement content advertising as well as for enforcing their terms and conditions for content moderation.<sup>17</sup> However, platforms merely host content, thus they do not have the editorial responsibilities that befall on publishers.<sup>18</sup> There have been renewed calls for traditional media responsibility standards to be applied to social media platforms as a result of their decisions on what news to display to whom, as news editors with responsibility for its topics. The question follows, whether social media platforms, through their algorithms that rank and curate third-party submissions, exert a form of editorial control traditionally performed by media professionals and therefore engage specific media responsibilities. Professional journalists, meanwhile, must navigate a delicate balance between engaging on social media — to extend journalism's reach — and avoiding pitfalls that undermine journalism's deontology.<sup>19</sup>

To counter information manipulation, online service providers have already implemented voluntary, self-regulatory measures to address FIMI operations and disinformation campaigns. These efforts include adhering to codes of conduct, community guidelines, enforcing terms of service (ToS), and employing fact checking companies, or automated tools for content moderation. It is likely that we are going to see more of automated tools for moderating content in the future than media professionals checking content. One illustrative example is Meta's acquisition back in 2016 and deployment of CrowdTangle, an analytics tool used by researchers, watchdog organizations, and journalists to monitor the spread of information—including misinformation—on Facebook and Instagram. In August 2024, Facebook and Instagram parent Meta Platforms shut down CrowdTangle.<sup>20</sup> In January 2025, Meta CEO Mark Zuckerberg that the social media company would stop working with third-party fact-checking organizations. Reportedly, Zuckerberg explained his views on content moderation have changed. Meta has made "too many mistakes" in how it applied its content policies, he said. "So we are going to get back to our roots, focus on reducing mistakes, simplifying our policies, and restoring free expression on our platforms," he added.<sup>21</sup>

### Discussion points and questions:

- Moderation on social media (functions, barriers...)
- The role that regulators and independent fact-checkers should play to minimize the impact of misinformation and disinformation (FIMI) on society.

---

<sup>17</sup> Alexandre De Stree, op.cit., p.43. Llansó, (2020),op.cit., p.2

<sup>18</sup> Charis Papaevangelou. 2023. "The non-interference principle: Debating online platforms' treatment of editorial content in the European Union's Digital Services Act." *European Journal of Communication* 38.5.

<sup>19</sup> Cherilyn Ireton and Julie Posetti (eds). 2018. *Journalism, 'Fake News' & Disinformation. Handbook for Journalism Education and Training*. UNESCO

<sup>20</sup> Meta says it will end fact checking as Silicon Valley prepares for Trump, January 7, 2025 By Huo Jingnan, Shannon Bond, Bobby Allyn, 7January 2025 . <https://www.npr.org/2025/01/07/nx-s1-5251151/meta-fact-checking-mark-zuckerberg-trump>

<sup>21</sup> Ibid.

- Fact-checking methods: Advantages and disadvantages of various solutions, particularly those involving humans.
- Information suppression: addressing the various challenges related to freedom of expression, self-censorship, and the provision of safe spaces.
- Role of fact-checkers on online platforms versus society.
- View of the impacts of misinformation and disinformation (FIMI) on society and democratic systems, especially related to social polarization.
- What is the role of fact checkers in online platforms?
- Can technological tools for detection and evaluation replace human checkers?
- What role should regulators and independent fact-checkers play in ensuring a fair and transparent process?

## Potential Participants:

- **Representatives from regulatory and governmental bodies** (e.g., officials working on the DSA in the EC).
- **Industry Experts and Moderation Policy Experts** from Meta, Google, Amazon, and other major platforms.
- **Legal Advisors** from platforms: lawyers or policy advisors familiar with the legal implications of the DSA on platform operations.
- **Tech Engineers:** those responsible for implementing moderation systems, data access for researchers, and other technical aspects of the platform's compliance with the DSA.
- **Digital Rights Advocacy Groups** (e.g., European Digital Rights - EDRI, Access Now, etc.).
- **Academics and researchers** working on social media and fact-checking or specializing in countering misinformation.
- **Journalists and Media Professionals** (e.g., tech journalists, investigative journalists).



### 4.3 The Munich Focus Group: AI tools for Identifying FIMI incidents online

AI is increasingly evoked to monitor online speech and for detecting and evaluating content deemed illegal or harmful, either by governments or digital platforms. Detecting and evaluating content, including FIMI, through AI presents technical and legal challenges that are at the centre of the Munich FG.

The development of AI systems to counter disinformation has been characterized as a 'double-edged sword' when it comes to information threats due to their dual nature<sup>22</sup>. On one side, AI systems have made it easier and quicker to create and distribute fake texts, images, videos, and audio pieces (e.g., deepfakes, bots)<sup>23</sup> that can appear real and ultimately aim to manipulate their target audience<sup>24</sup>. Generative AI (GenAI) tools make it more challenging to detect and counteract misleading or inaccurate narratives in the digital domain. On the other, AI technologies provide practical solution to the problem of detecting and evaluating disinformation against the enormous volume of information.

The question arises: "Can they?" Specialised literature indicates that there are some challenges. Technically, AI systems cannot distinguish content and intention<sup>25</sup>, which means that legitimate expressions can be mistakenly flagged<sup>26</sup>. A critical issue with text-based detection tools is the fact that they often face issues such as false positives/negatives and limitations in handling multiple languages<sup>27</sup>. In the context of FIMI, this challenge is intensified due to the absence of universally accepted definitions. Inconsistent legal standards complicate the development of detection criteria. AI models based on machine learning can be trained to classify articles as true or false using labelled data<sup>28</sup>. Current advancements in machine learning (ML), however, present drawbacks: they rely heavily on large datasets, can be prone to instability, and may not adapt well to new problems or datasets. Additionally, due to their reliance on complex hidden layers, they are often regarded as "black box" solutions, and therefore lacking in transparency and explainability<sup>29</sup>. If these tools rely on complex algorithms that distinguish content based on syntactic features rather than intent or context, they can make flawed decisions, further complicating the enforcement of fair and accurate moderation.<sup>30</sup> Moreover, automated

---

<sup>22</sup> Linda Slapakova. 2021. *Towards an AI-Based Counter-Disinformation Framework* (Hague diplomacy Blog, 24 March 2021), available at: <https://www.universiteitleiden.nl/hjd/news/2021/blog-post---towards-an-ai-based-counter-disinformation-framework> (last accessed: 18/06/2024).

<sup>23</sup> Cristos Velasco. 2022. "Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments" (2022), 23, *ERA Forum*, p. 112. <<https://link.springer.com/article/10.1007/s12027-022-00702-z?fromPaywallRec=false#citeas>>

<sup>24</sup> Raquel Miguel. 2024. 'Platforms' policies on AI-manipulated and generated misinformation' (EU DisinfoLab, 4 June 2024), available at: <https://www.disinfo.eu/publications/platforms-policies-on-ai-manipulated-and-generated-misinformation/> (last accessed: 19/06/2024).

<sup>25</sup> Linda Slapakova (2021), op.cit.

<sup>26</sup> Ibid.

<sup>27</sup> Ibid.

<sup>28</sup> Noémi Bontridder and Yves Pouillet, "The Role of Artificial Intelligence in Disinformation" (2021) 3 *Data & Policy*, p. 7-9, <https://www.cambridge.org/core/journals/data-and-policy/article/role-of-artificial-intelligence-in-disinformation/7C4BF6CA35184F149143DE968FC4C3B6>

<sup>29</sup> Ibid.

<sup>30</sup> Llansó, (2020),op.cit.

systems are limited in their capacity to accurately distinct incidents of expression where cultural or contextual cues are necessary. They may also lead to discriminatory outcomes due to biases embedded in historical data used to train the AI systems and could potentially result in de facto profiling of specific groups. When AI tools are used for moderating content posted online they automatically perform filtering, removing or blocking content online, deprioritising its visibility or disabling accounts, automatically<sup>31</sup>. The removal or filtering of content, including advertisements, is widely regarded as a highly effective strategy for addressing disinformation; however, it also raises concerns regarding prior restraint and the protection of free speech.

### Discussion questions

- Are AI-driven tools effective and reliable for detecting and analysing FIMI?
  - What kind of AI tools are being used so far in FIMI detection, and which is their level of effectiveness?
  - What are the technical limitations of AI in detecting foreign information manipulation targeted at diasporas?
- How can AI systems mitigate ethical risks, such algorithmic bias, overreach in content moderation, and chilling effects on speech?
  - How can AI systems be programmed to reduce false positives/negatives while minimising risks to equality, non-discrimination, and freedom of expression?
  - How can AI systems address linguistic and cultural biases that lead to unequal detection rates of FIMI across different regions or languages?
  - How can bias in AI systems be minimized to avoid discriminatory outcomes?

### Potential Participants:

- **AI and cybersecurity experts**
- **Legal professionals** specializing in AI regulation
- **Representatives from online platforms** developing AI moderation tools
- **Members of diaspora communities** with insights into targeted disinformation
- **Academics in AI ethics** and **international law**
- **Researchers participating in other related EU Horizon Projects** (e.g. VIGILANT).

---

<sup>31</sup> Ibid.

## 5. Focus Groups –Organisation

The following section provides an outline of the organisation the FGs (structure, format, timeline and methodology).

### ❖ **Format**

The DoA leaves partners free to choose the format, in person or online. Each partner responsible for each focus group is thus free to decide the best format.

### ❖ **Number of participants:**

RESONANT's DoA foresees three FGs of around 5 participants each in a. Madrid, b. Brussels, and c. Munich (Task 2.3).

It is Suggested the involvement of between 6 to 10 external experts, in addition to a moderator and assistant-note taker.

### ❖ **Duration of each session:** 2 hours maximum.

### ❖ **Suggested Timeline** - Focus groups are expected to take place between mid-February and March 2025.

### ❖ **Method** : Predefined topics and research questions guide the discussions (see section above), with flexibility allowing participants to raise specific issues.

### ❖ **Recruitment of participants:**

Each contributing partner invite candidate participants.

Incentives: In principle, no incentives are going to be given to the participants. However, each partner makes an independent decision on incentives for the participants in their respective country as envisioned in the project description and note this for the final report of D.6.3.

### ❖ **Focus group moderation**

A trained moderator (or facilitator), familiar with the topic under discussion, who coordinates the focus group and an assistant to help with note taking and assistance.

### ❖ **Recording and transcription**

The FGs will be carried out under Chatham House Rule (<https://www.chathamhouse.org/about-us/chatham-house-rule>).

Notes will be taken manually. The sessions will not be recorded.

### ❖ **Storage of output**

Results from FGs (transcriptions or reports) will be uploaded on the designated repository and downloaded by VUB to deliver D2.3. Partners are free to exchange results among themselves for RESONANT research and administrative purposes.