INACH

Bringing the Online In Line with Human Rights

licra

INACH First Shadow Monitoring Report 2021

Compiled by Maia Feijoo and Tamás Berecz

2021

TABLE OF CONTENTS

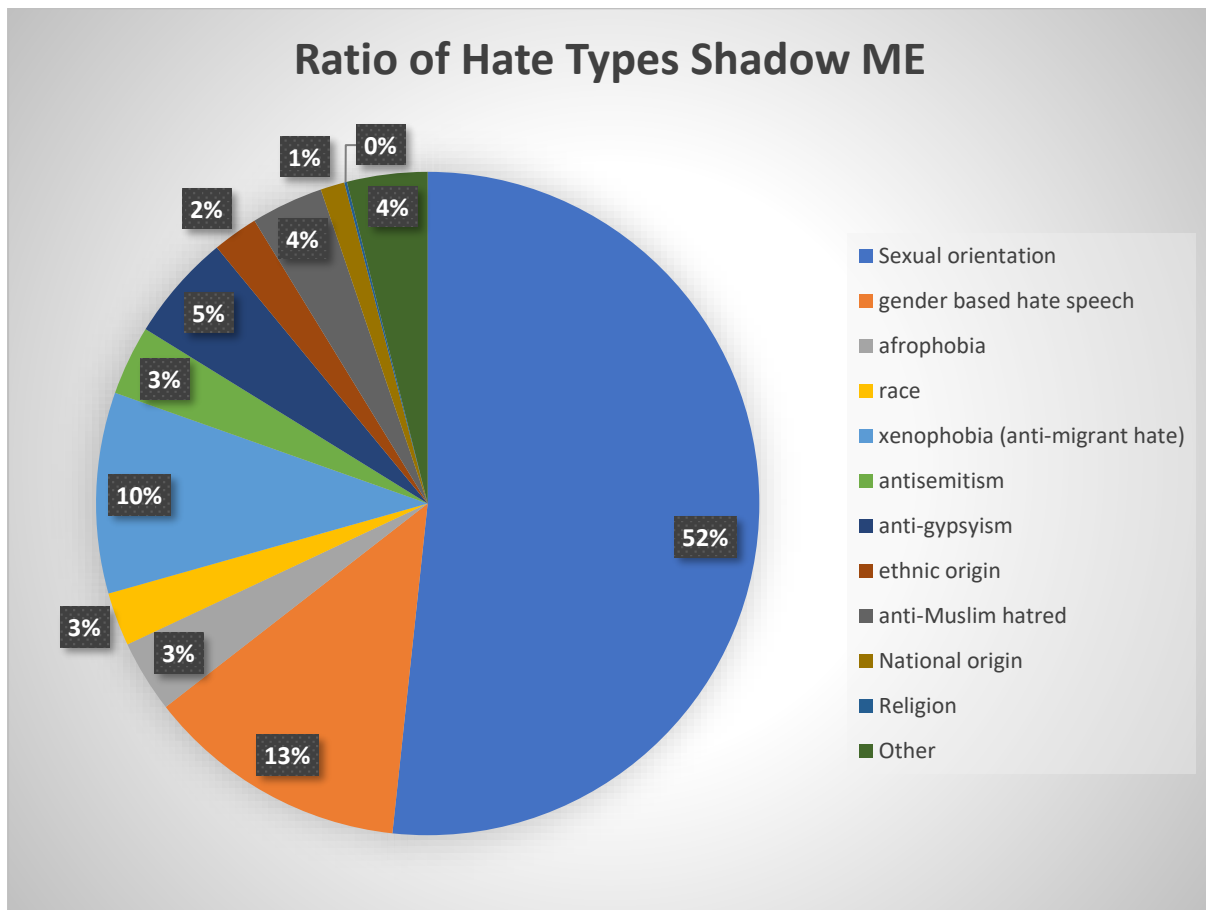## 1) Basic Information about the First 2021 Shadow Monitoring

Following the sixth monitoring exercise, the shadow monitoring has begun on the 19th of April 2021 and ended on the 7th of May. The goal of this exercise was to monitor the IT companies and verify if they are respecting the Code of Conduct, aside from the official period of monitoring, conducted by the European Commission.

The following NGOs took part in the shadow monitoring: ILGA, Háttér Society, CESIE, ROMEA and LGL. They represented five European countries: Portugal, Hungary, Italy, Czech Republic, and Lithuania. They monitored five IT companies: Facebook, Twitter, Instagram, YouTube, and Tik Tok; and they collected and reported 658 instances of illegal hate speech to the platforms.
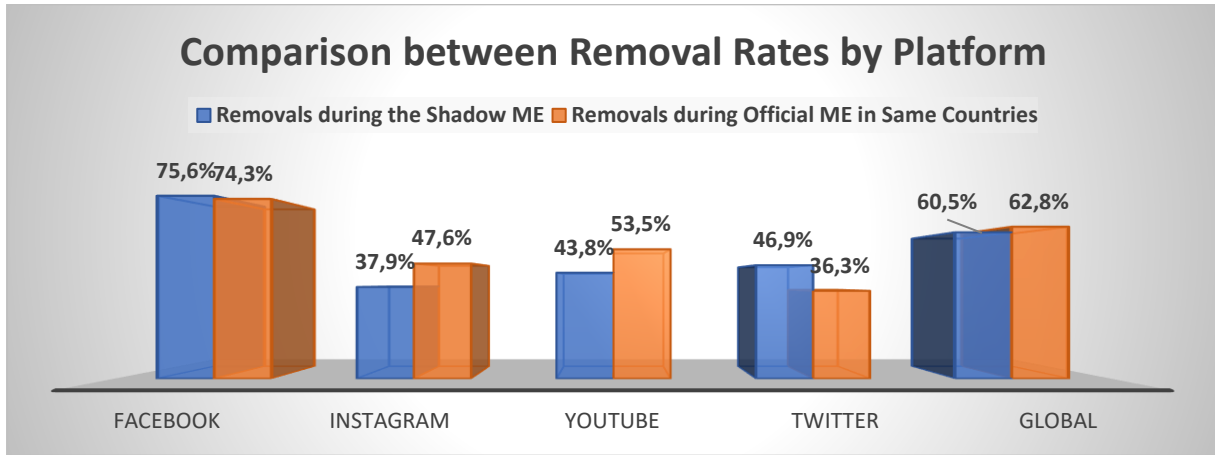
We observed that, during the shadow monitoring, the removal rate was lower than during the official monitoring. Reducing from 62,8 per cent to 60,5 per cent for removing online hate speech. Except in Hungary, where the removal rate has been higher during the shadow monitoring, increasing from 36,1per cent to 59,3 per cent. It was the same case in Czech Republic, with a total of 49,1 per cent during the shadow monitoring, whereas the removal rate during the official monitoring was at 45,5 per cent.

2) Findings of the First 2021 Shadow Monitoring

Due to the participation of two LGBTQ+ rights NGOs in the first shadow monitoring round, hate based on sexual orientation and gender were by far the most prominent hate types. However, anti-refugee hate, anti-Gypsyism, anti-Muslim hatred and antisemitism were also very prevalent, just like in the official 2021 ME.

Ratio of Hate Types Shadow ME

Legend:
- Sexual orientation — 52%
- gender based hate speech — 13%
- afrophobia — 3%
- race — 3%
- xenophobia (anti-migrant hate) — 10%
- antisemitism — 3%
- anti-gypsyism — 5%
- ethnic origin — 2%
- anti-Muslim hatred — 4%
- National origin — 1%
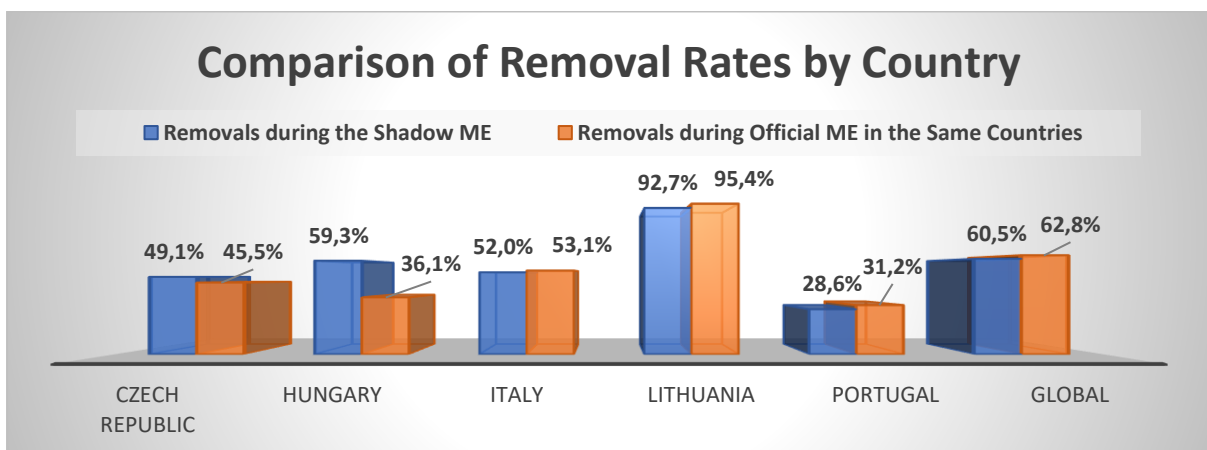- Religion — 0%
- Other — 4%

Instagram and YouTube did considerably worse in removing illegal content from their platforms during the shadow monitoring. Facebook removed almost exactly the same amount during the two exercises, and – as a surprising outlier – Twitter removed approximately 10 per cent more reported illegal content than during the official ME. Globally, it can be seen that the companies did somewhat better in removing reported illegal hate speech from their platforms during the official ME.

Comparison between Removal Rates by Platform

■ Removals during the Shadow ME ■ Removals during Official ME in Same Countries

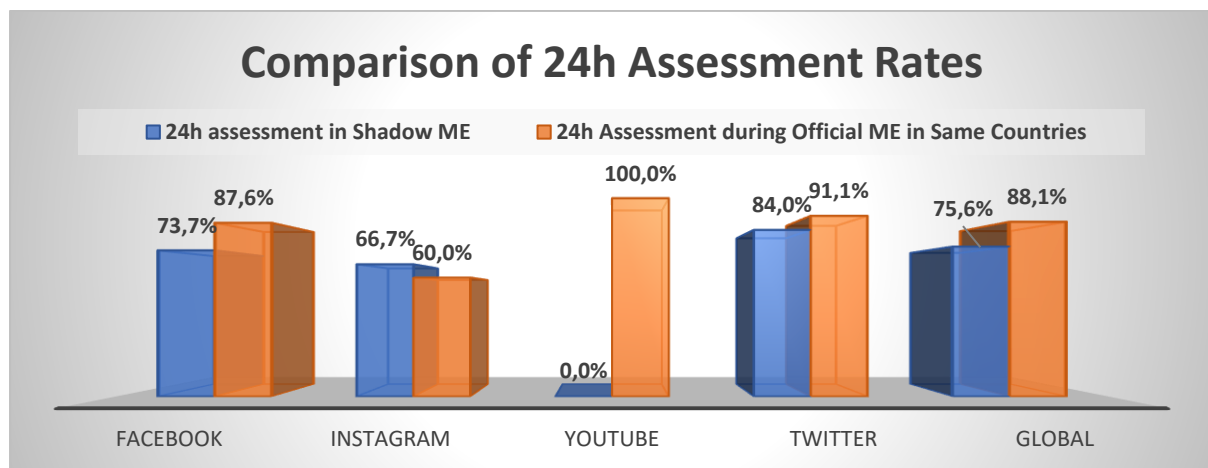|	FACEBOOK	INSTAGRAM	YOUTUBE	TWITTER	GLOBAL
Shadow ME	75,6%	37,9%	43,8%	46,9%	60,5%
Official ME	74,3%	47,6%	53,5%	36,3%	62,8%

When we examine the removal rates by country, we have to highlight two different issues, the first is the difference between the outcomes of the official ME and the shadow. As it can be seen, in Italy, Lithuania and Portugal, the platforms removed more reported illegal content during the official ME than the shadow one. In the Czech Republic, the removal rates were a little bit better during the shadow monitoring, whilst in Hungary, as a major outlier, the removal rate was more than 10 per cent higher during the shadow ME than the official one.
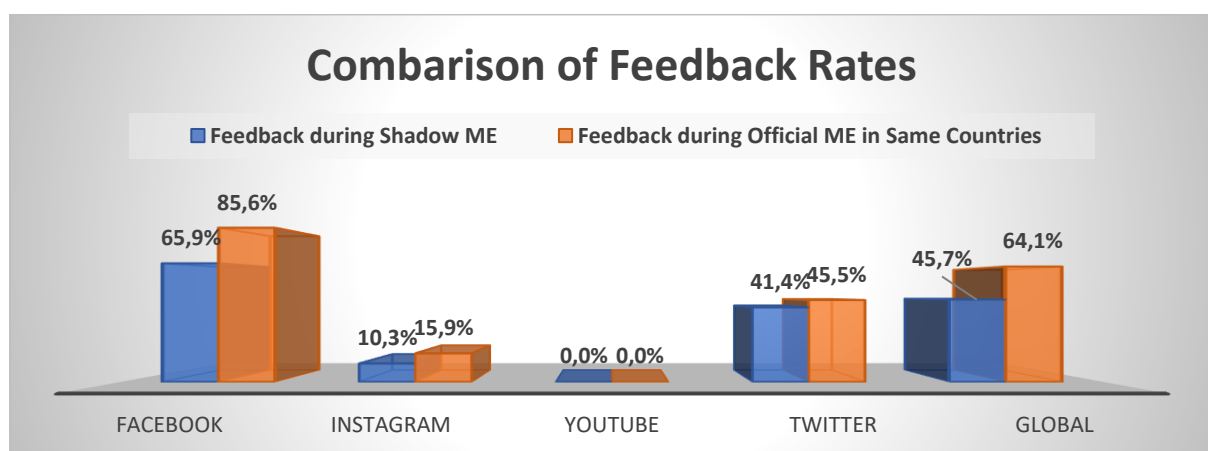
This leads us to the second issue we must shine light on: there are just too big discrepancies between the removal rates in different countries and in general. This is an issue that has been highlighted many times before, yet it has not been adequately addressed by the companies.

Comparison of Removal Rates by Country

■ Removals during the Shadow ME ■ Removals during Official ME in the Same Countries

	CZECH REPUBLIC	HUNGARY	ITALY	LITHUANIA	PORTUGAL	GLOBAL
Shadow ME	49,1%	59,3%	52,0%	92,7%	28,6%	60,5%
Official ME	45,5%	36,1%	53,1%	95,4%	31,2%	62,8%

The biggest differences between the findings of the shadow monitoring and the official ME can be observed if one takes a look at the 24-hour assessment rates of the platforms. Except for Instagram, all platforms did a lot worse in assessing reported content within 24 hours. Globally, the platforms assessed almost 13 per cent less illegal content within 24 hours as required of them by the Code of Conduct. This shows that conducting further shadow monitorings is essential to highlight the shortcomings of the platforms outside of the official monitoring periods.

Comparison of 24h Assessment Rates

- ■ 24h assessment in Shadow ME
- ■ 24h Assessment during Official ME in Same Countries

	Shadow ME	Official ME
FACEBOOK	73,7%	87,6%
INSTAGRAM	66,7%	60,0%
YOUTUBE	0,0%	100,0%
TWITTER	84,0%	91,1%
GLOBAL	75,6%	88,1%

In general, the platforms are not good at providing clear, concise, and timely feedback to reports. This became even more apparent during the shadow monitoring. Globally, the platforms provided almost 20 per cent less feedback during the shadow monitoring than the shadow one, with all platforms producing worse numbers. And the difference cannot even be described as marginal, especially when it comes to Facebook.

Combarison of Feedback Rates

- ■ Feedback during Shadow ME
- ■ Feedback during Official ME in Same Countries

	Shadow ME	Official ME
FACEBOOK	65,9%	85,6%
INSTAGRAM	10,3%	15,9%
YOUTUBE	0,0%	0,0%
TWITTER	41,4%	45,5%
GLOBAL	45,7%	64,1%

Conclusion

The findings of the first 2021 shadow monitoring clearly show the relevance and importance of continuing this type of monitoring activity. The companies did mostly worse in all indicators showed in this report, especially in the ones that they usually do worse in during the official monitoring periods. The differences in the assessment rates, and especially the feedback rates are worrisome and definitely something that the platforms should focus on. Therefore, we urge the Commission to put more pressure on the companies when it comes to assessment times and especially feedback rates.