

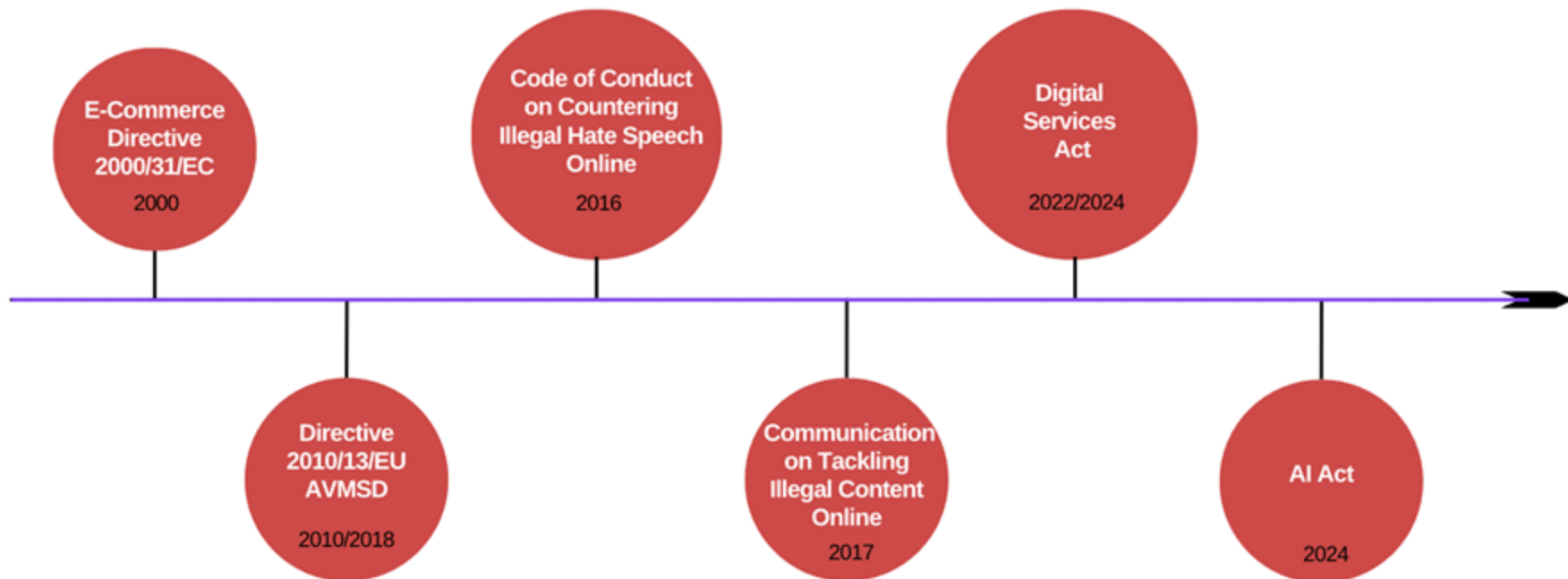


The Digital Services Act and the impact on hate speech monitoring and responses

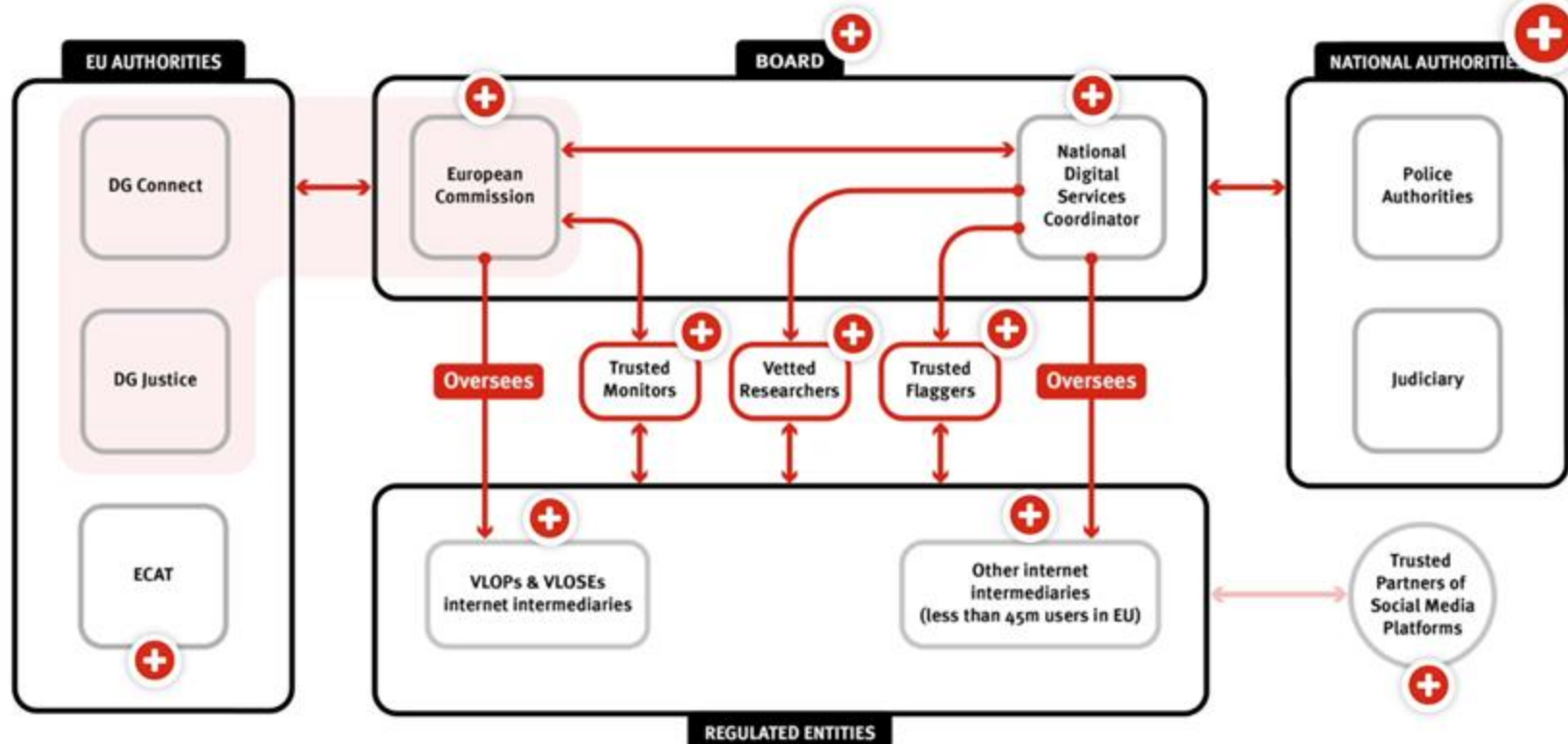


**Co-funded by
the European Union**

Focusing on online hate







- **Faded red arrow (between regulated entities and trusted partners):** represents a pre-existing relationship between trusted partners and VLOPs, which is not included in the DSA. The relationship remains active for organisation that do not plan to apply for Trusted Flagger status under the DSA, but still actively collaborate with some regulated entities through established channels.
- **One-way arrow:** indicates a dependency or an influence of one actor on another, meaning that the actions or functions of one actor rely on the oversight, regulation, or support of another actor. For example, VLOPs (Very Large Online Platforms) and VLOSEs (Very Large Online Search Engines) depend on the European Commission for oversight and regulatory guidance.
- **Two way arrows:** signify strong cooperation and mutual dependency on information sharing, which contributes to and supports the work of the actors involved.



*SafeNet, transparency reports of
regulated entities and CivilNet...*



DSA Transparency Report - October 2024

Introduction

This report covers the content moderation activities of X's international entity Twitter International Unlimited Company (TIUC) under the Digital Services Act (DSA), during the date range 1 April, 2024 to 30 September, 2024.

We may refer to "notices" as defined in the DSA as "user reports" and "reports".

Description of our Content Moderation Practices

Our content moderation systems are designed and tailored to mitigate systematic risks without unnecessarily restricting the use of our service and fundamental rights, especially freedom of expression. Content moderation activities are implemented and anchored on principled policies and leverage a diverse set of interventions to ensure that our actions are reasonable, proportionate and effective. Our content moderation systems blend automated and human review paired with a robust appeals system that enables our users to quickly raise potential moderation anomalies or mistakes.

Policies

X's purpose is to serve the public conversation. Violence, harassment, and other similar types of behaviour discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our Rules are designed to ensure all people can participate in the public conversation freely and safely.

X has policies protecting user safety as well as platform and account integrity. The [X Rules and policies](#) are publicly accessible on our Help Center, and we are making sure that they are written in an easily understandable way. We also keep our Help Center regularly updated anytime we modify our Rules.

For the purposes of the summary tables below, the X policy titles in use at the start of the reporting period have been retained, even if they changed throughout the period.

Enforcement

When determining whether to take enforcement action, we may consider a number of factors, including (but not limited to) whether:

Thank you!

Daniel Heller

Project Officer

daniel.heller@ceji.org

Follow our courses on Facing Facts Online

