

Are Online Platforms “Flunking” When it Comes to Safety and Viewpoint Neutrality?

It seems so simple: a grading system for online platforms to see how they’re performing in terms of safety and viewpoint neutrality. If an objective grading system did exist, should YouTube receive an A- or a C+? One grade for safety and another grade for neutrality? How about X, TikTok, Instagram or Google Search? Detailed and current information transparency is needed to keep the grading accurate and useful for comparisons of online platforms. With this level of transparency, a [wide range of academics, media](#), government agencies and companies can conduct peer-to-peer comparisons of the online platforms.

Rather than waiting to [receive a subpoena from House Judiciary Committee Chairman Jim Jordan](#) about the enforcement actions that result from their policies, Google and other platforms can simply provide full transparency of data reporting on a quarterly basis and avoid government regulatory actions and congressional hearings. These transparency reports should include all communications/requests from government and government-funded entities (with exception for specific national security and law enforcement issues).

This open transparency will provide an alternative to entities [like NewsGuard](#) who use their own private subjective opinions to render judgement on the truthfulness of news sites and online social media platforms. This is by-definition a subjective and false exercise due to the fundamental problem of “Who Decides?” what information is true or false, rather than a disagreement among scientists, politicians or simply differing opinions.

The largest platforms today offer [broad reports](#) (note – download the CSV file on this link which lists the general categories of enforcement actions) which contain no description of the specific enforcement actions taken such as user warning/ban, label on the content, content removal, or de-amplification of the content. These broad reports only mention extremely broad categories of reasons for the enforcement actions, such as government request, child porn, violence or undefined “misinformation”. Transparency requires details on the specific content categories where enforcement actions occurred including specific controversial or political topics such as climate, corruption, medical research, war/terrorism, historical analysis and all political policies/opinions. Transparency reports need to include communications from government and government-funded entities (with exception for actual law enforcement and national security) linked to any enforcement actions.

As enforcement actions at the scale of millions of users and billions of topics become more automated and based on [AI judgements](#), these measures of enforcement actions become more important. Only with far more detailed and granular data transparency can online safety and viewpoint neutrality be measured and graded.

With detailed transparency and reporting, each platform can pursue its own individual content moderation policies for online safety and viewpoint neutrality. Within the limits of what is legal, some platforms may choose to be the GreenPeace platform, family-friendly

platform, adult-content platform, or right-wing or left-wing political platform, pro-capitalism or pro-socialism, etc. As long as the platform is publicly transparent about its content enforcement actions, the population of users and advertisers can freely use whichever platforms they prefer - knowing that the content may be biased, unbiased, or curated using any number of published content rules and guardrails within the limits of what is legal.

In today's highly polarized political climate, it is nearly impossible for Congress to establish universal guardrails for all online platforms. Partisan politicians can't even agree on basic definitions of online safety and viewpoint neutrality, making any attempt at government-enforced content moderation deeply problematic. Introducing a government referee to oversee the guardrails of acceptable online speech sets a dangerous precedent—what one administration enforces today could be weaponized by the next. Given the historical pattern of election-driven power shifts in Washington, such regulatory authority is ripe for abuse, potentially leading to [unforeseen consequences](#) that could undermine free expression and fairness in the digital space.

Online platforms may or may not be ready to voluntarily expand their current quarterly reporting to include the specific enforcement actions taken, the specific content categories in which the enforcement actions were taken, and specific requests from named government entities. This is why a very simple federal mandate that only requires the platforms to transparently publish detailed information related to all their online enforcement actions is needed.

With detailed information transparency, multiple third-parties can grade whether YouTube and the other major platforms deserve an A+ or C- for online safety and an B+ or A- for viewpoint neutrality. The harsh light of publicity will have a profound effect on motivating these companies to improve their performance in online safety and viewpoint neutrality (or non-neutrality for platforms who publicly position and promote their planned biases).