

Anthropic’s Transparency Hub: Platform Security

A look at Anthropic's key processes, programs,
and practices for responsible AI development.

Platform Security

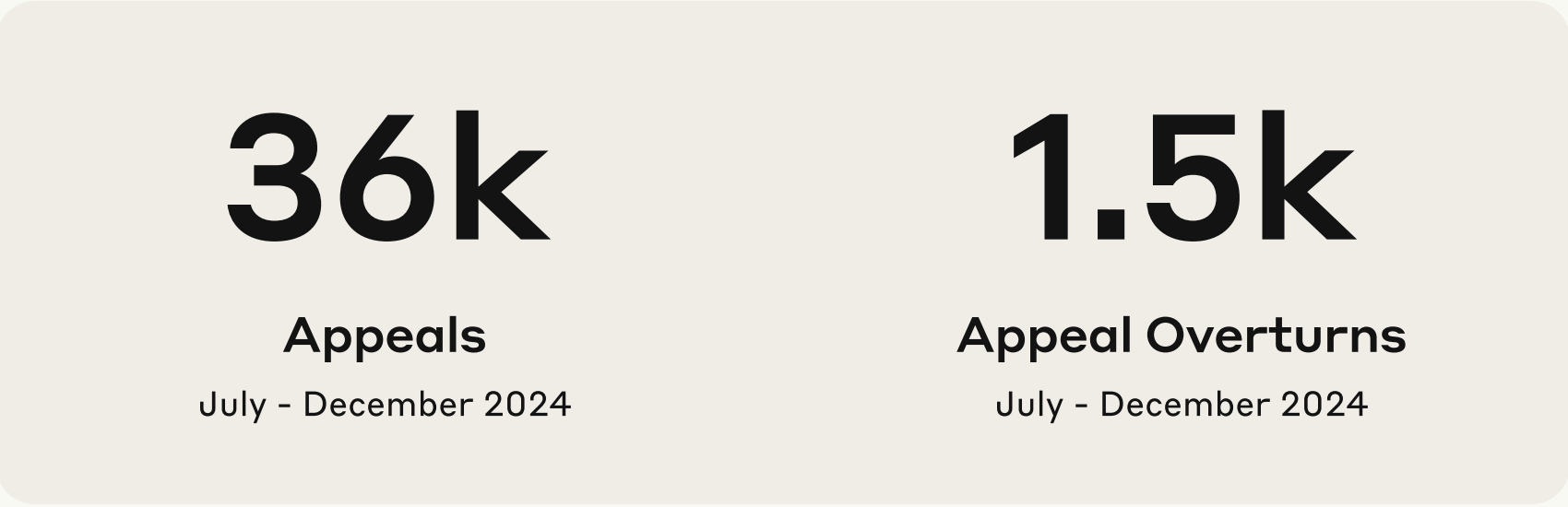
Last updated Feb 27, 2025

We are sharing more detail on our Usage Policy, some enforcement data, and how we handle legal requests to enable meaningful public dialogue about AI platform safety.

Banned Accounts

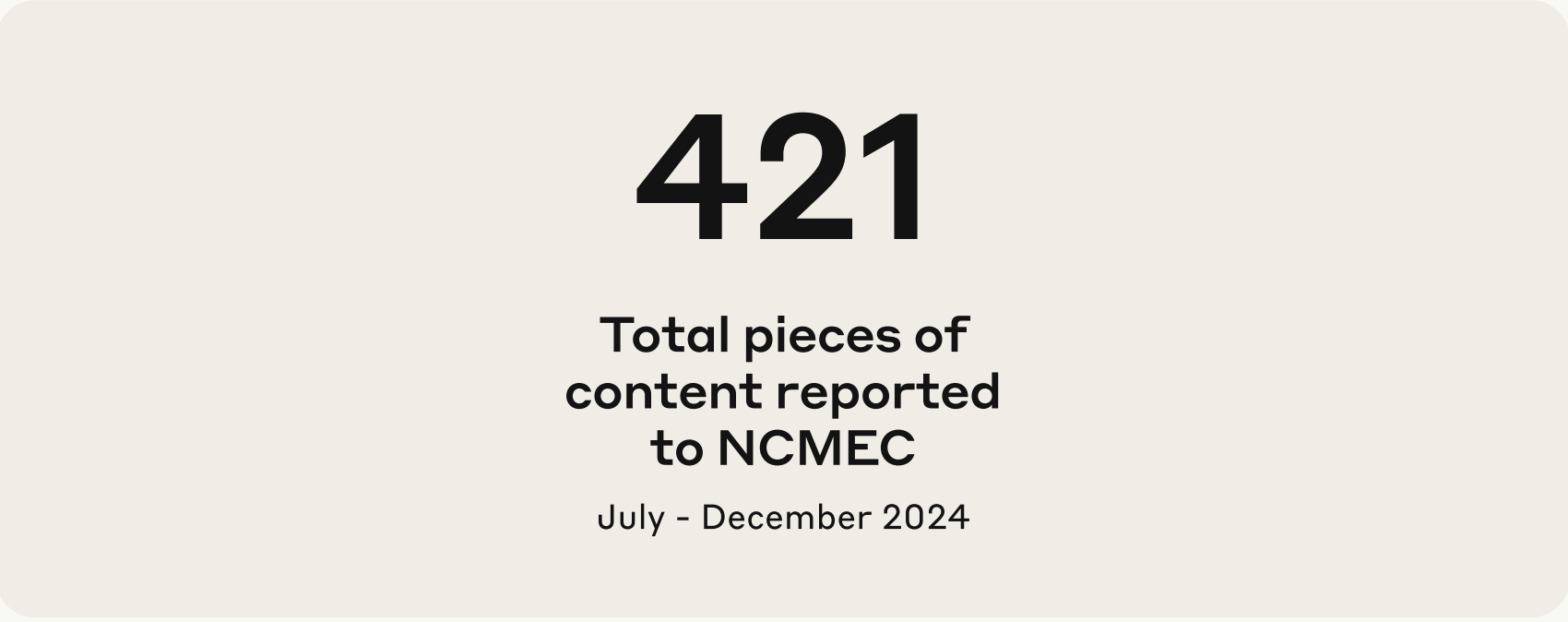


Anthropic’s Safeguards Team designs and implements detections and monitoring to enforce our Usage Policy. If we learn that a user has violated our Usage Policy, we may take enforcement actions such as warning, suspending, or terminating their access to our products and services.



Banned users may file [appeals](#) to request a review of our decision to ban their account.

Child Safety Reporting



Anthropic is committed to combating child exploitation through prevention, detection and reporting. On our first-party services, we employ hash-matching technology to [detect and report known CSAM](#) to NCMEC that users may upload.