

Publier le code source ne suffit pas...

Un court billet où Nicolas Kayser-Bril opère une mise au point : la loi européenne et les grandes entreprises du Web peuvent donner accès au code source, mais ce n'est qu'un facteur parmi d'autres qui s'avère souvent inutile à lui seul...

Article original : The ideology behind publishing Twitter's source code publié dans le bulletin d'information (en anglais ou allemand) d'Algorithmwatch auquel on peut s'abonner sur cette page.



Traduction Framalang **relue et révisée par l'auteur** : Squeeek, goofy, audionuma

L'idéologie derrière la publication du code source de Twitter

par **Nicolas Kayser-Bril**



*Photo par
Julia
Bornkessel
licence CC-BY
4.0*

Une fuite

Le 31 mars, Twitter a publié une partie du code source qui alimente son fil d'actualité. Cette décision a été prise quelques jours après qu'il a été rendu public que de grandes parties de ce code avaient déjà été divulguées sur Github [Gizmodo, 31 mars].

Les 85 797 lignes de code ne nous apprennent pas grand-chose. Les tweets ne contenant pas de liens sont mis en avant. Ceux rédigés dans une langue que le système ne peut pas reconnaître sont rétrogradés - discriminant clairement les personnes qui parlent une langue qui n'est pas reconnue par les ingénieurs californiens. Les *Spaces* (la fonction de podcasting en direct de Twitter) sur l'Ukraine semblent également être cachés [Aakash Gupta, 2 avril].

Le plus intéressant dans cette affaire reste le billet de blog rédigé par ce qu'il reste de l'équipe d'ingénieurs de Twitter. Il explique bien comment fonctionne un fil d'actualité d'un point de vue technique.

Comment (ne pas) ouvrir le code source

Une entreprise a été pionnière pour rendre son code source public : Twitter. Il y a deux ans, son équipe « Éthique, Transparence et Responsabilité » a publié le code d'un algorithme de recadrage d'images et a organisé une compétition permettant à quiconque d'y trouver d'éventuels biais [AlgorithmWatch, 2021]. Cette équipe a été l'une des premières à être licenciée l'année dernière.

Il ne suffit pas de lire un code source pour l'auditer. Il faut le faire fonctionner (l'exécuter) sur un ordinateur. En ce qui concerne l'Ukraine, par exemple, nous savons seulement que les *Spaces* Twitter étiquetés « UkraineCrisisTopic » subissent le même traitement que les articles étiquetés « violence » ou « porno ». Mais nous ne savons pas comment cette étiquette est attribuée, ni quels en sont les effets. Il semble que le code de ces fonctionnalités n'ait même pas été rendu public.

Dissimulation

Publier du code informatique sans expliquer comment le faire fonctionner peut être pire qu'inutile. Cela permet de prétendre à la transparence tout en

empêchant tout réel audit. Twitter n'est pas la première organisation à suivre cette stratégie.

La Caisse Nationale des Allocations Familiales a publié les 7 millions de lignes du code de son calculateur d'allocations suite à une demande d'informations publiques (demande CADA) [NextINpact, 2018]. On ne pouvait rien en tirer. J'ai fait une demande de communication des « documents d'architecture fonctionnelle », qui sont mentionnés dans des commentaires du code. La CNAF a répondu qu'ils n'existaient pas.

La loi européenne sur les services numériques prévoit que les « chercheurs agréés » pourront accéder aux « données » des très grandes plateformes, y compris éventuellement au code source [AlgorithmWatch, 2022]. Pour que la loi sur les services numériques fonctionne, il est essentiel que les entreprises traitent ces demandes comme le Twitter de 2021, et non comme le Twitter de 2023.

L'idéologie technologiste

Enfin, la focalisation sur le code source est au service d'un projet politique. J'entends souvent dire que le code est le cœur d'une entreprise, que c'est un secret commercial précieusement gardé. C'est faux. Si c'était le cas, les fuites de code source nuiraient aux entreprises. Suite à des intrusions ou des fuites, le code source de Yandex et de Twitch a été publié récemment [ArsTechnica, 2021 et 2023]. À ma connaissance, ces entreprises n'en ont pas souffert.

Le code source n'est qu'un facteur parmi d'autres pour une entreprise du Web. Parmi les autres facteurs, citons les employés, les relations avec des politiques, les procédures internes, la position sur le marché, l'environnement juridique et bien d'autres encore. Mettre le code sur un piédestal implique que les autres facteurs sont sans importance. Les propriétaires de Twitter et de Meta (et ils sont loin d'être les seuls) ont dit très clairement que les ingénieurs étaient beaucoup plus importants que le reste de leurs employé·e·s. Pour eux, tout problème est fondamentalement technique et peut être résolu par du code.

Je suis certain que la publication du code source de Twitter conduira certains technologues à prétendre que le harcèlement en ligne, les agressions et la désinformation peuvent désormais être « corrigés » par une *pull request* (lorsqu'un contributeur à un projet open source propose une modification du

code). Ce serait un pas dans la mauvaise direction.