Written evidence submitted by Sophie Zhang (OSB0214)

As a Facebook whistleblower, the Committee requested my testimony on the Bill, which I provided on October 18, 2021. The following is intended to cover proposals regarding specific areas of the draft bill that I believe could be improvements. I have no expertise in legislation, regulation, or British law, but hope nevertheless that my advice may be useful.

Proposal: Exemption for very small businesses.

The current Draft Bill requires regulated services to file risk assessments with OFCOM, with exemptions listed in Schedule 1. Services qualify as 'regulated' under Clause 3 if they are a user-to-user service or search service that have a significant number of United Kingdom users or United Kingdom users form one of the target markets for this service. This includes every new British social media company or search company upon creation (perhaps before they've even created the service.)

These risk assessments could potentially be taxing or difficult for new startup founders or small businesses, and potentially hinder Britain's innovation and entrepreneurship. I would hence suggest that an exemption be created for businesses under a certain total size – defined by the number of worldwide users, or total funding. The presumed goal would be to include large businesses that seek to expand into Britain or into search/user-to-user services, while not covering early-stage startups and the like.

Proposal: Enforcement for requirement that companies apply policies consistently

The current Draft Bill enumerates that companies are required to apply their terms of services consistently (Clauses 9.5b, 10.5b, 11.3b, 12.5b, 13.5b, 14.7b, 21.5, etc.) However, there does not appear to be enforcement for this clause.

As a result, I would propose legally requiring large companies (Category 1 under the OFCOM registry described in clauses 59 and 60) to report significant inconsistencies in enforcement to OFCOM. As an example, this requirement would have required FB to explain its Xcheck system in-depth to OFCOM before the testimony of myself and Frances Haugen regarding the subject, with the potential for penalties under Clause 72 if prior explanations were found to be inaccurate.

It may also be useful to require Category 1 services to separate their product policy and outreach/government affairs departments, in order to dissuade inconsistencies caused by conflicts of interest. Currently, these two departments are combined at Facebook – forcing a separation has been proposed by myself and multiple former senior FB employees, including Alex Stamos and Samidh Chakrabarti.

Proposal: Clarity regarding Clause 46

Clause 46 of the current Draft Bill enumerates that companies have a duty to protect users from "content that is harmful to adults" and not previously denoted as illegal. Parts of harmful content are defined in subsections 46.3 and 46.5 based on the service provider having "reasonable grounds to believe that the nature of the content is such that there is a material risk of the content having, or

indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities."

This appears to be a deeply subjective definition, in which a service provider could plausibly argue that many individual pieces of content are harmful or not harmful. The lack of clarity in this clause can create uncertainty among service providers, uncertainty regarding enforcement for OFCOM, and lead to severe differences in enforcement between companies as each interprets it in a different direction. The Committee should consider making Clause 46 more clear.

Proposal: Oversight regarding company risk assessments.

The current Draft Online Safety Bill appears to rely on self-provided company risk assessments (as enumerated in Clauses 7 and 19.) Oversight of these assessments appear to only take place via OFCOM investigations (Clause 75), a clause that one presumes will be reserved for rare and unusual circumstances.

As a result, I have two separate proposals for how to provide independent oversight and verification:

1) Empower OFCOM to perform red-team penetration tests on certain types of legal but harmful activity. I understand that it would be illegal for OFCOM to perform this with illegal activity (e.g terrorist content.) However, this could be a useful manner of helping enforce Clause 46 regarding content that is harmful to adults but legal. For instance, OFCOM could determine the ability of companies to deal with message harassment, by sending legal insults in controlled manners to test users created by OFCOM and reporting the message. OFCOM could also determine the ability of companies to avoid false positives in reporting by mass reporting innocuous activity by an OFCOM test user as violating, and seeing the response. Furthermore, I understand that fake accounts are not illegal in Britain unless they impersonate a real person; as a result, OFCOM could set up troll farms in very controlled circumstances to independently assess the company's abilities to catch them.

2) Require companies to provide data access to trusted researchers, and fund those researchers. This would have to be done extremely carefully, due to the potential for abuse of the data and attendant risks to privacy. University researchers are not perfect – as Aleksandr Kogan can illustrate.

Proposal: Expanded whistleblower protections for OFCOM investigations

If OFCOM opens an investigation under Clause 75, it may be very difficult to understand the inner workings of the investigated company without proper context, or determine which employees have pertinent information to the investigation. It may be useful to give whistleblower protections to any employee who voluntarily gives pertinent testimony or documentation regarding the area of investigation to OFCOM. I understand that existing British whistleblower protections fall under the Public Interest Disclosure Act of 1998, and requires the whistleblower to reasonably believe a severe violation has occurred (including criminal offense, failure to comply with legal obligations, or danger to the health/safety of employees.)

Proposal: Clarify relationship with encryption

A number of Britons have expressed concern that it is the intent of the Bill to stop end-to-end encrypted services by providing them with an impossible legal duty to limit illegal and harmful content. This uncertainty and lack of clarity is damaging to the Bill. The Draft Committee should clarify the intent of the Bill with regard to encryption via an official statement. My opinion is that a ban on E2E in Britain would be a drastic and unnecessary move.

9 November 2021