標題：Facebook 使命—貼文的安全保障

關鍵字：社群軟體、仇恨言論、版主

你相信在處於科技世代的今天，社群軟體的威力甚至高於政府，能控制各國民心和社會運動嗎？一天幾十億的貼文，充滿現代人的生活空隙，搭配推文演算法的努力，潛移默化中強化人類思想，這也難怪社群軟體會被指為民主損害與流血事件的元凶。而像 Facebook 這樣的大科技業者(Big-Tech)，如果不致力於每日的篩選工作，封鎖涉及危險的貼文，可能還會影響網站經營的續航力。「校正人們的說詞不是我們的工作，」資深工程師 Sara Su 說道，「但我們平台仍要堅持團體的某些標準，我們必須確保言論自由與社會安全之間的共識與平衡。」

規則的擬定

每個星期二的早晨，數個 Facebook 員工聚在總公司的玻璃會議室中，共同協議出二十億使用者的言論原則，這些人幾乎都是年輕的律師和工程師，試著將複雜的現實道德問題濃縮成淺顯易懂的是非題，當審核時發現某些特定關鍵字，該貼文就需要遭到封鎖。

紐約時報的 Max Fisher 公開 Facebook 提供的分類準則，他們對仇恨言論(hate speech)的定義猶如大雜燴般，充斥著兩百多條艱澀的專業行話，還有一系列政治黨派與人物的名單，隨時為平台的安全把關。這些文件讓 Facebook 儼然成為全球言論自由的裁決者，影響範疇甚至超乎公眾與 Facebook 自己的認知。

然而，擬定標準的過程未必精確，很可能造成完全偏差或錯誤的判決。例如印尼為火山罹難者舉辦的慈善募款活動曾經被 Facebook 移除，只因為宣傳活動的某間贊助廠商被列為黑名單。

另外，許多文件的內容沒有被及時更新，也沒有定期的文字勘誤，導致漏網之魚四處亂竄。舉例來說，巴爾幹半島 2016 限制標準忘記更新 Ratko Mladic 的身分，標示其為受極端分子推崇的波士尼亞逃犯，但事實上他早在 2011 就被捕入獄。而緬甸著名的恐怖組織在 Facebook 上跋扈多時，煽動駭人的種族滅絕，也該怪罪於一些愚蠢的文字疏忽。

政府的參與及施壓也會影響標準的寬鬆程度，不同國家或地區之間可能有著相當顯著的差別待遇。例如，德國政府有參與社群軟體的管控作業，因此 Facebook 阻攔大量的極右派分子；相對的，鄰居奧地利只被攔下一個組織。

版主的監督

Facebook 將開會擬出來的準則彙整成 PowerPoint 或 Excel 文件，分配給全球 15000 位版主，由他們擔任糾察隊監視來自四方的貼文。而這類的工作幾乎都被外包給人力成本相對低廉的國家，許多來自客服公司的員工，甚至沒有經過密集訓練。

時常需要仰賴 Google 翻譯的版主們，每天需要應付上千篇貼文，在短短的時間內回想不計其數的篩選標準早已成為他們必備的能力。然而，Facebook 提供的規則文件雜亂無章，術語之多令人望之卻步，不少貼文中包含的方言也不斷考驗 Google 翻譯的能力。而當版主發現新的威脅或漏洞時，還可能尋無通報管道，甚至缺乏糾正的動力。

由於有些版主的薪水高低取決於他們的審核速度及精準度，於是不少辦公室內流傳著一項潛規則：若遇到的文章超出所有人的理解範圍，就睜一隻眼閉一隻眼，姑且批准通過。

Facebook 全球業務副總裁 Justin Osofsky 表示：「這樣截彎取直的投機行為將被馳我們的原則，而成為外包中階主管自導自演的成品。」同時，鼓勵斯里蘭卡或緬甸種族清洗的貼文也將可能藏匿在平台中大大小小的角落，無時無刻洗腦大眾。

這凸顯出 Facebook 面臨的棘手問題—他們很難看透龐大的外包公司，掌握版主的所有行蹤，再加上 Facebook 的拓展計劃需要這些公司的援助，因此更難強硬插手他們的運作。

反求諸己

有些棘手的問題出在 Facebook 自己身上，像是推文演算法的使用。這類的演算法會將越多關注的貼文推送給更多用戶，倘若具有煽動性的貼文像滾雪球般越滾越大，很可能阻礙審核系統的有效作業。Facebook 或許該考慮延緩推文演算法的威力，同時也先暫停市場的拓增計畫，畢竟先解決當務之急，才有多餘的心力照顧更多國家的用戶。

回到前述，Facebook 不得忽視規則的定義問題，他們必須在規則的兼容性與判斷的精確度之間取得平衡，前者會牽制版主的行事效率，而後者會影響篩選的可信程度。

另外，Facebook 也應該正視自身對於社會的影響力，在更改條例時要避免僥倖。當某些事件發生，Facebook 便應該檢討核心的漏洞並進行修正，而非粗淺地做些微調再靜觀其變，這就像在飛航期間修理飛機一樣冒險。

Facebook 全球計劃領導 Monika Bickert 說：「我們一天有上億則貼文，潛在的危機源源不絕。在這樣的規格之下，就算我們有百分之九十九的正確率，錯誤判斷的數量還是相當可觀。」在未來的五至十年，機器學習系統逐漸參透人類的語言，版主的配置也會由勞力轉向機器，若目前的問題沒有一個好的下文，社群軟體的副作用將會加速氾濫，危害人類社會的安全。

參考資料：

1.Max Fisher, “ Inside Facebook’s Secret Rulebook for Global Political Speech”, The New York Times, 27 Dec 2018

2. Gideon Lichfield, “Facebook’s leaked moderation rules show why Big Tech can’t police hate speech“, MIT Technology Review, 28 Dec 2018