國立臺灣大學社會科學院新聞研究所

碩士論文

Graduate Institute of Journalism
College of Social Sciences
National Taiwan University
Master's Thesis

# 社群媒體時代的言論自由界線：以「監察委員會」對「立即危害」與「容許差別」之相關裁決為中心

# The Boundaries of Freedom of Speech in the Social Media Era: Focusing on the Oversight Board's Decisions Regarding "Imminent Harm" and "Allowance"

林彥廷
Yen-Ting Eddy Lin

指導教授：林明昕 博士
Advisor: Dr. iur. Ming-Hsin Lin

中華民國 113 年 1 月
January 2024

doi:10.6342/NTU202400262

# 國立臺灣大學碩士學位論文
# 口試委員會審定書
## MASTER'S THESIS ACCEPTANCE CERTIFICATE
## NATIONAL TAIWAN UNIVERSITY

**社群媒體時代的言論自由界線：以「監察委員會」對「立即危害」與「容許差別」之相關裁決為中心**

**The Boundaries of Freedom of Speech in the Social Media Era: Focusing on the Oversight Board's Decisions Regarding "Imminent Harm" and "Allowance"**

本論文係 林彥廷 R08342013 在國立臺灣大學 新聞研究所 完成之碩士學位論文，於民國 113 年 1 月 30 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Graduate Institute of Journalism on 30 (date) 1 (month) 2024 (year) have examined a Master's thesis entitled above presented by Yen-Ting Eddy Lin (name) R08342013 (student ID) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee：

林明昕　　林靈雲

（指導教授 Advisor）

系主任/所長 Director：社會科學院新聞研究所所長洪貞玲



doi:10.6342/NTU202400262

# 序　言

我在 2009 年 10 月註冊了 Facebook 帳號，若依照現行政策的年齡限制，當時 12 歲的國中生恐怕還不被允許出現在社群媒體上。自此之後，除了高三學測前的半年苦讀，每天瀏覽動態消息，藉此得知朋友同學和時事動態，毫無疑問成為生活中必不可少的一部分，說右手拇指已經養成往上滑動的肌肉記憶，也不為過。

我仍然記得：中學時玩得不亦樂乎的 Facebook 網頁遊戲、如今羞於見人的自拍大頭照；高中及大學投入學生自治，藉著社團和粉絲專頁，才能經營組織、分享政見、傳播資訊；出了社會，主持電視節目、輔選民意代表、創立 NGO，更無不是藉著 Facebook、Instagram 創造影響力，最終獲得了還算滿意的成果。2019 年，我去參觀了還名為 Facebook 的臺灣辦公室，在佈滿 emoji 的「神秘屋」拍下了一張照片，彼時的我絕對算社群媒體重度使用者：一半時間推動自己相信的理念與議題，剩下的另一半時間，大抵就花在社群媒體上。

世上沒有絕對的善與惡，社群媒體也未能跳脫此一框架，我至今仍然相信社群媒體影響世界的能力，遠超以往任何時期的傳播媒介，但普羅大眾卻往往只用而不知，這也是我加入 Meta 公共政策團隊的主要動機：能更近距離地一窺在西方世界普遍被認為的邪惡帝國面貌，同時在與以往不同的位置上，為臺灣的公民社會盡一份心，何其有幸，也備感壓力。

我不敢說這是一篇在學術上多有創見的論文，相比更多同儕深厚的法學、傳播學知識，我遠遠不及，但這確實是我總結了過往四年多來，在 University of Pennsylvania Law School、臺大新聞所的經驗與學習，也算是為兩年來在 Meta 的工作結束後，留下的些許念想。

我要特別感謝本篇論文的指導教授林明昕老師，從大二與老師結緣以來，不論是在學術、倡議、生活上，都深切感受到老師照顧學生的用心，往後若有機會也為他人之師，老師將是我的典範。我也非常謝謝兩位口試委員：我在新聞所的導師林麗雲老師，猶記得當時赴美前向老師辭別時，老師給我許多學業上的寶貴建議，以及後來在研究、工作上的提點；有幸在 Meta 時與政大傳院林日璇老師合作，我深深被老師在社群媒體、元宇宙的研究風采折服。三位老師能在百忙之中，指導我的研究大綱、論文口試，也不吝給予鼓勵，我深感榮耀。我也要在此特別向政大傳院陳憶寧院長致意，正是因為您在監察委員會的貢獻，啟發我投入本議題的研究。



doi:10.6342/NTU202400262

自 2018 年末開始備考新聞研究所始，我要感謝許多一路以來給予支持、一起打拼的朋友，特別是從臺大學生會到社團法人臺灣青年民主協會的夥伴：育萌、李欣、蔚慈、羽棻、琇文、奕柔、怡璇等，讓我堅定對公共議題的長遠理想；海螺會的兄弟：亮中、連毅、彤彥，啟蒙我在學術、思想上的發展；當然還有所有曾經、持續照顧我的臺大法律系、臺大新聞所同學，以及許多長久以來提攜我的前輩、好朋友，請允許我不會變成在這裡梳理人際網絡，容我再個別向大家說聲謝謝。

我也要向職涯上的先進們致上謝意：支持我兼顧唸書、工作的「呱吉」邱威傑及「752 研究室」全體同仁；在 Meta 照顧我的 George、Max、Ling、Aurora、Charles 及所有 Meta 同事；引領我踏足新聞傳播領域，並讓我持續相信媒體人精神的公共電視《青春發言人》舒逸琪製作人、賴彥如製作人、黃兆徽老師及製作團隊。因為有您們，才讓我順利完成了臺大新聞所的學業。

作為序言，實際上是我完成論文的最後一步。在校園相對平靜的春節假期，自羅斯福路正門走往辛亥路後門，法學院再到社科院，許多景物如此熟悉，幾年來奔忙趕課、舉辦活動、為校務穿梭的記憶又油然而生。冬陽斜映，從新聞所散步回公館的路上，凱君和我在椰林大道並肩拍了幾張照，謝謝她陪我走的返程這段路，聽著彼此絮絮叨叨，一如這幾年來她的陪伴與包容，是我們最日常而永恆的回憶，也將會一直、一起往前走。

最後，我要謝謝我的家人，阿公、阿嬤、爸爸、媽媽為我的教育和成長付出無數心血，給予我最大的空間和關懷，更無怨地接受我許多的自以為是，讓我始終能無後顧之憂地追逐夢想，請相信我從未視為理所當然，謝謝您們。

在晴空下像個孩子，在風雨中像個大人。

林秀廷

2024 年 冬末初春

於臺大校園



doi:10.6342/NTU202400262

# 摘　要

言論自由作為人民基本權利，也是民主體制的根基所在，但若目的為追求重要公共利益且手段符合比例原則時，國家仍得以法律限制之。然而，對於社群媒體而言，決定內容是否能顯示於平台之上，往往僅是根據私人科技公司自行制定的政策規範，卻儼然已形成國際間最廣泛、適用最多人的言論自由標準。

本文透過耙梳 Meta 公司為 Facebook、Instagram 制定的《社群守則》等內容管理政策、用於審查內容及執行政策的方式，及監督該公司對言論限制是否符合國際人權標準的「監察委員會」，來試圖建構、理解全球最大的社群媒體平台進行內容管理的體系。

此外，本文藉由分析至 2023 年 10 月，監察委員會所作成的 53 件案例裁決中涉及兩項不確定概念的「立即危害」及「容許差別」的案件，釐清前者作為社群媒體得移除、封鎖內容的主要原則之一，係源自聯合國《拉巴特行動計畫》的六項門檻測試。監察委員會並未著重於使用者的發佈意圖、言論與危險的時間差，而更關注與內容有關的危險是否在發佈時已發生或持續進行中，以及評估背景、環境影響下，危害風險實現的可能性。

本文進一步指出，「容許差別」為 Meta 公司創設出對於社群媒體言論自由界線的例外。監察委員會對於出現在政府專制、言論自由受壓迫的國家、地區，或與揭露侵害人權的行為相關的內容，傾向認定為具有新聞價值，在未達「立即危害」的情形下，縱然該內容違反《社群守則》，亦會例外允許繼續顯示於平台上；或雖不具新聞價值，但保留該則爭議內容有助於實現平台的核心價值時，也例外地不予移除，為「三段雙軌」的審查流程。

**關鍵詞：社群媒體、言論自由、Facebook、監察委員會、立即危害、容許差別**



doi:10.6342/NTU202400262

# Abstract

Freedom of speech, as a fundamental right of the people and a cornerstone of democratic systems, can still be legally restricted by states in pursuit of significant public interests, provided the means are proportionate. However, for social media, the decision to display content on platforms is usually based merely on policies set by private technology companies. Yet, these policies have seemingly become the most widely applied and universal standards for freedom of speech internationally.

This thesis analyzes the Community Standards and other content policies enacted by Meta for Facebook and Instagram, the measures used for content review and policy implementation, and the operation of the Oversight Board to oversee whether the company's content moderation decisions are aligned with international human rights standards. This study attempts to understand the content moderation system of the world's largest social media platforms.

Additionally, this thesis analyzes 53 case decisions made by the Oversight Board up to October 2023, focusing on cases regarding the concepts of "imminent harm" and "allowance." The former, a key principle for social media to remove or block content, is derived from the six-part threshold test of The Rabat Plan of Action by the United Nations. The Oversight Board does not strictly require the clarification of users' intent or the temporal gap between speech and danger. Instead, it places more emphasis on whether the related danger had already happened or was ongoing at the time of posting, and it considers the possibility of harm being realized in the context of the content's background and environment.

This study further points out that "allowance" serves as an exception developed by Meta to the boundaries of freedom of speech on social media. The Oversight Board tends to recognize content related to autocratic regimes, suppressing freedom of speech, or unveiling human rights violations as newsworthy. If it does not pose "imminent harm," such content is allowed to remain on the platform even if it violates the Community Standards. Alternatively, if it lacks newsworthiness, but retaining the content adheres to Meta's core values or if removing it would contradict them, it is also exceptionally retained, constituting a "three-level, two-track" review process.

**Keywords：Social Media, Freedom of Speech, Facebook, Oversight Board, Imminent Harm, Allowance**



doi:10.6342/NTU202400262

# 簡　目



doi:10.6342/NTU202400262

# 詳　目



doi:10.6342/NTU202400262

doi:10.6342/NTU202400262

# 第一章　緒論

**「如果目標是要確保言論不受壓制，更令人注意的擔憂必然是佔據主導地位的數位平台，……切斷言論的權利被有力地掌握在私人數位平台的手中[1]。」**

——美國聯邦最高法院大法官 Clarence Thomas

## 第一節　研究動機

2023 年中，美國科技公司 Meta Platforms 正式於財報中公布旗下的家族應用程式（family of apps）平均月活躍人數（daily active people）超過 38 億人，將近全球人口的二分之一。其中，社群媒體平台 Facebook 的日活躍用戶（daily active users）就超過 20 億人，月活躍用戶（monthly active users）也超過 30 億[2]，遠超全世界任何一個單一國家的總人口數。

回顧 Facebook 問世至今近 20 年，有被評價為社會帶來正向影響者，譬如「阿拉伯之春」始於埃及青年 Wael Ghonim 成立的粉絲專頁[3]，提供免費衛星網路協助印度資訊流通[4]，在俄烏戰爭中助力烏克蘭凝聚民心、創造正面形象、減少錯誤資訊傳播[5]；亦有因為管理失當而飽受批評時，如面對白人種族主義者威脅「Black Lives Matter」群眾時的消極、不作為，進而引發衝突傷亡[6]，又或未經授權洩漏用戶資訊，使劍橋分析（Cambridge Analytica）用以協助特定勢力操控英國脫歐、美國總統大選宣傳[7]。

---

[1] Biden v. Knight First Amend. Inst. At Columbia Univ., 209 L. Ed. 2d 519, 141 S. Ct. 1220, 1227 (2021)

[2] Meta Platforms, Inc. (2023). Meta Reports Third Quarter 2023 Results. *Meta Investor Relations*. https://investor.fb.com/investor-news/press-release-details/2023/Meta-Reports-Third-Quarter-2023-Results/default.aspx

[3] Vargas, J. A. (2012, February 17). Spring Awakening. *The New York Times*. https://www.nytimes.com/2012/02/19/books/review/how-an-egyptian-revolution-began-on-facebook.html

[4] Levy, S.（著），許恬寧（譯）（2022），《後臉書時代：完整解讀社群霸主從起步、成長、爭議到轉型，每一步的選擇與思考》，頁 229-258，天下雜誌。

[5] Bubna, V. (2022, April 10). How Instagram has played a role in accelerating the Russia-Ukraine war. *Modern Diplomacy*. https://moderndiplomacy.eu/2022/04/10/how-instagram-has-played-a-role-in-accelerating-the-russia-ukraine-war

[6] Frenkel, S. & Kang, C.（著），陳柔含、謝維玲（譯）（2022），《獲利至上：你的一舉一動，都是他們的賺錢工具！Meta 集團（Facebook, Instagram, WhatsApp）稱霸全球的經營黑幕》，頁 253-279，野人文化。

[7] Kaiser, B.（著），楊理然、盧靜（譯）（2020），《操弄【劍橋分析事件大揭祕】：幫川普當



doi:10.6342/NTU202400262

無獨有偶，Meta 公司於 2012 年收購，以圖片分享、影音創作為導向的 Instagram，也被揭露導致青少年產生焦慮、憂鬱等心理健康問題[8]；原名 Twitter 的 X，雖然在許多國家廣為政治人物、媒體記者、評論員所用，但鬆散的監管使得色情、毒品、血腥等不宜內容充斥平台[9]；短影音平台 TikTok 更出現許多致命的「挑戰」，吸引用戶模仿危險動作，使得青少年接二連三發生悲劇[10]。

縱然一時之間，社群媒體對於社會整體發展的影響看似毀譽參半，但其已帶來的衝擊卻是無庸置疑。上至撼動專制政權，卻也動搖民主根基；再至重塑行銷產業，同時重創媒體收益；於普羅大眾而言，更是根深蒂固地改變了傳統的閱聽習慣及人際交流模式。

如今在社群媒體上，我們可以看到有人透過直播展演才藝、分享生活點滴，也有商家利用其推廣業務、銷售商品，亦有樂於發佈文章砭議時政、分享新聞者。社群媒體儼然已成為當代實現言論自由的主要管道，其性質既可以是娛樂舞台、日記手札，也可以是行銷媒介、訊息載體。在美國法院中，政府及部分政治人物的社群媒體帳號，更已經被認定為提供社會大眾溝通意見、思想交流的「公共論壇（public forum）」[11]，而不得任意限制民眾瀏覽、留言，否則如同政府以公權力侵害言論自由，為美國憲法第一修正案所不許。

---

選、讓英國脫歐，看大數據、Facebook 如何洩露你的個資來操弄你的選擇？》，頁 142-161、221-230，野人文化。

[8] Wells, G., Horwitz, J., & Seetharaman, D. (2021, September 14). Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

[9] 鏡週刊（01/18/2023），〈【色毒滿推特】裸拍、援交、毒趴全不設防　推特成情色犯罪溫床〉，https://www.mirrormedia.mg/story/20230115soc013

[10] Carville, O. (2022, November 11). TikTok's Viral Challenges Keep Luring Young Kids to Their Deaths. *Bloomberg*. https://www.bloomberg.com/news/features/2022-11-30/is-tiktok-responsible-if-kids-die-doing-dangerous-viral-challenges

[11] 於 Knight First Amend. Inst. at Columbia Univ. v. Trump 一案中，美國聯邦第二巡迴上訴法院引用了聯邦最高法院 1985 年於 Cornelius v. NAACP Legal Def. & Educ. Fund, Inc.的判斷標準，來認定系爭場域是否為美國憲法第一修正案保障下的「公共論壇」，分別是「政府政策或施政（the policy and practice of the government）」、「產權本質（the nature of the property）」和「與表現活動相關性（compatibility with expressive activity）」。第二巡迴上訴法院認為，時任美國總統 Donald J. Trump 的 Twitter 帳號已被用作官方的政策公告使用，且該帳號具備的「互動特色」，是「無限制開放給公眾進行公開討論」，已構成「公共論壇」而不應被任何來自政府的政治立場、觀點歧視所限制。*See* Knight First Amend. Inst. at Columbia Univ. v. Trump, 928 F.3d 226, 232 (2d Cir. 2019) and Cornelius v. NAACP Legal Def. & Educ. Fund, Inc., 473 U.S. 788, 802 (1985)



doi:10.6342/NTU202400262

以上種種，皆在在說明了社群媒體在當今言論市場中，扮演著既關鍵又多樣的角色。近年來，台灣政府及公民社會也隨著國際趨勢，針對網路平台（online platforms）與新聞媒體的分潤、議價機制進行協商，以圖減緩 Google、Meta 等跨國科技公司在廣告市場的強勢地位，並確保新聞媒體作為民主基石永續發展[12]。亦有欲效仿歐盟《數位服務法（Digital Service Act）》精神，要求網路平台依其規模，承擔相應的管理、揭露義務，而推出《數位中介服務法》草案，卻在業者反彈、政治考量夾擊下無疾而終[13]。

誠然，新聞媒體與平台議價、問責網路服務供應商（Internet Service Provider）皆為非常重要的議題，也有儘速立法、明確規管的必要性。但從過往涉及網路平台、社群媒體的科技政策討論中，我們也不難發現，不分國內外的立法者及專家學者，往往未能在初期即有效掌握科技公司實際的商業模式、運作機制，導致討論進程延滯，甚或落得需要調整論述的挑戰。舉凡美國參議員在對 Facebook 創辦人暨執行長 Mark Zuckerberg 的聽證會上，許多問題被評價為「對 Facebook 所知甚少」[14]，又或我國立法委員在呼籲平台應與新聞媒體公平分潤時，仍不時可見停留在以「使用者付費」為基礎的立論[15]。

究其原因，首先當然應該歸責於科技公司不願完全揭露平台設計決策（如：演算法、政策、程序和技術功能）[16]，致使外界縱然願意了解，仍不得其門而入，只能透過終端使用者體驗（end user experience）或吹哨者（whistleblower）

---

[12] 新聞民主與平台議價論壇（2022），〈《新聞媒體與數位平臺強制議價法草案》總說明〉，https://www.twjour.org/post/ 新聞媒體與數位平臺強制議價法草案-總說明（最後瀏覽日：01/29/2024）。

[13] 中央社（08/19/2022），〈中介法業界歧見大　NCC 暫緩辦理 25 日公聽會〉，https://www.cna.com.tw/news/afe/202208190225.aspx

[14] Tibken, S. (2018, April 11). Questions to Mark Zuckerberg show many senators don't get Facebook. *CNET*. https://www.cnet.com/news/politics/some-senators-in-congress-capitol-hill-just-dont-get-facebook-and-mark-zuckerberg/

[15] 網路平台應依「使用者付費」原則，向實際產出內容的新聞媒體支付合理報酬，此一主張出現於台灣進行媒體議價機制討論的前期，但於 Meta、Google 分別辯以新聞連結係由用戶主動分享發佈、搜尋結果為著作權上合理使用，且流量最終皆導向新聞媒體網站後，各媒體公學會轉向以廣告市場失衡、健全民主機制為主要立論基礎。參見新聞民主與平台議價論壇（2023），〈2 月 15 日立法院新聞有價公聽會〉，https://www.twjour.org/post/2 月 15 日立法院新聞有價公聽會（最後瀏覽日：01/29/2024）；中央社（12/21/2022），〈媒體數位平台首輪對話落幕 學者籲新聞業統整訴求〉，https://www.cna.com.tw/news/afe/202212210150.aspx

[16] 監察委員會（2021），〈案例裁決 2021-001-FB-FBR〉， https://oversightboard.com/decision/FB-691QAMHJ/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

提供的資訊[17]，加上由 Meta 公司的「政策公開透明平台（Transparency Center）」主動發佈《社群守則執行狀況報告（Community Standards Enforcement Report）》[18]，拼湊社群媒體的運作邏輯。或許也正因此限制，當前以社群媒體為主題的中文研究，多數著眼於已發生且一般使用者得見的問題、現象，與政策相關者，譬如虛假訊息[19]、假帳號管制[20]、隱私權[21]等是。

值得慶幸者是，2020 年 5 月，Facebook 創辦人暨執行長 Mark Zuckerberg 宣布成立監察委員會（Oversight Board），旨在獨立審查 Meta 公司於旗下平台所作出的內容處置（content decision）是否得當，且符合國際人權標準[22]，此一機制或能在未來持續為研究者提供與以往不同的研究取徑。

由於監察委員會得於審查案件過程中，要求 Meta 公司額外提供在決定內容處置、執行動作（enforcement action）時的有關資訊[23]，外界得以透過監察委員會所發佈如同法院裁判書的案例裁決（case decisions）及其中利害關係人（如：監察委員會成員、外部參與者）現身說法[24]，一窺社群媒體平台如何進行個案政策適用、解釋《社群守則（Community Standards）》，以及內容審查系統的實際運作模式。

---

[17] 其中最著名者，為 2021 年 Facebook 前員工 Frances Haugen 向《華爾街日報》揭露數萬份內部翻攝文件，集結而成的調查報導《Facebook 文件（The Facebook Files）》，指稱該公司對於 Facebook、Instagram 上的仇恨言論、與 COVID-19 疫苗相關的錯誤資訊、政治極化、青少年心理健康等問題，明知而不作為。*See* The Facebook Files. (2021, September). *The Wall Street Journal*. https://www.wsj.com/articles/the-facebook-files-11631713039

[18] Meta Platforms, Inc. (2023). Community Standards Enforcement Report. *Transparency Center*. https://transparency.fb.com/reports/community-standards-enforcement/

[19] 黃銘輝（2019），〈假新聞、社群媒體與網路時代的言論自由〉，《月旦法學雜誌》，292 期，頁 5-29。

[20] 張智瑋（2023），《社群媒體假帳號之管制研究》，國立中央大學法律與政府研究所碩士論文。

[21] 劉靜怡（2012），〈社群網路時代的隱私困境:以 Facebook 為討論對象〉，《臺大法學論叢》，41 卷 1 期，頁 1-70。

[22] 陳憶寧（2021），〈社群平台的治理：從 Facebook 成立監察委員會兼談如何處理虛假資訊〉，《中華傳播學刊》，第 39 期，頁 138。

[23] Oversight Board. (2023, February). Oversight Board Bylaws Article 2, Section 2.2. https://www.oversightboard.com/sr/governance/bylaws

[24] Douek, E. (2020). "What Kind of Oversight Board Have You Given Us?". *University of Chicago Law Review Online*, 2020, 1-11.



doi:10.6342/NTU202400262

本文作者於2022年加入Meta公司[25]，經手公共政策、政府關係等業務後，經常需與政策利害關係人說明Meta公司如何治理Facebook、Instagram，並解釋《社群守則》、《廣告刊登準則》等，無可諱言，也曾經遭受許多來自公民社會合理的挑戰與質疑。

作為兩平台執行內容管理（content moderation）主要依據的《社群守則》，雖看似已針對不同類型的內容，盡可能列舉並設置標準，卻也如同法律，難免有模糊抽象、難以預測，具備高度詮釋空間者，譬如如何界定該內容為「讚揚」危險組織和人物[26]。此外，《社群守則》也未對某些可能應例外允許的特殊情況，如生理女性跨性別者進行平胸手術後，是否仍落入「包含女性乳頭的圖像」的禁止範圍[27]，進行更詳細的定義。

綜結以上，吾人應該體認，在當前實踐言論自由場域已顯著轉向社群媒體的時代，如何管理平台上的內容，已經不該單單由科技公司說的算。如同國家對於言論表達，縱然可以透過區分言論價值高低，來進行不同程度地限制，但都必須通過中度或嚴格的憲法審查標準，來檢視是否過度侵害人民權利[28]。

秉持知己知彼的態度，盡可能地了解社群媒體現有的制度、政策，才能更有效地發掘潛在問題，據以批判並提出解方。而由Meta公司捐助成立的監察委員會於近年所作出的案例裁決，恰恰提供了一個能近距離觀察言論自由在社群媒體上如何被詮釋、劃限的機會，更是公民社會應該積極關注、討論的課題，結合本文作者對於法學、傳播學的學術背景與工作上的專業經驗，本研究因此而生。

---

[25] 本文係作者個人之學術研究，並不代表任何公司、組織，所引述資料皆為合法取得，併予敘明。

[26] 監察委員會（2020），〈案例裁決 2020-005-FB-UA〉， https://oversightboard.com/decision/FB-2RDRCAVQ/（最後瀏覽日：01/29/2024）。Meta公司於該案例裁決公布後，依照監察委員會建議，進一步定義「讚揚」的樣態包含「以正面角度談論指定實體或活動」、「賦予指定實體或活動成就感」、「宣稱指定實體的仇恨、暴力或犯罪行為合法、合乎道德、正當或可以接受，藉此合理化指定實體的理念」、「在意識形態上支持指定實體或活動」等四種，並分別以案例說明，試圖具體化此一不確定概念。參見Meta Platforms, Inc.（2023），〈Facebook《社群守則》－危險組織和人物〉，https://transparency.fb.com/zh-tw/policies/community-standards/dangerous-individuals-organizations/（最後瀏覽日：01/29/2024）。

[27] 監察委員會（2020），〈案例裁決 2022-009-IG-UA and 2022-010-IG-UA〉，https://oversightboard.com/decision/BUN-IH313ZHJ/（最後瀏覽日：01/29/2024）。

[28] 林子儀（2002），〈言論自由導論〉，李鴻禧（等著），《台灣憲法之縱剖橫切》，頁103-179，元照。



doi:10.6342/NTU202400262

## 第二節　研究標的

### 壹、社群媒體平台

本研究欲探討言論自由在社群媒體上如何被界定，首先要處理的問題為研究標的，即社群媒體平台的選擇。參採各公司於 2023 年公布的月活躍用戶數據，Facebook 有逾 30 億用戶[29]，其次則為 YouTube 有逾 25 億用戶[30]，Instagram 約 20 億，WeChat、TikTok 則各有 13 億、10 億用戶，X 則僅逾 5 億用戶[31]。就規模而言，考量《社群守則》皆適用於 Meta 公司旗下的 Facebook、Instagram 兩平台[32]，相比其他社群媒體，其涵蓋的用戶更為普及，也因而更具代表性。

就平台性質而言，YouTube、TikTok 實際上皆以影音創作者的影像為主要內容，多數用戶的發言內容係透過留言方式呈現，相較 Facebook、Instagram 人人皆可發佈內容、分享、留言的互動模式為單一。亦即，如要分析規範與被規範客體之間的互動時，前二平台將無可避免地著重在討論如何管理由特定類型用戶，即影音創作者所發佈的影音內容，不若 Meta 公司旗下兩平台的規範所轄對象、內容形式全面。

又，雖然 YouTube 有《社群規範》[33]，TikTok 亦設《社群自律公約》[34]等相類於《社群守則》的內容管理依據，但並無如《社群守則》呈現歷次修訂紀錄，供查閱者比較不同版本間差異。更為關鍵者，即其他社群媒體平台皆未設有如監察委員會此類額外的審查機制，研究者既無法據此獲得更多平台運行機制的資訊，也無法從獨立、第三方的審查意見，分析外部觀點與社群媒體內容管理政策的互動過程。

---

[29] Meta Platforms, Inc., *supra* note 2.

[30] Wong, B. (2023, May 18). Top Social Media Statistics And Trends Of 2023. *Forbes*. https://www.forbes.com/advisor/business/social-media-statistics/

[31] Dixon, S. (2023, October 11). Most popular social networks worldwide as of July 2023, ranked by number of monthly active users. *Statista*. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[32] 監察委員會（2020），〈案例裁決 2020-004-IG-UA〉，https://www.oversightboard.com/decision/IG-7THR3SI1/（最後瀏覽日：01/29/2024）。

[33] YouTube，《社群規範》，https://www.youtube.com/intl/zh-TW/howyoutubeworks/policies/community-guidelines/（最後瀏覽日：01/29/2024）。

[34] TikTok，《社群自律公約》，https://www.tiktok.com/community-guidelines/zh-tw/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

此外，若從平台治理層面比較，X 於 2022 年易主、更名後，Elon Musk 解散原 Twitter 聘請專家組成的外部政策諮詢單位「信任與安全委員會（Trust and Safety Council）」[35]、移除隸屬國家、政府的媒體帳號識別標籤（government and state-affiliated media account labels）[36]、停止打擊新冠肺炎（COVID-19）錯誤資訊的內容政策等一系列減少管理的舉措，被公民社會批評「既沒有實質的公共透明化機制，也沒有任何證據顯示有效地執行政策」[37]。對外部研究者而言，此情形並不利於資料取得，也恐無分析其管理政策的研究價值。WeChat則考量其用戶地域集中，且《微信個人賬戶使用規範》又明言「嚴格遵守中華人民共和國相關法律法規」[38]，實未見有深入探討其內容管理依據的必要之處。

綜觀來說，在考量研究量能、篇幅有限等條件下，無論是依據規範普及程度，又或平台本身性質、規範呈現形式、已知的治理模式等，選擇 Meta 公司旗下社群媒體平台 Facebook、Instagram 作為本研究的主要研究標的，無疑是最為妥適、效率的作法，也更有機會為後繼研究創造參考價值。

## 貳、研究文本

本研究核心議題既為探討言論自由的界線在社群媒體上如何被劃設，在擇定以 Meta 公司旗下社群媒體平台 Facebook、Instagram 為主要研究標的後，緊接必須要討論的，即是應透過分析哪些文本資料，來系統化理解、分析有關規範，並據以與美國、台灣的法制進行比較。

---

[35] Haggin, P. (2022, December 12). Elon Musk's Twitter Disbands Trust and Safety Council. *The Wall Street Journal*. https://www.wsj.com/articles/elon-musks-twitter-disbands-trust-and-safety-council-11670898329

[36] Twitter 原先的用意在於透過此標籤，讓用戶於接觸到中國國營媒體所發佈的內容時，能夠有更多的脈絡（context），但實際上包含美國、英國、俄羅斯、法國等國的國營或政府出資的媒體，亦會被此標籤標識，具有此標籤的帳號將不會被主動系統推薦，此一作法亦見於 Facebook、Instagram、YouTube。*See* X. (n.d.). About government and state-affiliated media account labels on X. *X Help Center*. https://help.twitter.com/en/rules-and-policies/state-affiliated-china

[37] Myers, S. L., Thompson, S. A., & Hsu, T. (2023, October 28). The Consequences of Elon Musk's Ownership of X. *The New York Times*. https://www.nytimes.com/interactive/2023/10/27/technology/twitter-x-elon-musk-anniversary.html

[38] 騰訊，《微信個人賬戶使用規範》，https://weixin.qq.com/cgi-bin/readtemplate?&t=page/agreement/personal_account（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

## 一、Facebook、Instagram 內容管理政策

首先，如前所述，Facebook、Instagram 用以管理內容的主要依據，依照內容性質不同，用戶原生內容[39]（organic content）適用《社群守則》，而廣告內容（ad content）除《社群守則》外，亦需符合《廣告刊登準則（Advertising Standards）》。概括而論，《廣告刊登準則》的限制較《社群守則》為多，譬如以廣告形式推廣加密貨幣必須經過事前書面同意、廣告內容不得推廣大麻等[40]，原生內容則無此限制。

此一區別，可能源於 Meta 公司對於原生內容，得援引美國《通訊端正法（Communication Decency Act）》第 230 條（Section 230），不但使社群媒體免責於用戶在平台上所發佈的言論[41]，且鼓勵平台扮演「善良撒馬利亞人（Good Samaritan）」盡善意（good faith）防免惡意言論[42]。然而，廣告內容係透過對價及使用 Meta 公司提供的廣告工具推廣、發佈，平台亦有「共同開發者（co-developer）」的性質而無從免責，因此需進行更嚴謹的管理，近期美國聯邦第九巡迴上訴法院（United States Court of Appeals for the Ninth Circuit）也作出肯定判決[43]。

對於《社群守則》、《廣告刊登準則》等 Facebook、Instagram 主要的內容管理政策可鑽研之處甚多，容本文於後詳述。但本研究透過耙梳相關規範，無異於先建立一個鳥瞰言論自由如何在社群媒體上被界定的最佳、必要視角。

---

[39] 原生內容即 Facebook、Instagram 用戶於個人帳號（profile）、粉絲專頁（Facebook Page）、社團（Facebook Group）上，主動發佈的貼文（post）、影片、留言（comment）、Reels（連續短片）、限時動態（Stories）、分享（share）、直播（Live）等內容。原生內容不同於廣告內容，不需付費給 Meta 公司，即可獲得出現在動態消息（Feed）或其他對應版位的機會。

[40] Meta Platforms, Inc. (n.d.). Introduction to the Advertising Standards. *Transparency Center*. https://transparency.fb.com/policies/ad-standards/

[41] Chander, A. (2022). Section 230 and the International Law of Facebook. *Yale Journal of Law and Technology*, *24*, 393-420.

[42] 盧建誌（2023），〈假訊息管制與言論自由的平衡：美國網路中介責任的邊界探察與反思〉，《中華傳播學刊》，第 43 期，頁 172。*See also* Reidenberg, J. R., Debelak, J., Kovnot, J., & Miao, T. (2012, April 25). Section 230 of the Communications Decency Act: A survey of the legal literature and reform proposals. *Fordham Law School- Center on Law and Information Policy, Fordham Law Legal Studies Research Paper*, No. 2046230.

[43] Vargas v. Facebook, Inc., No. 21-16499, 2023 WL 6784359 (9th Cir. Oct. 13, 2023)



doi:10.6342/NTU202400262

## 二、監察委員會作為 Facebook、Instagram 的「最高法院」

然而，在 Meta 公司創設監察委員會後，如果本研究僅僅討論前揭幾項規範及對應的內容處置手段，則未免顯得狹隘且可惜。蓋如同通常之法學研究，除盤點、介紹既有法規外，也應當說明國家或私人如何適用該規範，或法院如何作出裁判、憲法審查機關如何進行違憲審查等。

而監察委員會引用國際人權標準，審查 Meta 公司依據《社群守則》所作出限制用戶的內容處置，似與美國、台灣的合憲性審查（constitutional review）架構皆有可比之處，使得其作出的案例裁決在社群媒體的言論自由框架中，扮演著詮釋、細緻化規範的重要角色。

此處必須先強調的是，本研究並非主張將監察委員會案例裁決所建構出的規範體系，與美國、台灣或其他國家的憲政體制進行比較。因 Meta 公司既非國家行為者（state actor），就監察委員會所審查的《社群守則》等內容政策而言，也僅屬私人企業與用戶間協議[44]，故權力分立、基本權對第三人效力（Drittwirkung der Grundrechte）等憲法原則於此皆不適用。

之所以提及美國、台灣的合憲性審查架構，乃因監察委員會的審查流程、裁決效力等與之相類，故在研究方法採擇上，本文認為得藉由常見之法學研究方法，即透過梳理判決、文獻，分析監察委員會於特定主題的立場、論理邏輯。

申言之，國家在執行合憲性審查時，以美國聯邦最高法院（Supreme Court of the United States）與我國憲法法庭為例，即是透過選案程序，對於下級法院裁判或爭議案件，依據憲法及判決先例（precedents）進行審理，如未能合乎憲法規範，則發生原判決遭推翻（reverse）（含廢棄發回）[45]，或適用之法律違憲等效果[46]。

而監察委員會亦是從成千上萬對 Meta 公司內容處置的「上訴」中選案，再以國際人權標準作為根本規範進行審查，決定維持（uphold）或推翻（overturn）用戶原先在 Facebook、Instagram 受到的處置決定，及要求 Meta 公司修正違反國際

[44] 監察委員會，前揭註 16。

[45] 蘇彥圖（2019），〈美國聯邦最高法院審理流程概論〉，發表於《司法院大法官一〇八年度學術研討會議——憲法訴訟的新時代》，司法院（主辦），臺北。

[46] 吳信華（2022），《憲法訴訟基礎十講》，頁 1-9，元照；陳陽升（2023），〈憲法法庭審查大法庭裁定之裁判效力〉，《憲政時代》，第 47 卷第 2 期，頁 216。



doi:10.6342/NTU202400262

人權標準的《社群守則》[47]。如是架構，正是 Mark Zuckerberg 參考美國聯邦最高法院的設計，因而讓監察委員會得名「Facebook 最高法院（Facebook Supreme Court）」[48]。

| 審查機關 | 美國－聯邦最高法院 | 台灣－憲法法庭 | Meta－監察委員會 |
|---|---|---|---|
| 審查案件 | 1. 州法院終局裁判<br>2. 聯邦法院下級裁判<br>3. 大使或州為當事一方之案件 | 1. 法規範及裁判<br>2. 機關爭議<br>3. 總統、副總統彈劾<br>4. 政黨違憲解散<br>5. 地方自治保障<br>6. 統一解釋法律、命令 | 1. Facebook、Instagram內容處置最終決定<br>2. 政策諮詢建議請求 |
| 審查標的 | 1. 法律<br>2. 裁判 | 1. 法律<br>2. 裁判 | 1. 《社群守則》<br>2. 政策執行 |
| 審查依據 | 1. 憲法<br>2. 聯邦最高法院判例 | 1. 憲法（大法官解釋）<br>2. 憲法法院判決 | 1. 國際人權標準<br>2. 案例裁決 |
| 未能通過產生結果 | 1. 推翻判決（含廢棄判決發回）<br>2. 宣告法規範違憲、修正要求、警告 | 1. 推翻判決（含廢棄判決發回）<br>2. 宣告法規範違憲、修正要求、警告 | 1. 推翻處置<br>2. 《社群守則》及執行政策修正要求 |
| 意見書 | 有（協同、不同） | 有（協同、不同） | 無（併入理由書） |

**表 1：美國、台灣合憲性審查與監察委員會審查架構比較**

（本表為作者自製）

既擇定以法學研究視角，剖析監察委員會所作案例裁決如何形塑社群媒體言論自由，其次則要檢驗相關文本、主題選擇是否得當。

[47] Douek, *supra* note 24.

[48] Klonick, K. (2020). The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, 129(8), 2418-2499.



doi:10.6342/NTU202400262

截至 2023 年 10 月底，監察委員會共發佈 53 則案例裁決，涉及 60 件案件[49]，另有三則政策諮詢意見。其中，僅 12 則案例裁決維持原先 Meta 公司作出的內容處置，40 則推翻，一則無法審查。

| 裁決編號 | 裁決名稱 | 結果 | 涉及平台 | 涉及政策 | 涉及主題 |
|---|---|---|---|---|---|
| — | 案例無法審查 | — | Facebook | 仇恨言論 | 政治、暴力、宗教 |
| 2020-002-FB-UA | 關於穆斯林的緬甸貼文 | 推翻 | Facebook | 仇恨言論 | 政治、宗教、暴力 |
| 2020-003-FB-UA | 亞塞拜然的亞美尼亞人 | 維持 | Facebook | 仇恨言論 | 文化、歧視、宗教 |
| 2020-004-IG-UA | 乳癌症狀和裸露 | 推翻 | Instagram | 成人裸露和性行為 | 健康、安全 |
| 2020-005-FB-UA | 納粹引述 | 推翻 | Facebook | 危險人物和組織 | 政治 |
| 2020-006-FB-FBR | 宣稱的新冠肺炎療法 | 推翻 | Facebook | 暴力和煽動仇恨 | 健康、錯誤資訊、安全 |
| 2020-007-FB-FBR | 抵制法國的印度抗議 | 推翻 | Facebook | 暴力和煽動仇恨 | 宗教、暴力 |
| 2021-001-FB-FBR | 美國前總統<br>川普帳號遭停權 | 維持 | Facebook | 危險人物和組織 | 表達自由、政治、安全 |
| 2021-002-FB-UA | 裝扮「黑彼得」 | 維持 | Facebook | 仇恨言論 | 文化、攝影、<br>兒童／兒童權利 |
| 2021-003-FB-UA | 旁遮普媒體對<br>印度 RSS 組織的擔憂 | 推翻 | Facebook | 危險人物和組織 | 政治 |
| 2021-004-FB-UA | 俄羅斯支持<br>Navalny 抗議活動 | 推翻 | Facebook | 霸凌和騷擾 | 表達自由、新聞事件、<br>政治 |
| 2021-005-FB-UA | 「兩顆按鈕」梗圖 | 推翻 | Facebook | 殘暴和麻木不仁 | 表達自由、幽默、政治 |
| 2021-006-IG-UA | 奧賈蘭的隔離監禁 | 推翻 | Instagram | 危險人物和組織 | 表達自由、偏鄉社區、<br>錯誤資訊 |
| 2021-007-FB-UA | 緬甸 Bot<br>（機器人程式） | 推翻 | Facebook | 仇恨言論 | 表達自由、政治 |
| 2021-008-FB-FBR | 巴西的新冠肺炎封城 | 維持 | Facebook | 暴力和煽動仇恨 | 政府、健康 |
| 2021-009-FB-UA | 被分享的<br>半島電視台貼文 | 推翻 | Facebook | 危險人物和組織 | 新聞事件、新聞學、<br>戰爭和衝突 |
| 2021-010-FB-UA | 哥倫比亞示威抗議 | 推翻 | Facebook | 仇恨言論 | 社區組織、表達自由、<br>陳情抗議 |
| 2021-011-FB-UA | 南非毀謗用語 | 維持 | Facebook | 仇恨言論 | 政府、偏鄉社區、政治 |

[49] 涉及案件數量較案例裁決多，係因其中五件案例裁決所審查的內容處置決定所涉及背景、內容類似，監察委員會合併審查並作出裁決。參見監察委員會，〈案例裁決和政策諮詢意見〉，https://oversightboard.com/decision/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

| | | | | | |
|---|---|---|---|---|---|
| 2021-012-FB-UA | 貝殼織帶 | 推翻 | Facebook | 仇恨言論 | 藝術／寫作／詩詞、文化、偏鄉社區 |
| 2021-013-IG-UA | 熬煮死藤水 | 推翻 | Instagram | 管制商品 | 文化活動、健康、宗教 |
| 2021-014-FB-UA | 宣稱在拉亞科博發生的犯罪 | 維持 | Facebook | 仇恨言論 | 表達自由、戰爭和衝突 |
| 2021-015-FB-UA | 開立 Adderall® | 推翻 | Facebook | 管制商品 | 歧視、健康、安全 |
| 2021-016-FB-FBR | 瑞典記者報導未成年人遭受性暴力 | 推翻 | Facebook | 成人裸露和性行為 | 兒童／兒童權利、安全 |
| 2022-001-FB-UA | Knin 卡通片 | 推翻 | Facebook | 仇恨言論 | 歧視、種族和民族、表達自由 |
| 2022-002-FB-MR | 蘇丹血腥暴力影片 | 維持 | Facebook | 血腥暴力內容 | 新聞事件、安全 |
| 2022-003-IG-UA | 要求收回使用某些阿拉伯字詞 | 推翻 | Instagram | 仇恨言論 | LGBT、偏鄉社區、性別和性別認同平等 |
| 2022-004-FB-UA | 哥倫比亞警方卡通 | 推翻 | Facebook | 危險人物和組織 | 表達自由、政府、虐待 |
| 2022-005-FB-UA | 新聞報導中提及塔利班 | 推翻 | Facebook | 危險人物和組織 | 新聞學、政治、新聞事件 |
| 2022-006-FB-MR | 提格雷通訊事務局 | 維持 | Facebook | 暴力和煽動仇恨 | 政府、暴力、戰爭和衝突 |
| 2022-007-IG-MR | 英國鑽頭音樂 | 推翻 | Instagram | 暴力和煽動仇恨 | 表達自由、藝術／寫作／詩詞、政府 |
| 2022-008-FB-UA | 俄羅斯文詩詞 | 推翻 | Facebook | 仇恨言論 | 藝術／寫作／詩詞、戰爭和衝突 |
| 2022-009-IG-UA<br>2022-010-IG-UA | 性別認同與裸露 | 推翻 | Instagram | 性誘惑 | 性別和性別認同平等、健康 |
| 2022-011-IG-UA | 奈及利亞教堂攻擊後的影片 | 推翻 | Instagram | 血腥暴力內容 | 虐待、安全、戰爭和衝突 |
| 2022-012-IG-MR | 印度性騷擾影片 | 維持 | Instagram | 成人性剝削 | 表達自由、偏鄉社區、新聞事件 |
| 2022-013-FB-UA | 伊朗抗議標語 | 推翻 | Facebook | 暴力和煽動仇恨 | 陳情抗議、政府、性別和性別認同平等 |
| 2022-014-FB-MR | 斯里蘭卡處方藥物 | 維持 | Facebook | 管制商品 | 政府、健康、安全 |
| 2023-001-FB-UA | 巴西將軍的演講 | 推翻 | Facebook | 血腥暴力內容<br>預謀傷害和宣揚犯罪<br>暴力和煽動仇恨 | 選舉、政府、陳情抗議 |



doi:10.6342/NTU202400262

| | | | | | |
|---|---|---|---|---|---|
| 2023-002-IG-UA<br>2023-005-IG-UA | Violence against women | 推翻 | Instagram | 仇恨言論 | 性別和性別認同平等、<br>表達自由 |
| 2023-003-FB-MR | 柬埔寨首相 | 推翻 | Facebook | 暴力和煽動仇恨<br>預謀傷害和宣揚犯罪 | 選舉、政治、陳情抗議 |
| 2023-004-FB-MR | 亞美尼亞戰俘影片 | 維持 | Facebook | 預謀傷害和宣揚犯罪 | 戰爭和衝突、<br>表達自由、安全 |
| 2023-006-FB-UA | 性別暴力的圖像 | 推翻 | Facebook | 霸凌和騷擾 | 性別和性別認同平等、<br>幽默、暴力 |
| 2023-007-FB-UA<br>2023-008-FB-UA<br>2023-009-IG-UA | 土耳其大選前的<br>政治爭論 | 推翻 | Facebook<br>Instagram | 仇恨言論 | 選舉、新聞學、<br>自然災害 |
| 2023-010-IG-MR | 提倡將 K 他命應用<br>在非 FDA 核准的療法 | 推翻 | Instagram | 管制商品 | 表達自由、健康 |
| 2023-011-IG-UA<br>2023-012-FB-UA<br>2023-013-FB-UA | United States posts<br>discussing abortion | 推翻 | Facebook<br>Instagram | 暴力和煽動仇恨 | 性別和性別認同平等、<br>健康、表達自由 |
| 2023-014-IG-UA | 呼籲古巴婦女抗議 | 推翻 | Instagram | 仇恨言論 | 陳情抗議、表達自由、<br>性別和性別認同平等 |
| 2023-015-FB-UA | 針對某位女性的<br>非人化言論 | 推翻 | Facebook | 仇恨言論 | 性別和性別認同平等 |
| 2023-016-FB-UA | 針對秘魯總統 Pedro<br>Castillo 的隱喻陳述 | 推翻 | Facebook | 暴力和煽動仇恨 | 政治 |
| 2023-017-FB-UA | 反殖民領袖<br>Amílcar Cabral | 推翻 | Facebook | 危險人物和組織 | 表達自由 |
| 2023-019-FB-UA<br>2023-020-FB-UA | 果汁飲食法 | 維持 | Facebook | 自殺和自殘 | 健康、兒童／兒童權利 |
| 2023-024-IG-UA | 對反猶太主義的回應 | 推翻 | Instagram | 危險人物和組織 | 表達自由 |
| 2023-025-FB-UA | 印尼執法機關貪腐情形 | 推翻 | Facebook | 暴力和煽動仇恨 | 政府 |
| 2023-026-FB-UA | 衣索比亞的旅館 | 推翻 | Facebook | 暴力和煽動仇恨 | 暴力、戰爭和衝突 |
| 2023-027-IG-UA | 黎巴嫩社會運動人士 | 推翻 | Instagram | 危險人物和組織 | 表達自由 |

**表 2：監察委員會所發佈的案例裁決列表（截至 2023 年 10 月）**

（本表為作者自製）



doi:10.6342/NTU202400262

綜觀監察委員會所作成的案例裁決，可知60件案件所涉及的國家、主題、背景、內容管理政策，乃至內容所使用的語言各不相同，甚或可以說是廣泛、多元。其中，計有40則案例裁決、43件案件出現在Facebook上，源於Instagram的案件有17件，形成案例裁決共15則。

此外，案件所涉及的《社群守則》，以仇恨言論、暴力和煽動仇恨兩類最多，各有16則案例裁決，危險人物和組織則有10則，亦見數則分別與性、血腥暴力內容、管制商品等內容管理政策相關者。

## 三、以「立即危害」與「容許差別」為中心

值得注意的是，在詳閱各篇案例裁決後，本文發現，縱然案件所涉主題、內容管理政策不同，監察委員會在一定條件情況下，多次引入「立即危害（imminent harm）」、「容許差別（allowance）」兩標準進行審查。

舉例而言，在〈案例裁決 2020-006-FB-FBR〉中，監察委員會認為，Facebook 對於與新冠肺炎相關且「可能造成人身傷害，但不具立即性的危害」的錯誤資訊（misinformation），應採取干涉度較低的處置手段[50]。同樣地，在〈案例裁決 2021-001-FB-FBR〉中，對於時任美國總統 Donald Trump 在國會暴動期間發佈可能激化衝突的貼文，監察委員會要求 Facebook 於「貼文很有可能造成立即性傷害」時，「應該採取行動，迅速執行規則」[51]。

「容許差別」於〈案例裁決 2021-002-FB-UA〉中，首次獲得解釋。監察委員會肯認，縱然仇恨言論違反《社群守則》，但若某些情況下，該內容能「帶來的公眾利益大於可能造成傷害的風險」[52]，則應該被例外地允許。相同原則也適用於「為喚起大眾關注、或紀錄人權迫害行為」[53]，但有悖於《社群守則》的血腥暴力內容。

---

[50] 監察委員會（2020），〈案例裁決 2020-006-FB-FBR〉，https://oversightboard.com/decision/FB-XWJQBU9A/（最後瀏覽日：01/29/2024）。

[51] 監察委員會，前揭註 16。

[52] 監察委員會（2021），〈案例裁決 2021-002-FB-UA〉，https://oversightboard.com/decision/FB-S6NRTDAJ/（最後瀏覽日：01/29/2024）。

[53] 監察委員會（2022），〈案例裁決 2022-002-FB-MR〉，https://oversightboard.com/decision/FB-AP0NSBVC/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

進一步耙梳，本文統整出截至 2023 年 10 月，監察委員會共有 18 則案例裁決使用「立即危害」作為審查標準，判斷該內容是否應予移除；另有 16 則案例裁決，監察委員會以「容許差別」進行衡量，評估維持顯示違規內容是否符合公共利益。

兩標準乍看之下不難理解，但究竟「立即」所指具體時間為何？「危害」須達何種程度？「公眾利益」多寡又該如何衡量？以上要件既未見於《社群守則》等內容管理政策，監察委員會也沒有在單一的案例裁決中說明具體內涵。或因如此，「立即危害」與「容許差別」也是在監察委員會所有案例裁決中，唯二多次出現，且具有延伸討論的審查標準。

本文認為，事實上「立即危害」與「容許差別」，一如法學中的不確定法律概念（unbestimmte Rechtsbegriff），僅能描述而未有明確定義，需要透過經驗、事實加以認定[54]。亦即，本文得透過比較、分析監察委員會採用「立即危害」與「容許差別」審查之案例裁決，為兩標準描繪出更為清晰的輪廓，進而理解言論自由在社群媒體上的界線與例外。

## 第三節　研究架構

行文至此，宜再次強調本文的問題意識與研究目標，即透過盤整 Meta 公司對於旗下平台 Facebook、Instagram 所制定的內容管理政策，以及分析對於內容處置決定具有終審性質、解釋權限的監察委員會所作成之案例裁決，嘗試以較全面、系統性的方式，探討社群媒體如何管理言論。

同時，考量篇幅限制、研究量能，本文將以監察委員會採用「立即危害」、「容許差別」兩項不確定概念作為審查標準的案例裁決為中心，具體化社群媒體對於言論自由在特定條件下的詮釋，並試圖比較美國、台灣近似的法律規範，期能為社會大眾對於基本權隨時代、科技變遷，開啟更多想像空間。

準此，本文研究架構分為五個主要章節。第一章即本章，以社群媒體作為當代言論自由實現的重要場域為引，簡介平台在內容管理上的現況與爭議，進而闡述本文為何選擇 Meta 公司作為主要研究標的，且如何利用其內容管理政策、機制

[54] 李惠宗（2020），《行政法要義》，頁 151-172，元照。



doi:10.6342/NTU202400262

的特殊之處，取徑法學研究方法以進行分析。

本文第二章以 Facebook、Instagram 為中心，探究目前社群媒體平台如何管理內容，包含說明《社群守則》等政策規範、平台對於用戶的內容處置手段，並討論監察委員會的運作、裁決過程及對平台治理所產生的影響，旨在綜覽社群媒體對於限制言論、落實管理所採用的策略與方法。

第三章聚焦於監察委員會採用「立即危害」作為移除內容與否的判斷標準及其具體內涵，本文將挑選出數則具指標性的案例裁決，透過文本相互對話，探討監察委員會在考量網路及社群媒體特性後，如何援引聯合國《拉巴特行動計劃（the Rabat Plan of Action）》六項門檻評估危害風險[55]，平衡用戶的言論自由與平台希望能達成的安全理念。

而後，本文將一探與之相類的「明顯而立即危險（Clear and Present Danger）」原則，回顧美國聯邦最高法院於 1969 年的 Brandenburg v. Ohio 案，設立限制言論自由須滿足該言論「旨在煽動或造成迫在眉睫且可能發生的不法行為（directed to inciting or producing imminent lawless action and is likely to incite or produce such action）」的標準[56]，並發展出主觀上之故意、急迫性、可能性三個衡量指標[57]。本文亦將於該章討論美國如何適用該標準劃設言論自由界線，據以比較國家、社群媒體對於危害言論容忍程度的異同。

第四章則側重分析「容許差別」。由於「新聞價值容許差別（newsworthy allowance）」及「政策精神容許差別（“spirit of the policy” allowance）」實皆為 Meta 公司自行創設的概念，因此本文將透過不同背景、脈絡的案例裁決，評析監察委員會如何援引國際人權標準，權衡爭議內容所伴隨的公共利益與傷害風險，為社群媒體既有的言論自由界線創造例外空間。本文認為「容許差別」的核心功能，即為言論自由與其他價值發生衝突時，社群媒體用以判斷何者應優先保障、何者應退讓的準則。

---

[55] United Nations, Human Rights Council. (2013). *Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred*. A/HRC/22/17/Add.4. https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf

[56] Brandenburg v. Ohio, 395 U.S. 444, 89 S. Ct. 1827, 23 L. Ed. 2d 430 (1969)

[57] 謝佳穎（2014），《明顯而立即危險原則之回顧與前瞻》，頁 47-55，國立臺灣大學法律學研究所碩士論文。



doi:10.6342/NTU202400262

末章則彙總各章結論，回應本研究的問題意識，並提出對 Meta 公司旗下 Facebook、Instagram 內容管理政策的觀察，以及「立即危害」、「容許差別」兩項不確定概念的定義。本文亦將說明二平台在 Meta 公司、監察委員會的治理機制下，所面臨到的限制與挑戰，以及當代言論自由在社群媒體上未來可能的發展方向。



doi:10.6342/NTU202400262

# 第二章　社群媒體如何管理內容？

包含言論自由在內的表現自由（freedom of expression）屬制度性保障之自由，國家形成制度以規制、保障人民是項權利[58]。亦即，如欲討論言論自由在特定國家法體系的具體內涵為何，必然須透過研析其制度本身，始能窺其全貌。

而所謂制度，則不侷限於憲法、法律所載文字，憲法審查機關、法院針對個案形成的判決，往往才能夠體現國家對於言論自由更為細緻的詮釋。譬如我國大法官藉釋字第 414 號解釋，將不同性質言論類型化，並肯認得對商業言論作出更多限制；抑或美國聯邦最高法院在 Chaplinsky v. State of New Hampshire 案中，揭示言論自由並非絕對的「雙階理論（two-tier theory）」，即防止、懲罰淫穢（lewd）、猥褻（obscene）、誹謗（libelous）等低價值言論亦為憲法所許[59]。

因此，本文欲分析言論自由在 Facebook、Instagram 的界線如何被劃設，則無疑需要先理解其管理內容所依循的政策，以及在 Meta 公司成立監察委員會後，該會的運作模式、對平台治理的實際影響，以及所作案例裁決如何援引國際人權標準，界定社群媒體應負的公共責任。

## 第一節　《社群守則》等內容政策

Facebook 原型為 Mark Zuckerberg 於 2004 年架設、提供美國 Harvard University 學生搜尋同學、查找校園人際網絡的 Thefacebook。在 2005 年開放公眾註冊之前，Thefacebook 僅有一位全職人員負責人工審查內容，直至出現一個「充滿死屍和侮辱同志」的社團，該公司才決定建立一個類似普通法（common law）、臚列各種應禁止內容的文件，作為團隊決定是否移除爭議內容的指引[60]。

時至今日，Meta 公司服務已遍及全球，在 Facebook、Instagram 上的內容以數十種語言呈現[61]，不斷增修、以《社群守則》為首的內容政策，便形成規範何種言論可以出現、留存於二社群媒體平台上的主要原則。

---

[58] 李惠宗（2019），《憲法要義》，頁 203-224，元照。

[59] Chaplinsky v. State of New Hampshire, 315 U.S. 568, 62 S. Ct. 766, 86 L. Ed. 1031 (1942)

[60] Levy, S.，前揭註 4，頁 268-277。

[61] Meta Platforms, Inc. (n.d.). Facebook Community Standards. *Transparency Center*. https://transparency.fb.com/policies/community-standards/



doi:10.6342/NTU202400262

## 壹、《社群守則》

原生內容[62]作為 Facebook、Instagram 上最大宗的內容型態，其須依循的《社群守則》可說是形塑社群媒體言論自由最關鍵的角色。Meta 公司主張，制定《社群守則》的目標，乃是在保障用戶「言論自由（Expression）」的同時，也能維護該公司重視的「真實無欺（Authenticity）」、「用戶安全（Safety）」、「確保隱私（Privacy）」、「尊嚴無損（Dignity）」四項核心價值[63]。

植基於此，《社群守則》被分類為六章，包含「暴力和犯罪行為」、「用戶安全」、「令人反感的內容」、「誠信和真實性」、「尊重智慧財產權」、「內容相關要求和處置」，各章下則依主題分為若干節，共計 24 節，譬如「令人反感的內容」中，即有「仇恨言論」、「血腥暴力內容」、「成人裸露和性行為」、「性誘惑」四節。

再以「仇恨言論」一節為例，從中可知 Meta 公司於《社群守則》中定義為「針對他人（而非概念或機構）受保護的特徵進行直接攻擊，包括種族、民族、國籍、身心障礙、宗教信仰、種姓、性傾向、性別、性別認同和重大疾病」[64]，並以正面表列的形式，列舉何謂「攻擊」的樣態，舉凡以非人化、貶低、表達厭惡、咒罵言論對待受保護特徵他人等是。

值得注意的是，除了抽象的原則規範，《社群守則》亦具體舉例該當要件的禁止內容，甚至從語言版本的差異中，可以發現其試圖融入當地背景進行調整。譬如咒罵行為的定義包含「以生殖器或肛門稱呼目標對象」，在英文版《社群守則》舉例為「asshole（肛門）」，香港繁體中文版亦直譯為「屁眼」，但在台灣繁體中文版《社群守則》中，該定義則舉例台灣較為常見的「雞掰人」。反觀日文版或考量當地並類似意義的粗鄙用語，並未翻譯前述舉例內容，而逕以英文呈現[65]。

---

[62] 參見前揭註 39。

[63] Meta Platforms, Inc., *supra* note 61.

[64] Meta Platforms, Inc. (2023). Hate Speech. *Transparency Center*. https://transparency.fb.com/policies/community-standards/hate-speech/

[65] *Id*.



doi:10.6342/NTU202400262

《社群守則》除表列禁止內容的類型及具體案例，於多節政策中，亦可見雖爭議內容該當禁止要件，但 Meta 公司認為「需要提供額外資訊或背景資料」[66]，始能判斷如何進行處置。譬如「成人裸露和性行為」一節中，雖明定用戶不得發佈含有成年人性器官及女性乳頭的圖片，但如發佈內容背景涉及分娩、疾病預防，Facebook、Instagram 並不會直接移除該則內容，而是會加上「敏感內容（sensitive content）」警告[67]，用戶須點擊提示選項後始能查看，符合特定條件的血腥暴力內容亦採取相同舉措[68]。

除從語言版本、規範程度角度觀察外，如前所述[69]，《社群守則》與其他平台所制定之內容規範最大區別，在於 Meta 公司也一併公開歷年來的變更紀錄，有心研究者因而得以透過比較不同時期的《社群守則》差異，分析其內容管理政策的演進歷程。以「管制商品和服務」一節為例，於 2018 年 8 月時，僅見兩平台禁止販售、交易非醫療藥物（non-medical drugs）、槍枝及其零件兩類常見的管制商品[70]；然而，時至 2023 年，已細分到瀕危物種、活體動物、人血、減重商品、歷史文物等，其中，對於比較罕見的宗教致幻劑（entheogens），更進一步區分為禁止購買、交易，但允許年滿 18 歲的成年用戶瀏覽承認或提倡使用有關物質的內容[71]。

雖然 Meta 公司並未言明《社群守則》各節政策歷次修改的原因，但對於增列宗教致幻劑，本文從監察委員會作成的〈案例裁決 2021-013-IG-UA〉見得端倪。在該則案例裁決中，監察委員會認為，可能造成精神刺激的「死藤水（Ayahuasca）」雖可被認定為非醫療藥物，但考量其在南美洲原住民文化中，有用於傳統宗教、儀式的背景，應允許「與可能造成傷害並無直接立即關聯（no direct and immediate connection between the content, …, and the possibility of harm）」

[66] Meta Platforms, Inc., *supra* note 61.
[67] Facebook. (n.d.). Why am I seeing a warning before I can view a photo or video on Facebook? *Help Center*. https://www.facebook.com/help/814083248683500
[68] Meta Platforms, Inc. (2023). Violent and Graphic Content. *Transparency Center*. https://transparency.fb.com/policies/community-standards/violent-graphic-content/
[69] 參見本文第一章第二節第壹項。
[70] Meta Platforms, Inc. (2018). Restricted Goods and Services. *Transparency Center*. https://transparency.fb.com/policies/community-standards/restricted-goods-services/
[71] *Id*.


doi:10.6342/NTU202400262

的內容，並建議 Meta 公司修改內容政策[72]。後者在「政策公開透明平台」的「監察委員會建議（Oversight Board recommendations）」頁面中，亦表示〈案例裁決 2021-013-IG-UA〉的政策建議已獲「全面實施（implementing fully）」[73]，亦可印證此一推論。

## 貳、《廣告刊登準則》與其他內容政策

除《社群守則》外，Meta 公司亦針對特定形式（如：廣告）、特殊時期（如：新冠肺炎、俄烏戰爭）、特定主題（如：政治、社會性議題）、使用特定功能（如：Facebook 粉絲專頁、社團、AI 廣告工具）的內容，制定特別、相對應的內容政策。

就規範付費廣告內容的《廣告刊登準則》而言，與原生內容最大不同者，在於 Meta 公司會於廣告刊登前，以自動化系統進行事前審查[74]，確認該廣告內容符合規範後始得發佈。此外，對於出現在 Facebook、Instagram 的廣告內容本身，《廣告刊登準則》也設下較原生內容為嚴格的限制。除前文舉例的大麻、加密貨幣[75]，對於已經第三方事實查核機構驗證的錯誤資訊，若為用戶原生內容，雖可能遭調降其在動態消息的出現排序，並被加註該內容為錯誤資訊的標籤[76]，但仍不致被移除；然而，相同內容若為廣告形式，則無法通過事前審查而自始不能發佈，或於被系統偵測後，因違反《廣告刊登準則》而被移除[77]。

在特殊時期，亦可見 Meta 公司對於《社群守則》進行擴張或特定解釋，譬如在新冠肺炎疫情流行期間，將「預謀、描述、承認或宣傳由您或您的同伴主動且刻意散播傳染性疾病」[78]及「明確宣稱民眾將因為參與投票感染新冠肺炎（或其

[72] 監察委員會（2021），〈案例裁決 2021-013-IG-UA〉， https://oversightboard.com/decision/IG-0U6FLA5B/（最後瀏覽日：01/29/2024）。

[73] Meta Platforms, Inc. (2023). Oversight Board recommendations. *Transparency Center*. https://transparency.fb.com/zh-tw/oversight/oversight-board-recommendations/

[74] Meta Platforms, Inc., s*upra* note 40.

[75] 參見本文第一章第二節第貳項。

[76] Meta Platforms, Inc. (n.d.) About fact-checking on Facebook and Instagram. *Business Help Center*. https://www.facebook.com/business/help/2593586717571940?id=673052479947730

[77] Meta Platforms, Inc., s*upra* note 40.

[78] "Coordinating, depicting, admitting to or promoting active and deliberate spread of communicable diseases by you or your associates." Meta Platforms, Inc. (2022). Coordinating Harm and Promoting Crime. *Transparency Center*. https://transparency.fb.com/policies/community-standards/coordinating-harm-



doi:10.6342/NTU202400262

他傳染性疾病）」[79]的內容，認定為「預謀傷害和宣揚犯罪」一節中的「對他人造成傷害」而被禁止。

值得留意的是，前揭二限制在 2022 年 4 月 24 日的修正版本中皆被移除，也就是說，目前若有用戶發佈上述原先禁止的內容，則可能可以被允許存在於二平台上。本文認為，此一政策變更或可說明社群媒體在評估爭議言論帶來的的風險、傷害時，會將現實世界的環境變化納入考量，以及其管理政策未必僅有朝向收緊、縮限的趨勢，亦有可能放寬、解禁原先限制的內容。

針對新冠肺炎相關的錯誤訊息，Meta 公司也隨著疫情流行程度調整政策。例如在 2023 年 5 月前，因新冠肺炎仍為世界衛生組織認定的「國際關注的公共衛生緊急事件（Public Health Emergency of International Concern）」[80]，公共衛生機關對錯誤資訊判斷是否「不實」且「可能導致立即的人身傷害」，即為 Facebook、Instagram 移除與否的標準。更有甚者，兩平台對於「系統預測可能違反，尚未確定違反本新冠肺炎錯誤資訊政策」的內容，即祭出縮減散布範圍（distribution）的處置[81]。

隨著新冠肺炎的急迫、緊張情況不再，Meta 公司發佈的〈新冠肺炎政策更新及保護措施〉即不復見上述「未審先判」的作法。縱然在少數還有疫情管控的國家中，亦要求須經當地公共衛生機關循「不實」且「可能導致立即的人身傷害」標準，判定確為錯誤資訊後，始得移除該爭議內容[82]。

除新冠肺炎外，Meta 公司也會針對個別國家、地區發生重大事件期間，推出適用於當地的內容政策，包含因應俄烏戰爭標註俄羅斯國營媒體帳號、設立 24 小時的特別行動中心（special operations center），駐有俄語、烏克蘭語為母語的內

---

promoting-crime/

[79] "Explicit claims that people will be infected by COVID-19 (or another communicable disease) if they participate in the voting process." *Id*.

[80] World Health Organization. (2023, May 5). Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic

[81] 林彥廷（2022），〈疫情中的社群平台——以 Facebook 的內容政策為例〉，發表於《新興科技應用之法治對策期末研討會》，國立臺灣大學法律學院（主辦），臺北。

[82] Facebook. (n.d.). COVID-19 and Vaccine Policy Updates & Protections? *Help Center*. https://www.facebook.com/help/230764881494641



doi:10.6342/NTU202400262

容審查人員[83]；在美國國會暴動期間，任何呼籲在華盛頓哥倫比亞特區進行抗議的言論，縱為號召和平集會[84]，也被二平台禁止。至於本研究撰文當下尚未止歇的以哈戰爭（Israel-Hamas War），由於哈瑪斯仍挾有人質，Meta公司也暫時擴張解釋《社群守則》，稱在人質安全作為最優先考量下，即使內容目的是為了譴責哈瑪斯所為，只要使得人質身分能被識別，一律會遭到禁止、移除[85]。

對於牽涉社會議題、選舉或政治（social issues, elections or politics）的廣告內容，Meta 公司則稱，基於維護選舉公正（election integrity）的目的，若要針對包含台灣在內的世界上多數國家、地區投放時，須額外提前進行個人、單位身分驗證的授權程序，以確定廣告主居住於該國家、地區內，並簽署免責聲明後，才能將該廣告送交審查[86]。有論者認為，此一措施是為了回應外界批評 Facebook、Instagram 充斥誤導、不實的政治廣告，並減少特定國家、勢力干預，影響各國選舉的可能性[87]。

又，因應人工智慧（artificial intelligence）漸成熟普及，除《社群守則》既有對深度偽造（deepfake）技術變造影音的限制外[88]，Meta 公司進一步要求社會議題、選舉或政治廣告如有使用任何形式的數位變造（digital alteration），廣告主必須額外聲明，使看到廣告的用戶收到提醒[89]。而 Meta 公司所開發的生成式人工智慧（generative AI）廣告工具，則直接禁止健康、金融服務、就業服務、社會議題、選舉、政治等與公共利益高度相關領域的廣告主使用[90]。

---

[83] Meta Platforms, Inc. (2022). Meta's Ongoing Efforts Regarding Russia's Invasion of Ukraine. https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/

[84] Meta Platforms, Inc. (2021). Our Response to the Violence in Washington. https://about.fb.com/news/2021/01/responding-to-the-violence-in-washington-dc/

[85] Meta Platforms, Inc. (2023). Meta's Ongoing Efforts Regarding the Israel-Hamas War. https://about.fb.com/news/2023/10/metas-efforts-regarding-israel-hamas-war/

[86] 陳澍、陳奕儒、Aidan Hoy（2020），《捍衛台灣選舉公正》，頁 8，Facebook。

[87] BBC. (2019, November 2). General Election 2019: Facebook takes down 'political' advertising. *BBC News*. https://www.bbc.com/news/uk-50270727

[88] 使用深偽或其他技術變造的影音內容，並不一定會在 Facebook、Instagram 被移除，仍須滿足《社群守則》中規定「透過一般人難以察覺的方式進行編輯或合成」且「可能會誤導一般人相信影片當事人說了沒說過的話（以假亂真）」的要件，才會遭到處置。*See* Meta Platforms, Inc. (2023). Misinformation. *Transparency Center*. https://transparency.fb.com/policies/community-standards/misinformation/

[89] BBC. (2023, November 9). Meta requires political advertisers to mark when deepfakes used. *BBC News*. https://www.bbc.com/news/technology-67366311

[90] Paul, K. (2023, November 6). Meta bars political advertisers from using generative AI ads tools. *Reuters*. https://www.reuters.com/technology/meta-bar-political-advertisers-using-generative-ai-ads-tools-2023-



doi:10.6342/NTU202400262

## 第二節　政策執行

Meta 公司根據前述包括《社群守則》在內的內容政策，對出現於 Facebook、Instagram 上且違反政策的內容作出處置，即政策執行（policy enforcement）。依照其發動原因，政策執行可分為被動（reactive）及主動（proactive）兩種情境；而根據其執行時點，又可分為事前（ex ante）及事後（ex post）兩類[91]。

### 壹、被動或主動政策執行

無論政策執行是被動或主動發動，考量兩平台上每天新增的內容數量以數十億則[92]計算，內容審查工作其實都與機器學習的人工智慧模型、演算法密切相關，其區別在於可能違反政策的內容是否由系統自動識別、偵測而發現，或 Meta 公司人員有額外提供指令，要求系統執行調查。

Meta 公司聲稱，超過九成的違規內容都是由系統主動偵測並作出處置[93]，然而，在其系統並未開源（open source）的情形下，吾人只能透過各類型公開文件透露的資訊一窺究竟。大抵來說，此一途徑的流程為審查系統先將 Facebook、Instagram 上的已發佈內容依據「嚴重性（severity）」、「傳播性（virality）」、「違規可能性（likelihood of violating）」三指標進行綜合排序[94]，排序越靠前者，將由系統針對其所包含的文字、圖片、影音進行實質審查，確認該內容是否有違反《社群守則》等政策，再進行處置。經此途徑審查的案件，通常不會經過人工審查程序，亦不論是否為近期發佈的內容，因而時有聽聞用戶抱怨、在台灣被謔稱為「被祖（克柏刪除內容）」[95]者，通常就是此類被系統偵測、判斷違反

11-06/

[91] *See* Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, 131(6), 1598-1670.

[92] Meta Platforms, Inc. (2023). How technology detects violations. *Transparency Center*. https://transparency.fb.com/zh-tw/enforcement/detecting-violations/technology-detects-violations/

[93] *Id*.

[94] Meta 公司並未公開系統衡量此三指標的具體方式，但說明「嚴重性」代表「內容在網路及現實世界造成傷害的可能性多大？」；「傳播性」代表「內容分享的速度有多快？」；「違規可能性」代表「問題內容確實違反政策的可能性多大？」*See* Meta Platforms, Inc. (2022). How Meta prioritizes content for review. *Transparency Center*. https://transparency.fb.com/policies/improving/prioritizing-content-review/

[95] 「被祖」為台灣用戶以諧音該內容在 Facebook、Instagram 上「被阻」，也取 Mark Zuckerburg 姓氏的中譯「祖克柏」首字雙關，嘲諷該審查決定可能是由 Mark Zuckerburg 作出的。參見遠見雜誌（05/09/2021），〈你今天「被祖」了嗎？臉書文字獄引眾怒，揭曉背後審查模式〉，



doi:10.6342/NTU202400262

《社群守則》，而被限制、移除的情形。

雖然批評者眾，本文也欲指出，在現實世界人與人交流過程中，都難免因為「說者無心，聽者有意」或不同語言的意境、文化脈絡產生誤解，更何況要求機器堪當判讀社群媒體上常見的諷刺、玩笑是否確屬仇恨、騷擾言論？當然，本文無意為科技巨頭開脫其應善加治理平台的責任，但畢竟全數交由人工審查或不會誤判的審查系統，兩作法皆不現實且過於極端，如何加速改善現有的人工智慧模型，並確保各地區、種族、語言、宗教、文化脈絡的資料庫皆能被充分訓練，再輔以人工審查作為困難案件的再審、驗證機制，或許是未來更值得期待的解決方式。

除審查系統主動偵測外，主動進行政策執行的途徑亦有可能透過Meta公司人為啟動，譬如針對極端或恐怖主義內容，即會有專門團隊透過關鍵字、發佈地區、語言等訊號進行規模化排查、審核，再加以移除[96]。但本文也發現，過往文獻似乎鮮有提及人為啟動調查的具體作法。

事實上，Meta 公司在遭遇重要事件，如前文提及的俄烏戰爭、美國國會暴動、各國選舉期間[97]，或廣受關注的惡意使用、威脅，如中國、俄羅斯認知作戰（covert influence operations）等時[98]，都會以其定義的「協同性造假行為（Coordinated Inauthentic Behavior）」為判斷標準，從「行為者（actor）」、「行為（behavior）」、「內容（content）」三方面評估爭議帳號、內容是否應予移除[99]。亦即，當Meta公司主動在Facebook、Instagram二平台上，以該事件、威脅相關的訊號排查時，會透過IP位址、登入裝置、朋友名單等數據，認定可疑帳號是否互有關聯或具相同隸屬；是否在短時間內，發送大量相同內容的訊息、留言，或有非一般使用者的使用行為；或各帳號所傳送、發佈的內容是否違反《社群守則》等[100]。

---

https://www.gvm.com.tw/article/79423

[96] Klonick, *supra* note 48, at 2430.

[97] Clegg, N. (2023). How Meta Is Planning for Elections in 2024. *Meta*. https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024/

[98] Meta Platforms, Inc. (2023). Raising Online Defenses Through Transparency and Collaboration. https://about.fb.com/news/2023/08/raising-online-defenses/

[99] Meta Platforms, Inc. (2018). Coordinated Inauthentic Behavior Explained. https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/

[100] 此處所舉各例，乃研究者結合過往自身經驗彙整而成，通常個案處理並無絕對標準，而是在綜



doi:10.6342/NTU202400262

過往人為啟動政策執行的案例，包含Meta公司移除逾八千個被認定以「協同性造假行為」散播錯誤消息、為中國進行認知作戰的 Facebook 帳號、粉絲專頁[101]；在台灣，亦有發生Meta公司獲報有多個使用繁體中文的Facebook社團以Risu為名，轉傳未經同意散布私密影像（non-consensual intimate image）的外部連結，雖此類社團原先未經審查系統主動偵測、發現違反《社群守則》[102]，但經團隊通報、調查後，仍將有關社團全面移除。

對多數使用者而言，應該較為熟悉被動的政策執行途徑，即是在使用者發現可能違反《社群守則》等內容政策的情形後，使用檢舉（report）功能向 Meta 公司反映，並要求其審查、移除該帳號或內容。然而，用戶檢舉後的流程實更為繁瑣、難以一概而論，部分被檢舉內容會交由審查系統進行，卻也有部分案件進入 Meta 公司原不曾公開承認、直至監察委員會揭露的「交叉檢查（cross check）」機制[103]，也就是當高知名度、敏感帳號被審查系統判定有違反《社群守則》情形時，並不會立即對該帳號、內容作出處置，而是將該案件移交人工審查，降低因機器誤判而導致錯誤執行政策的風險[104]。

此外，「交叉檢查」又可分為「一般後續審查（General Secondary Review）」、「初步反應後續審查（Early Response Secondary Review）」兩種類別[105]，前者係經由系統根據該內容的傳播力、瀏覽次數等指標進行排序，以決定是否需送交外包審查人員進行審查；後者則直接依據Meta公司的內部名單，待審查案件若是源於名單內的帳號、粉絲專頁，都將移交人工審查[106]。

---

合考量各方面上的違反程度，再由有關團隊作出處置決定。參見前揭註 86。

[101] 中央社（08/29/2023），〈Facebook 抗中國大外宣垃圾訊息戰 刪除逾 7700 個帳號〉，https://www.cna.com.tw/news/aopl/202308290429.aspx

[102] 中央社（06/06/2023），〈Risu 臉書社團洩私密影像 Meta：已移除違反政策社團〉，https://www.cna.com.tw/news/asoc/202306060325.aspx

[103] 監察委員會，前揭註 16。

[104] Meta Platforms, Inc. (2023). Reviewing high-impact content accurately via our cross-check system. *Transparency Center*. https://transparency.fb.com/enforcement/detecting-violations/reviewing-high-visibility-content-accurately/

[105] *Id.*

[106] 監察委員會（2023），〈案例裁決 2023-007-FB-UA, 2023-008-FB-UA, 2023-009-IG-UA〉，https://oversightboard.com/decision/FB-T8JDDDJV/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

## 貳、事前或事後政策執行

相較區辨主動或被動的政策執行，如何判斷政策執行為事前或事後，則顯得相對明確、易於理解，其關鍵時點即是用戶是否成功經由 Facebook、Instagram 發佈內容或傳送訊息。如前所述，由於 Meta 公司審查機制係將已發佈的內容進行排序後，再將系統認定有違反《社群守則》之虞的內容交予審查，因而絕大部分引用《社群守則》進行審查並作出處置的案件，都屬事後的政策執行。

Facebook、Instagram 既不會在用戶撰寫文字的同時進行審查，如何運用事前的政策執行防堵有害內容，達成最有效的遏阻手段？相關研究指出，在使用者等待上傳圖片、影片的過程中，系統其實正在同步針對特定訊號進行排查，譬如違反《社群守則》的血腥、裸露等，藉此避免如兒童色情的內容一經發佈，Meta 公司可能會連帶負擔美國聯邦法律的刑責[107]。

除此之外，Meta 公司亦有開發一個稱作「黑洞（blackhole）」的功能，即針對被判定違反《社群守則》的網域，在兩平台上進行全面移除並事前封鎖有關內容，其來源包含 Google Safe Browsing 資料庫認定為可疑、存在威脅的網站[108]。過往台灣也有前立委陳柏惟突遭 Facebook 系統通知，稱其個人網域「3qi.tw」違反《社群守則》，致使其過往所有包含該網址的競選、募款貼文皆被移除[109]，而欲新發佈包含該網域在內的貼文時，也會顯現「你無法分享此連結」、「無法分享你的貼文」等語而發佈失敗[110]，即是 Meta 公司以「黑洞」功能，事前阻斷特定內容的發佈可能性。

## 參、政策執行作法

判定內容違反《社群守則》之後，Meta 公司進而實施的處置手段，可被歸類為移除、縮減、告知三類[111]。移除的範圍包含違規內容本身、發佈內容的帳號、

---

[107] Klonick, *supra* note 91, at 1636.

[108] Facebook. (n.d.). How to check if a link I'm viewing on Facebook is secure. *Help Center*. https://www.facebook.com/help/713015355575644

[109] 鏡週刊（11/28/2023），〈陳柏惟臉書被抄家「200 多篇文消失」 疑被判定色情相關！帳號火速歸隊〉，https://www.mnews.tw/story/20231128nm018

[110] 陳柏惟（11/27/2023），〈3qi.cn 陳柏惟 成功發出？〉，Facebook，https://www.facebook.com/wonda.chen/posts/pfbid0sbXxhWTBtGn9NJDB1UB1GvYXmzrUsKv2Mhzm42zMnYiQiHjzxy5UwALUN62uzKHpl

[111] Meta Platforms, Inc. (n.d.). Taking action. *Transparency Center*.



doi:10.6342/NTU202400262

粉絲專頁，甚或多次出現違規內容的社團；縮減則可理解為台灣慣稱的「降低觸及」，即減少該爭議內容（如：經第三方事實查核機構查證為不實的資訊）[112]出現在用戶動態消息上的機會，並下降其出現的排序；所謂告知，則包含前文提及的「敏感內容」警告[113]、錯誤資訊標籤等，讓使用者於閱覽時，有更多關於該內容的脈絡、資訊，並避免不適，提升整體使用體驗。

申言之，移除的作法其實未必皆是將違規內容在 Facebook、Instagram 上全面移除，蓋因部分內容雖無違反《社群守則》，卻未見得容於當地法律規範。又各國法令限制不一、時有修正，Meta 公司遂與各國政府機關合作，建立「消費者政策管道（Consumer Policy Channel）」，或我國政府稱呼的「綠色通道」[114]，使部分涉及行政專業（如食品、藥物管理）且與政治關聯度較低的政策執行，以當地權責機關的認定為主要依歸，在該國境內將違反當地法令的內容實施「地區限制（geo-blocking）」。也就是說，在 Meta 公司的審查系統尚未判定該內容是否亦違反《社群守則》之前，他國用戶仍然可見，藉此作為合乎當地法規阻絕爭議內容，又避免各國政府過度干預言論自由的權衡舉措。

## 第三節　監察委員會

與其他社群媒體相比，Facebook、Instagram 特有之處，即是自 2020 年監察委員會成立後，為當前唯二號稱透過外部獨立機構，公開解釋、監督內容管理政策的平台。本節將回顧監察委員會的成立背景、運作模式，並分析其於審查個別內容處置決定時，所引據的國際人權標準為何。

### 壹、成立起源

2021 年 10 月 28 日，Facebook 創辦人暨執行長 Mark Zuckerberg 宣布將公司名稱改為 Meta Platforms, Inc.，宣示重塑企業品牌，並將朝向成為「元宇宙

https://transparency.fb.com/enforcement/taking-action/

[112] Meta Platforms, Inc. (2022). Penalties for sharing fact-checked content. *Transparency Center*. https://transparency.fb.com/enforcement/taking-action/penalties-for-sharing-fact-checked-content/

[113] 參見本章第一節第壹項。

[114] 中央社（05/30/2023），〈黃仁勳遭冒名詐騙 陳建仁：感謝公私協力迅速下架〉，https://www.cna.com.tw/news/aipl/202305300111.aspx



doi:10.6342/NTU202400262

（Metaverse）社群公司」發展。有論者認為，此舉是為了讓Facebook擺脫過往用戶資料外洩、造成美國等地的政治動盪的惡名[115]。

然而，在此之前，Facebook早已多次因為內容管理政策失當，而飽受外界批評。如2018年，路透社將該公司稱為「Hatebook」，因其對於平台上充斥歧視羅興亞人，進而引發種族衝突的仇恨言論置若罔聞[116]。同年，聯合國特別調查員亦公開批評Facebook對於歧視言論的縱容，「煽動大量暴力與仇恨」[117]。

各界對於網路內容管理的關注，促使聯合國於 2019 年發佈《促進和保護意見和表達自由權特別報告》，指出在網路、社群平台上限制言論，可能導致公眾沈默，又或一但無法處理煽動行為時，將助長線下暴力和歧視[118]，並進一步主張「所有資訊和通訊技術行業的公司，都應該適用聯合國《工商企業與人權指導原則》，並通過設計將人權預設納入其產品」[119]。

為了回應國際社會的譴責，Mark Zuckerberg 公開承諾設置一個獨立的問責機制，來監督該公司在 Facebook、Instagram 上的內容管理政策[120]，並於 2020 年接續宣布捐款 1.3 億美金成立信託，用以籌組、運作監察委員會。其並稱，期待監察委員會作為對 Facebook、Instagram 兩平台內容處置決定的最終裁決單位，能夠讓社群媒體的內容管理更符合國際人權標準下的言論自由[121]。

---

[115] Kelly, S. (2021, October 28). Facebook changes its company name to Meta. *CNN*. https://www.cnn.com/2021/10/28/tech/facebook-mark-zuckerberg-keynote-announcements/

[116] Stecklow, S. (2018, August 15). Why Facebook is losing the war on hate speech in Myanmar. *Reuters*. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

[117] BBC. (2018, March 13), UN: Facebook has turned into a beast in Myanmar. https://www.bbc.com/news/technology-43385677

[118] United Nations, Human Rights Council. (2020). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/44/49/Add.2, at 11. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G20/190/89/PDF/G2019089.pdf

[119] *Id*. At 14.

[120] Fung, B. (2020, October 22). Facebook's Oversight Board is finally hearing cases, two years after it was first announced. *CNN*. https://www.cnn.com/2020/10/22/tech/facebook-oversight-board/index.html

[121] Gilbert, B. (2020, February 16). Facebook is spending $130 million to create a 'Supreme Court' that can overrule Mark Zuckerberg — here's everything we know about it. *Insiders*. https://www.businessinsider.com/facebook-moderation-independent-oversight-board-supreme-court-mark-zuckerberg-explained-2020-2



doi:10.6342/NTU202400262

## 貳、運作模式

依據監察委員會《規章制度和行為規範（Oversight Board Bylaws）》，該會在理想情況下由 40 位委員組成，每位成員任期三年，總任期不得超過九年，由信託董事會獨立任命，但須經 Meta 公司推派或公眾推薦[122]。據美國媒體 The New Yorker 報導，每位委員每年可獲得十萬以上美元薪資，由前述信託支應，此一節亦使監察委員會獨立性受外界質疑[123]。目前，監察委員會組成包含法律學者、傳播學者、NGO 工作者、諾貝爾和平獎得主、退休政治領袖等，台灣則有國立政治大學傳播學院陳憶寧教授參與其中，為唯一一位中文母語人士。

作為「Facebook 最高法院」，監察委員會如同多數國家的憲法審查機關，要求當事人應先窮盡救濟手段（exhaust accessible remedies），亦即當用戶受 Facebook、Instagram 的內容處置影響後，應先窮盡 Meta 公司的內部申訴程序[124]，如仍不同意該公司重新審查後所作出的決定，才能在收到決定後 15 日內向監察委員會提出申訴[125]。

除用戶提出申訴外，監察委員會的另一種案件來源，是由 Meta 公司提出的政策諮詢，範圍不侷限於爭議內容能被允許或應移除[126]。此目的係讓該公司有機會在面臨困難個案時，能主動確保相關舉措符合國際人權標準，不致造成言論自由的不當限縮。具體個案如〈案例裁決 2021-008-FB-FBR〉，該案例中一家巴西公共醫療機構發佈了一則批評當地防制新冠肺炎政策的貼文，但錯誤引用了國際衛生組織（WHO）官員的發言，謬稱「封城不但無法拯救生命，還會導致窮人潦倒。」由於該則言論在公共衛生緊急情況下，極難判定是否應援引惡意錯誤資訊（disinformation）政策而施行移除，Facebook 將該個案移送至監察委員會請求作出政策指導[127]。

---

[122] 然而，研究者注意到，2021 年以前的《規章制度和行為規範》為規定監察委員會應由 11 到 40 位委員組成，且每位委員至多連任三屆。Oversight Board, *supra* note 23. Article 1, Section 1.4.

[123] Klonick, K. (2021, February 12), Inside the Making of Facebook's Supreme Court. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court

[124] Wong, D. & Floridi, L. (2023). Meta's Oversight Board: A Review and Critical Assessment. *Minds and Machines*, 33, 261-284.

[125] Oversight Board, *supra* note 23. Article 3, Section 1.1.

[126] *Id*. Article 2, Section 2.1.

[127] 監察委員會（2021），〈案例裁決 2021-008-FB-FBR〉，https://oversightboard.com/decision/FB-B6NGYREK/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

由用戶或 Meta 公司提出申訴或諮詢請求後，監察委員會具有選擇審查案件的裁量權（discretion），其標準為「可能影響未來政策制定且與全球用戶相關的困難案例」[128]，或「重要」且「困難」。成員指出，監察委員會其實會先依靠幕僚團隊從數以萬計的申訴案件中，挑出數十件可能符合前述標準的案件，再由四位委員組成挑選案件小組（Case Selection Committee）[129]。然而，可議者係，並無任何規範要求監察委員會公布所收受的申訴案件總數，以及案件未獲審查的原因，因而外界無法據此了解上述挑選標準是否被落實，以及該制度的執行效率。

而後，五位委員將組成一組，並確保一定有來自該案例所在、影響區域的委員，再以多數決方式審查挑選出的案例，決定維持、推翻原先Meta公司所作出的內容處置。至於對案例裁決的協同、不同意見書，則不若美國、台灣的憲法審查方式獨立呈現，而融合在裁決理由書中。此外，在特定情形，如牽涉少數族群（原住民、LGBTQ+）時，監察委員會則會透過蒐集公眾意見、委託外部語言、文化機構提供當地意見，來試圖彌補委員背景所導致的侷限。

綜觀監察委員會所作出的裁決，不外乎是維持、推翻 Facebook、Instagram 的內容處置，即要求Meta公司應恢復原先在兩平台移除的內容，或肯認其作法符合國際人權標準。而如原處置遭推翻，Meta 公司必須要在監察委員會公布裁決後七天內依其意旨執行決議[130]，惟監察委員會對《社群守則》等內容政策所提出的改善建議，則無上述的強制拘束力。

學者亦有批評，在此制度下，監察委員會的管轄權（jurisdiction）僅限於 Meta 公司有作為的情形，也就是發生違規內容遭移除，或對適用《社群守則》有疑義而主動移送案件時。然而，在其放任、消極或不作為的情況，譬如對於仇恨言論未及時作出處置，因無人能透過 Facebook、Instagram 申訴功能要求再行審查，自然也無從上訴至監察委員會[131]。

---

[128] 監察委員會，〈提出申訴，協助打造 Facebook 和 Instagram 的未來〉，https://oversightboard.com/appeals-process（最後瀏覽日：01/29/2024）。

[129] 陳憶寧，前揭註 22，頁 132。

[130] 本文亦發現，在 2021 年之前《規章制度和行為規範》中，僅規範 Facebook、Instagram 應「迅速執行裁決（implement the ruling promptly）」，並無目前的七日期限。Oversight Board, *supra* note 23. Article 2, Section 2.3.

[131] Douek, *supra* note 24, at 5.



doi:10.6342/NTU202400262

但本文發現，此一見解或許在〈案例裁決 2023-026-FB-UA〉後宜作修正，因該案件始於「用戶對 Meta 保留一則 Facebook 貼文的處置提出申訴」[132]。有疑者係，監察委員會運作依循的《規章制度和行為規範》，僅允許用戶在「內容因違反內容政策被移除」[133]的情況下始得上訴，該案如何使得監察委員會注意，又或監察委員會成員在發現爭議案件後，有主動發動審查的空間？由於該則案例裁決屬於 2023 年後新增設的「摘要裁決（summary decision）」[134]，並未實際進入審查，故暫時無從得知該案確切經過的程序，值得往後研究持續觀察監察委員會管轄權是否有擴張趨勢。

## 參、審查標準

在 53 則案例裁決中，除唯一一件無法受理的案件外，皆可見監察委員會無一例外地在裁決理由中，先檢視爭議之內容處置是否正確援引適用於 Facebook、Instagram 的《社群守則》[135]，並考量該處置是否遵循「Facebook 內容政策理念（Facebook's Values）」[136]，再援引《公民及政治權利國際公約》第 19 條的「合法性（legality）」、「正當目的（legitimate aims）」、「必要性與比例原則（necessity and proportionality）」三段式審查標準[137]，以檢驗 Facebook、Instagram 所作出的內容處置決定是否妥當。

### 一、《社群守則》

綜觀所有案例裁決，內容處置有無依循《社群守則》既有規範執行，可謂為該處置是否符合人權保障的首要標準。亦即《社群守則》是 Meta 公司執行內容管

[132] 監察委員會（2023），〈案例裁決 2023-026-FB-UA〉，https://oversightboard.com/decision/FB-IULHG7JK/（最後瀏覽日：01/29/2024）。
[133] Oversight Board, *supra* note 125.
[134] 依據監察委員會的說明，摘要裁決係「檢視 Meta 在監察委員會提請該公司注意某項內容後，該公司撤銷原始內容處置的案例」。
[135] 監察委員會（2023），〈案例裁決 2023-002-IG-UA, 2023-005-IG-UA〉，https://oversightboard.com/decision/IG-H3138H6S/（最後瀏覽日：01/29/2024）。
[136] 雖 Meta 公司已更名，但截至 2023 年 12 月，監察委員會網站中仍以「Facebook 內容政策理念」指涉該公司所制定、應用於 Facebook、Instagram 上的內容政策追求、保障之價值，參見本文第二章第一節第貳項。
[137] United Nations, General Assembly. International Covenant on Civil and Political Rights. *Treaty Series*, *999*, 171 (1966).


doi:10.6342/NTU202400262

理的主要依據，因此當監察委員會開始審查申訴時，會首先會確定該內容處置是否逾越了《社群守則》的授權範圍，或援引的條文是否足夠明確，且容易讓使用者理解[138]。

在〈案例裁決 2020-005-FB-UA〉中，一位用戶發佈了一則錯誤宣稱某內容為納粹德國時期的政府宣傳，然而該引言實則並非來自納粹政府。隨後，Facebook 以違反其《社群守則》中的「危險人物和組織」條款為由，移除了該則貼文，而監察委員會推翻該內容處置[139]。

監察委員會認定，儘管《社群守則》禁止用戶「支持」或「讚美」危險人物、組織，但 Facebook 並沒有公開其認定為危險人物、組織的名單。換言之，使用者在發佈與這些高度爭議性對象相關的內容時，無從預知是否可能觸犯《社群守則》而遭受到貼文被限制、移除的不利益處分。

監察委員會亦發現，Meta 公司的內部文件中，確實將納粹列為仇恨組織，但此一資訊並未對外揭露，所謂「支持」和「讚美」的定義也含糊不清[140]，致使一般用戶無從得知、理解。這些發現因而構成了 Facebook 在執行此項內容處置時，並未遵守《社群守則》的結論。

**二、Facebook 內容政策理念**

而後，監察委員會將會審查系爭內容處置，是否是為了遵循「Facebook 內容政策理念」，包括「言論自由」、「真實無欺」、「用戶安全」、「確保隱私」、「尊嚴無損」[141]。由於案件是審查 Meta 公司對於言論管理的處置是否合乎人權，而移除內容或限制帳戶等處置手段，無疑都會干涉用戶在 Facebook、Instagram 上表達意見的權利，因而在此一審查標準中，實皆為言論自由與其他四個政策理念的衝突。

53 則案例裁決中，監察委員會僅有一則裁決，認為 Meta 公司違反了其理念。在〈案例裁決 2021-006-IG-UA〉，一位 Instagram 用戶發佈了一張武裝暴力

[138] 監察委員會，前揭註 26。
[139] 同前註文。
[140] 同前註文。
[141] 此五內容政策理念，與本章第一節第壹項所介紹的《社群守則》核心價值並無不同，但其在官方中文翻譯中並未統一，Expression 亦有譯為「意見表達」、Authenticity 為「真實性」、Safety 僅作「安全」、Privacy 為「隱私」、Dignity 作「尊嚴」。


doi:10.6342/NTU202400262

組織領袖的照片，鼓勵公眾討論對該組織領袖施加單獨監禁，是否涉嫌違反人權。雖然如此，因為該武裝暴力組織和領袖都被 Facebook 歸類為「危險人物和組織」，該貼文因而被認定為支持危險人物而被移除[142]。

然而，Meta 公司在此則案例中，未能證明移除該則貼文、限制該用戶的言論自由後，能進而達成其餘四項政策理念，因為該則內容並不具有「恐嚇、排除或使其他用戶保持沉默」等潛在可能性，也與保護安全的政策理念無涉[143]。

### 三、國際人權標準

從已發佈的裁決中可以發現，在國際人權標準的審查階段中，監察委員會最常引用審查依據，包括：《公民與政治權利國際公約》及其一般性意見書、聯合國《工商企業與人權指導原則》、《拉巴特行動計劃（Rabat Plan of Action）》[144]、聯合國人權調查報告等國際公約或其他相關文書。

如前所述，聯合國《工商企業與人權指導原則》旨在鼓勵私部門尊重並避免侵犯人權[145]。學界認為，這些原則對私部門不具有法律約束力，因為企業並不如國家參與公約，是協議的締約方，也不具備國家行為者的要素，而且原則中使用的措辭是「應（should）」，與具備義務性質的「必須（shall）」有別[146]。然而，該原則仍為企業提供了保障人權的指導方針和基本標準，並對私部門施加社會責任。作為世界上最大的科技公司之一，Meta 公司基於社會責任，稱自願遵守《工商企業與人權指導原則》，授權監察委員會採用與言論自由相關的國際人權標準，對其推出的產品及服務進行審查。

準此，監察委員會始有立足點採納《公民與政治權利國際公約》第 19、20 條意旨，即言論自由可以受到一定的限制，前提是這些限制具有法律依據，並且

---

[142] 監察委員會（2021），〈案例裁決 2021-006-IG-UA〉，https://oversightboard.com/decision/IG-I9DP23IB/（最後瀏覽日：01/29/2024）。

[143] 同前註文。

[144] United Nations, Human Rights Council, *supra* note 55.

[145] United Nations, Office of the High Commissioner for Human Rights. (2011). *Guiding Principles on Business and Human Rights*. HR/PUB/11/04, at 13. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

[146] Knox, J. H. (2011). The Ruggie Rules: Applying Human Rights Law to Corporations. In Mares, R. (Ed.), *The UN Guiding Principles on Business and Human Rights: Foundations and Implementation*, pp. 51-83. Martinus Nijhoff Publishers.



doi:10.6342/NTU202400262

是為了尊重他人的權利或名譽、保護國家安全、公共秩序、公共健康或道德、禁止戰爭宣傳、禁止煽動歧視、敵意或暴力的民族、種族或宗教仇恨等目的。在《公民與政治權利國際公約》第 19 條一般性評論第 34 號中，聯合國人權委員會進一步要求這些限制必須以最嚴格審查標準進行檢視，以確保其「符合必要性和比例原則」[147]。

監察委員會將這些原則演變為三段式標準，包括：「合法性」、「正當目的」、「必要性與比例原則」。要通過合法性測試，該處置所引用的《社群守則》必須明確、可得預見且公開可近用（clear, foreseeable, and accessible）[148]。如前所述，在〈案例裁決 2020-005-FB-UA〉中，由於 Meta 公司並未向用戶公開說明納粹是在平台上被禁止的危險組織，所以也未能通過此處的合法性標準[149]。

正當目的的審查，則是為了確保 Facebook、Instagram 所作出的處置，符合《公民與政治權利國際公約》允許的目的。亦即，除非系爭內容牽涉國家安全、公共秩序、公共健康或與種族和宗教仇恨有關，否則 Meta 公司不得隨意制定新的《社群守則》來限制用戶於平台上的言論。

在國際人權標準審查階段的最後一部分，監察委員會並未明確揭露如何推演「必要性與比例原則」，但對不同的內容處置，以對用戶干涉、影響的程度進行劃分。如同〈案例裁決 2020-006-FB-FBR〉，監察委員會引用了《公民與政治權利國際公約》第 34 號一般性意見，要求 Meta 公司「應證明其已選擇了最不具干涉性的手段，來解決合法的公共利益目標。」此外，監察委員會也引用聯合國《促進和保護意見和表達自由權特別報告》，要求 Meta 公司證明「欲實現的公共利益目標已無法通過不侵犯言論的手段來解決」、「在侵犯言論的措施中，已選擇了最不具干涉性的手段」以及「所選擇的手段確實有助於實現目標」[150]。

從多個案例裁決中可以歸納得知，監察委員會認為 Meta 公司可用的處置手段包括移除內容、增加警告標語，以及封鎖或暫停帳戶等。在〈案例裁決 2020-006-FB-FBR〉中，監察委員會認定限制內容散布範圍和增加警告標語，相較直接

[147] United Nations Human Rights Committee. (2011). *General comment no. 34, Article 19, Freedoms of opinion and expression*. CCPR/C/GC/34. https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf
[148] 監察委員會（2020），〈案例裁決 2020-003-FB-UA〉，https://oversightboard.com/decision/FB-QBJDASCV/（最後瀏覽日：01/29/2024）。
[149] 監察委員會，前揭註 26。
[150] 監察委員會，前揭註 50。


doi:10.6342/NTU202400262

移除內容，是干涉程度較小的手段[151]。在〈案例裁決 2021-001-FB-FBR〉，監察委員會則認為無限期對用戶停權，是一項不被允許的手段，因為此舉嚴重侵犯用戶在社群平台上表達意見的權利，但如果符合特定情況，將用戶暫時停權一個合理、得確定的期間，則可能被視為適當的政策執行作法[152]。

## 第四節　小結

綜上所述，我們可以初步理解社群媒體所用以管理內容的依據、機制既廣且雜，又更動頻繁，或許必須如此才足堪應付平台上時刻新增、難以計數、主題各異的言論。本文亦認為，《社群守則》等內容政策與 Meta 公司用以執行政策的審查系統，如同平台內容管理的兩大支柱，卻也會雙向互動、彼此影響：審查系統以內容政策為指引進行作業，但如審查系統執行結果發生錯誤、矛盾，Meta 公司的內容政策團隊將再對政策作出調整、修正。

而於近五年成立的監察委員會，則好比是兩支柱的外部建築監工，不但手握 Facebook、Instagram 對言論自由想像的設計原圖（《社群守則》及 Facebook 內容政策理念），亦要求其構造必須符合國際標準。然而，監察委員會是否有能力從表面監督到深入探查，根本改善社群媒體治理為人詬病之處，卻似嫌不足。

考量美國《通訊端正法》其實給予平台相當寬裕的免責空間[153]，值得加以討論的，或許是社群媒體願意投入資源進行內容管理的根本動機為何，也是未來能繼續促使其加速改善的主要誘因。

學者有認為源於企業責任、經濟考量兩面向[154]，企業責任包含員工對產品的期待、認同，促使其對既無益於社會、又斲傷公司聲譽的內容自發採取行動；經濟考量則旨在留住使用者，也反映社群媒體具有「雙邊市場（two-sided market）」的特性[155]，即用戶會傾向使用有較多與自己相關、有興趣內容的平台，用戶數量多，也才能為平台創造廣告收益。然而，若是移除太多內容，平台

[151] 同前註文。
[152] 監察委員會，前揭註 16。
[153] 參見本文第一章第二節第壹項。
[154] Klonick, *supra* note 91, at 1625.
[155] Thépot, F. (2013). Market Power in Online Search and Social Networking: A Matter of Two-Sided Markets. *World Competition*, *36*(2), 195-221.


doi:10.6342/NTU202400262

會失去用戶信任；保留全部內容，又會讓用戶有不愉快的使用體驗，進而影響營收表現[156]，一個完善、符合使用者期待的內容管理機制，因此成為最大化平台成長的關鍵角色。

[156] Klonick, *supra* note 91, at 1627.



doi:10.6342/NTU202400262

# 第三章　社群媒體言論自由的重要界線：立即危害

監察委員會創設目的，是為了牽制Meta公司不致任意制定《社群守則》等內容政策，放任或過度限制 Facebook、Instagram 上的言論表達，而促使應在聯合國公約、決議、調查員報告等國際人權標準的框架下，盡可能維護平台言論自由[157]。因此，監察委員會如何詮釋原先用以約束國家的人權責任，並加諸社群媒體之上，即是本研究所欲討論者。

本章將聚焦於截至 2023 年 10 月，監察委員會所發佈的 18 則以「立即危害」作為檢驗Meta公司的內容處置決定是否符合比例原則的案例裁決，以研討如何在考量社群媒體特性、無法確知結果的情形下，適當地評估一則內容可能會導致現實世界產生危險，並比較其與美國實務上的「明顯而立即危險」原則異同之處。

## 第一節　與「立即危害」相關的案例裁決

由於本研究撰寫當下，國內外中、英文研究皆尚無以本文欲分析的「立即危害」為主題，對監察委員會所作之案例裁決進行整理，故作者必須先將 53 則案例裁決盡數閱讀後，從中提取以「立即危害」作為審查判準的案例，始能進行後續討論。

本文在耙梳資料過程發現，監察委員會在撰寫案例裁決時，若有將爭議內容可能產生的危害風險立即性納入考量，所採用的用字遣詞，在英文、中文實未有固定。比如在〈案例裁決 2020-006-FB-FBR〉中以「立即傷害（imminent harm）」[158]稱之；在〈案例裁決 2021-013-IG-UA〉中則未見，而是以「可能造成傷害……直接立即的關聯（direct and immediate connection between… and the possibility of harm）」[159]表述；而中文版本的〈案例裁決 2022-007-IG-MR〉，則將「imminent harm」譯作「立即危害」[160]，鑑於監察委員會有逐漸完善其裁決體系的趨勢，本文遂以較新用語從之。

---

[157] *See* Wong, D. & Floridi, L., *supra* note 124, at 267.

[158] 監察委員會，前揭註 50。

[159] 監察委員會，前揭註 72。

[160] 監察委員會（2022），〈案例裁決 2022-007-IG-MR〉，https://oversightboard.com/decision/IG-PT5WRTLW/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

然而，前揭不同敘述所要表達的概念實質上並無二致，皆是在爭議內容被認為可能導致現實世界傷害，但無法確知結果的情況下，監察委員會用以衡量 Meta 公司限制該內容出現在平台上是否確有必要，或可能只是過度連結、假想難以實現的風險，反而不成比例地限縮社群媒體的言論自由空間，所使用的判斷標準。

在此脈絡下，本文從 53 則案例裁決中，整理出 18 則監察委員會在裁決理由中，引用內容是否會導致「立即危害」作為審查標準的案例裁決，製表如下。

| 裁決編號 | 裁決名稱 | 結果 | 涉及平台 | 涉及政策 | 涉及主題 |
|---|---|---|---|---|---|
| 2020-005-FB-UA | 納粹引述 | 推翻 | Facebook | 危險人物和組織 | 政治 |
| 2020-006-FB-FBR | 宣稱的新冠肺炎療法 | 推翻 | Facebook | 暴力和煽動仇恨 | 健康、錯誤資訊、安全 |
| 2020-007-FB-FBR | 抵制法國的印度抗議 | 推翻 | Facebook | 暴力和煽動仇恨 | 宗教、暴力 |
| 2021-001-FB-FBR | 美國前總統<br>川普帳號遭停權 | 維持 | Facebook | 危險人物和組織 | 表達自由、政治、安全 |
| 2021-008-FB-FBR | 巴西的新冠肺炎封城 | 維持 | Facebook | 暴力和煽動仇恨 | 政府、健康 |
| 2021-013-IG-UA | 熬煮死藤水 | 推翻 | Instagram | 管制商品 | 文化活動、健康、宗教 |
| 2021-014-FB-UA | 宣稱在拉亞科博<br>發生的犯罪 | 維持 | Facebook | 仇恨言論 | 表達自由、戰爭和衝突 |
| 2021-015-FB-UA | 開立 Adderall® | 推翻 | Facebook | 管制商品 | 歧視、健康、安全 |
| 2022-006-FB-MR | 提格雷通訊事務局 | 維持 | Facebook | 暴力和煽動仇恨 | 政府、暴力、<br>戰爭和衝突 |
| 2022-007-IG-MR | 英國鑽頭音樂 | 推翻 | Instagram | 暴力和煽動仇恨 | 表達自由、藝術／<br>寫作／詩詞、政府 |
| 2022-008-FB-UA | 俄羅斯文詩詞 | 推翻 | Facebook | 仇恨言論 | 藝術／寫作／詩詞、<br>戰爭和衝突 |
| 2023-001-FB-UA | 巴西將軍的演講 | 推翻 | Facebook | 血腥暴力內容<br>預謀傷害和宣揚犯罪<br>暴力和煽動仇恨 | 選舉、政府、陳情抗議 |
| 2023-002-IG-UA<br>2023-005-IG-UA | Violence against women | 推翻 | Instagram | 仇恨言論 | 性別和性別認同平等、<br>表達自由 |
| 2023-003-FB-MR | 柬埔寨首相 | 推翻 | Facebook | 暴力和煽動仇恨<br>預謀傷害和宣揚犯罪 | 選舉、政治、陳情抗議 |
| 2023-004-FB-MR | 亞美尼亞戰俘影片 | 維持 | Facebook | 預謀傷害和宣揚犯罪 | 戰爭和衝突、<br>表達自由、安全 |



doi:10.6342/NTU202400262

| | | | | | |
|---|---|---|---|---|---|
| 2023-006-FB-UA | 性別暴力的圖像 | 推翻 | Facebook | 霸凌和騷擾 | 性別和性別認同平等、幽默、暴力 |
| 2023-010-IG-MR | 提倡將 K 他命應用在非 FDA 核准的療法 | 推翻 | Instagram | 管制商品 | 表達自由、健康 |
| 2023-026-FB-UA | 衣索比亞的旅館 | 推翻 | Facebook | 暴力和煽動仇恨 | 暴力、戰爭和衝突 |

**表 3：與「立即危害」相關的監察委員會案例裁決列表（截至 2023 年 10 月）**

（本表為作者自製）

從列表中，我們可以初步發現，有一定比例以「立即危害」作為移除、限制理由的案例裁決，與該內容和煽動暴力、犯罪相關，亦有涉及疫情、藥物等可能損害人身健康的案件，本研究將擇取其中較具代表性，且對「立即危害」有詳加闡述的案例裁決，如下析論之。

## 第二節　初見「立即危害」作為審查標準

監察委員會首次明確提及移除內容適當與否，應評估是否會造成「立即危害」，出現於〈案例裁決 2020-006-FB-FBR〉[161]。而在此之前的〈案例裁決 2020-005-FB-UA〉僅較抽象提及內容的「回應、用戶所在地點，以及貼文在競選活動期間的發佈時間」[162]，即與內容相關的背景、迴響等可能影響所造成的結果，皆應納入考量，強調處置決定的正當性不僅侷限於內容本身。

### 壹、〈案例裁決 2020-006-FB-FBR〉：宣稱的新冠肺炎療法

〈案例裁決 2020-006-FB-FBR〉中的爭議內容被認定違反《社群守則》「暴力和煽動仇恨」一節，其中禁止「可能導致立即暴力行為或人身傷害的錯誤資訊和無法證實的謠言」。該則內容發佈於法國新冠肺炎疫情流行期間，以影片形式聲稱羥氯奎寧搭配阿奇黴素能有效治癒新冠肺炎，並批評法國政府推廣使用另一名為瑞德西韋的藥物為錯誤決策[163]。由於彼時世界衛生組織尚未認定任何新冠肺

[161] 監察委員會，前揭註 50。
[162] 監察委員會，前揭註 26。
[163] 監察委員會，前揭註 50。



doi:10.6342/NTU202400262

炎的有效療法，Facebook 因此認定該內容可能使得用戶逕作嘗試，導致錯誤用藥，進而造成人身傷害，因而決定從平台上移除該影片[164]。

須先指明的是，此案中監察委員會所審查、決議者，並不在於往後審查流程是否皆應納入內容可能導致「立即危害」的判斷，而是因該內容所違反的《社群守則》中，本身即有「立即人身傷害」的要件，因此監察委員會加以詮釋、定義，在何種條件下始該當此要件。

監察委員會認為，若某內容被證實為謠言、錯誤資訊後，是否符合「立即」危害而得移除，應考量「貼文發佈者的狀態和可信度、其言論的觸及對象、使用的確切用語，以及容易誤信訊息的用戶能否取得訊息中聲稱的療法及治療藥物。」又因在法國，取得羥氯奎寧、阿奇黴素兩藥物需由醫師開立處方箋，一般民眾實無法輕易取得，且該療法未經法國官方驗證，難認常人會盡信其說法而擅自服用[165]。

監察委員會於是裁決，縱然該內容為錯誤資訊，但因不具有立即性的危害，以移除作為處置手段恐過分侵害用戶在平台上的言論自由空間，且 Facebook 尚有對該內容增加提醒標籤、縮減散布範圍等干涉性較低的處置手段，原處置決定因而遭到推翻，同時 Facebook 應讓該影片恢復顯示[166]。

從上述理由中，我們可以發現，監察委員會認為社群媒體如要在合乎國際人權標準的情況下，對潛在有害的內容進行管理，即不宜單純透過內容政策規範來限制，亦要綜合考量爭議言論的背景、脈絡，且積極證明其確有立即導致危害的可能性，才能以干涉性最大的移除手段限制內容。

## 貳、〈案例裁決 2020-007-FB-FBR〉：抵制法國的印度抗議

無獨有偶，在隨後發佈的〈案例裁決 2020-007-FB-FBR〉中，監察委員會亦是針對違規內容是否有該當《社群守則》所禁止的「會造成實際人身傷害，或對公共安全有直接威脅」進行審查。值得注意的是，此案件中不如〈案例裁決 2020-006-FB-FBR〉所引用的《社群守則》中即有「立即危害」的要件，而是在

[164] 同前註文。
[165] 同前註文。
[166] 同前註文。


doi:10.6342/NTU202400262

案例裁決理由中，由監察委員會肯認 Facebook 主張所謂內容「會造成實際人身傷害」即應是「立即人身傷害」而言[167]，近一步將 Facebook、Instagram 兩平台上凡是可能造成傷害的內容，限縮定義至應同時該當「立即危害」才能被移除。

〈案例裁決 2020-007-FB-FBR〉所審查的內容，係一則包含圖片、文字的貼文，圖片含有譯為「如果異教徒開始對先知出言不遜，那麼寶刀就該出鞘」的北印度文，而文字內容則以英文呈現，談及先知對貼文發佈者的神聖性。Facebook 判定對異教徒寶刀出鞘，似有隱藏的暴力威脅，因而將其移除。

然而，監察委員會最終裁決，「由於目標對象廣泛」，「（發佈貼文的）用戶應非國家組織或公眾人物」且「未暗示要在特定時間或地點威脅採取或煽動行動」，因此發生人身傷害或暴力的可能性並不急迫，其風險並未達到合理取代言論自由的程度[168]，Facebook 應恢復顯示該則貼文。

有疑者係，雖然監察委員會在這三個指標中，都認定系爭貼文並未該當「立即危害」，但如果其中一個指標增加風險實現的可能性，例如對象特定、使用者具備號召力或明示特定時間或地點，但其餘指標並未顯現危險的立即性，是否仍然會具備移除的正當性？也就是說，三個指標究竟是法律上的構成要件（elements），必須同時具備方能成立，或僅為要素（factors），最終仍須綜合考量、個案判斷？

監察委員會在此則案例裁決中並未明確回答，但本文亦發現，裁決理由中提及，審查過程中確有參考聯合國《拉巴特行動計劃》的建議，始認為系爭貼文未達到「國家需禁止的煽動歧視、敵視或暴力之程度」[169]。

## 第三節　川普案與《拉巴特行動計劃》

從監察委員會甫成立的 2020 年所作成的七件案例裁決可知，雖已有觸及危險內容應達立即危險始得禁止、立即危害應考量現實世界可能性等兩原則，且試圖以內容所提及的對象、時間、地點是否特定，及發佈者是否具有影響力，作為社群媒體內容是否能達成「立即危害」的評估要素，但似乎尚未言明如何具體操作

[167] 監察委員會（2020），〈案例裁決 2020-007-FB-FBR〉，https://oversightboard.com/decision/FB-R9K87402/（最後瀏覽日：01/29/2024）。
[168] 同前註文。
[169] 同前註文。


doi:10.6342/NTU202400262

此標準，而隨之公布、備受矚目的〈案例裁決 2021-001-FB-FBR〉，提供了更進一步的解釋。

## 壹、〈案例裁決 2021-001-FB-FBR〉：美國前總統川普帳號遭停權

2021 年 1 月 6 日，正當位於美國華盛頓哥倫比亞特區的國會山莊進行 2020 年美國總統大選選舉人票清點程序時，史無前例地發生抗議群眾闖入國會大廈，並導致暴力行動、造成人員傷亡。示威者主張民主黨及總統當選人拜登（Joe Biden）在選舉中有舞弊情事，並試圖透過進入國會大廈以阻止其宣布拜登當選，咸認為受到時任美國總統川普（Donald J. Trump）在同日集會上呼籲支持者「We’re going to the Capitol.（我們將前往國會）」煽動[170]。

在暴動期間，川普在其 Facebook 粉絲專頁上發佈一則影片[171]，旨在呼籲示威者「But go home and go home in peace.（回家，和平地回家）」，卻也在影片中再次強調對手陣營「take it away（奪走）」選舉，拜登當選的結果為「fraudulent election（選舉舞弊）」等語。逾一小時後，川普又發佈一則貼文，稱「election victory is so unceremoniously & viciously stripped away from great patriots（選舉勝利被粗暴、惡劣地從偉大的愛國者手中奪走）」，卻也再次呼籲支持群眾「Go home with love in peace.（帶著愛與和平回家）」[172]。

值得注意的是，在川普於粉絲專頁發佈第二則貼文前，第一則影片已因違反《社群守則》「危險人物和組織」一節中的禁止「對參與違法活動的團體、領導者或個人表示支持與頌揚」內容而遭到 Facebook 移除，隨後第二則貼文亦因相同理由而被移除，並被限制在 Facebook、Instagram 二平台上無法發文 24 小時[173]。

然而，在 1 月 7 日時，因 Facebook 認為川普「利用我們的平台來煽動對民主選出的政府進行暴力叛亂」[174]，宣布將延長封鎖川普粉絲專頁、Instagram 帳號至

[170] 監察委員會，前揭註 16。

[171] 原影片已遭 Facebook 移除，請參見 CNN. (2021, January 7). Trump tells rioters to 'go home' while repeating election lies. *YouTube*. https://youtu.be/3_JxN9CwIMU

[172] 原貼文亦已遭 Facebook 移除，請參見 Delkic, M. (2022, May 10). Trump’s banishment from Facebook and Twitter: A timeline. *The New York Times*. https://www.nytimes.com/2022/05/10/technology/trump-social-media-ban-timeline.html

[173] 監察委員會，前揭註 16。

[174] 同前註文。



doi:10.6342/NTU202400262

少兩週，或無限期直至政權和平轉移。事實上，此一封鎖後來於 2023 年 1 月 25 日，Meta 公司才以「公眾應該要能夠聽到政治人物所言，才能在充足資訊下作出決定」[175]為由，解除對川普在 Facebook、Instagram 帳號的限制。

由於美國國會暴動、川普社群媒體帳號被封鎖為世界所矚，且川普作為時任美國總統，其粉絲專頁、帳號應屬美國憲法第一修正案所保護的「公共論壇」[176]，私人科技公司以《社群守則》等服務規範逕為限制是否得當？無限期地封鎖政治人物的 Facebook、Instagram 帳號，是否有過度干涉言論自由，並侵害一般民眾知的權利？如是問題，促使 Facebook 主動將此一處置決定轉介予監察委員會進行審查。

宜先辨明者係，在〈案例裁決 2021-001-FB-FBR〉中，監察委員會對於 Facebook 引用《社群守則》「危險人物和組織」作出移除兩則內容的處置決定並無疑義，因其言論「支持」闖入國會、暴力示威等「違法活動」至為明確。然而，Facebook 對川普於兩平台的帳號進行 24 小時的限制，以及後續追加無期限的封鎖，此二者才是監察委員會運用國際人權標準進行審查的主要標的。

## 貳、〈案例裁決 2021-001-FB-FBR〉引用《拉巴特行動計劃》

所謂《拉巴特行動計劃》是聯合國人權高級專員辦事處（Office of the United Nations High Commissioner for Human Rights）於 2012 年通過、2013 年發佈，匯集人權專家對於各國如何遵照《公民與政治權利國際公約》第 19、20 條意旨，制定禁止鼓吹仇恨、歧視、暴力主張言論法律的立法建議[177]。

蓋因《公民與政治權利國際公約》第 20 條明文：「任何鼓吹戰爭的宣傳，應以法律禁止之。任何鼓吹民族、種族或宗教仇恨之主張，構成煽動歧視、敵視或強暴（violence）者，應以法律禁止之」[178]。然而，言論自由既為民主社會最重要

[175] The public should be able to hear what politicians are saying so they can make informed choices. Retrieved from Clegg, N. (2023). Ending Suspension of Trump's Accounts With New Guardrails to Deter Repeat Offenses. *Meta*. https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/

[176] 參見前揭註 11。

[177] Office of the United Nations High Commissioner for Human Rights. (n.d.) OHCHR and freedom of expression vs incitement to hatred: the Rabat Plan of Action. *United Nations*. https://www.ohchr.org/en/freedom-of-expression

[178] 英文為「Any propaganda for war shall be prohibited by law. Any advocacy of national, racial or



doi:10.6342/NTU202400262

基石，如何在最嚴謹、有限度地情況下，以適當的法律限制對公眾有害的言論，《公民與政治權利國際公約》則未有明確指引。

因而在後續的《拉巴特行動計劃》中，人權專家結論建議，縱然是欲限制煽動仇恨（incitement to hatred）的言論，該言論也必須要達到最嚴重（most severe）且令人深切感受恥辱（deeply felt form of opprobrium），國家始能以刑事訴追方式進行限制。其衡量標準有六，分別為「背景（context）」、「發言者（context）」、「意圖（intent）」、「內容和形式（content and form）」、「程度及範圍（extent of the speech act）」及「危害可能、急迫程度（likelihood, including imminence）」

申言前揭的六項門檻測試（six-part threshold test），《拉巴特行動計劃》建議，在分析「背景」時，「應將言論置於當時主流的社會與政治脈絡」下考量；「發言者」的社會地位、對應的受眾皆可能是變因；「意圖」強調過失及重大過失（negligence and recklessness）並不足以構成違反《公民與政治權利國際公約》第 20 條的鼓吹或煽動；「內容和形式」則視乎該言論多挑動（provocative）、直接（direct）；「程度及範圍」應考量言論是否公開，或只在有限的範圍內傳播；「危害可能、急迫程度」要求法院必須確定言論與煽動實際行動間有「合理的可能性（reasonable probability）」始能作出限制[179]。

監察委員會亦引入此六項門檻測試，作為在案例裁決第三階段的國際人權標準審查時，評估無限期封鎖川普帳號的處置決定是否妥適的標準。但必須提前指出的是，事實上監察委員會在第一階段[180]，審查該處置決定是否合乎《社群守則》時，即指出在 Meta 公司所制定的內容政策中，並無提示使用者帳號有可能遭到無限期限制，因而此一舉措已「違反了維護表達自由的原則」[181]。

本文發現，在監察委員會類似於憲法審查的流程中，不若美國聯邦最高法院在進行「目的」及「手段與目的之關聯」兩判準審查時，如「目的」非合法、正當即屬違憲，或我國時而僅論「目的合憲性」與「手段適當性」，時而著重「狹

---

religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.」，此處採聯合國官方中文翻譯。United Nations, General Assembly, *supra* note 137, Article 20.

[179] United Nations, Human Rights Council, *supra* note 55, at 11.

[180] 參見本文第二章第三節第參項。

[181] 監察委員會，前揭註 16。



doi:10.6342/NTU202400262

義比例性」[182]。在所謂監察委員會的三階段審查標準，即《社群守則》、Facebook 內容政策理念與國際人權標準中[183]，如其中一階段未通過，監察委員會仍會續行審查，此案雖在《社群守則》階段就被認定有違，但仍於國際人權標準審查時詳加論述，即為一例。

監察委員會認定，在「背景」部分，由於是項暴動成因與川普密切相關，且二則貼文發佈時，暴力行為尚在持續進行，後續局勢仍然不穩；川普作為時任美國總統，「發言者」權威、影響力不言可喻；「意圖」層面，監察委員會雖無法斷定，但認為川普理應認知其發言將會帶來風險；「內容和形式」部分，則有支持、讚揚暴力示威者的語句；「程度及範圍」則考量川普的 Facebook 粉絲專頁追蹤者達 3500 萬、Instagram 帳號追蹤者達 2400 萬，亦會受媒體關注、報導；「危害可能、急迫程度」部分，監察委員會則強調川普發文當下，抗議者正在攻擊國會大廈[184]。

準此，監察委員會裁決 Facebook 於 1 月 6 日對川普帳號所施加的 24 小時限制有其正當性，有效避免可能產生的暴力危險，但無限期停權的處置方式，則因未有清楚、明確的政策依據，而被認定有失允當，責令 Facebook 應盡快修訂政策。

值得再深究者，係《拉巴特行動計劃》的六項門檻測試，於〈案例裁決 2021-001-FB-FBR〉中，監察委員會究竟認定其為構成煽動暴力嚴重風險的「要件（elements）」抑或「因素（factors）」？本文認為，此項辨別有其必要之處，在於若以前者作為認定，則缺一不可，包含發言者的意圖皆須確認，始該當《公民與政治權利國際公約》第 20 條「構成煽動歧視、敵視或強暴」的鼓吹仇恨；然若為後者，則縱六項門檻有一無從證明，或與其他門檻有相反呈現，則仍可權衡認定。

本文發現，雖然監察委員會試圖將大部分適用於國家行為者的各項國際人權標準套用於 Meta 公司，但此處恐為有意或必須與《拉巴特行動計劃》作出不同解釋。蓋因《拉巴特行動計劃》乃是為國家行為者制定法律設想，對可能導致暴力

[182] 湯德宗（2009），〈違憲審查基準體系建構初探——「階層式比例原則」構想〉，廖福特（主編），《憲法解釋之理論與實務》，頁 8，中央研究院法律學研究所籌備處。
[183] 需注意的是，監察委員會在國際人權標準的審查階段也有三段式審查標準，即「合法性」、「正當目的」、「必要性與比例原則」，參見本文第二章第三節第參項。
[184] 監察委員會，前揭註 16。


doi:10.6342/NTU202400262

的言論以刑事法規相繩，實為對言論自由最直接、嚴重的干涉。因而，手段適法與否，應傾向最嚴格的認定，避免過度侵害人民基本權利，始能符合如美國憲法對高價值言論內容管制須符合「嚴密剪裁（narrowly tailored）」的嚴格審查標準（strict scrutiny）[185]。實際上，《拉巴特行動計劃》亦將六項門檻指為 elements（要件），並在相對難證明的發言者「意圖」一項，強調若言論係出於過失及重大過失皆不足以構成國家刑事體系下的犯罪[186]，足證本文見解。

然而，對於社群媒體來說，無論是從經濟成本考量或實務可行性出發，如要求平台必須將六項門檻測試視作「要件」來審查內容，則恐過於苛求且不切實際。因為社群媒體不若法院，無從透過訊問、調查證據等法律程序探求真實情況，尤其貼文發佈者的「意圖」，勢必只能透過分析其帳號互動、內容脈絡等資訊，終以心證進行推論其動機，如同監察委員會在本件案例裁決中坦言「無法對川普的意圖妄下結論」[187]。而若每則爭議內容皆欲深切探究是否該當六項門檻，在現況審查違反其他《社群守則》內容的人力尚且不足的情況下，遑論再等待細緻分析適法與否，恐怕重大風險已經實現且難以挽回。

或因如此，監察委員會在〈案例裁決 2021-001-FB-FBR〉皆稱六項門檻為 factors（要素），留予自己與 Meta 公司在未來認定內容是否達立即危害時，得以在無法確證六項門檻時，能有綜合判斷的空間及彈性。事實上，在後續的〈案例裁決 2021-002-FB-UA〉中，監察委員會亦肯認，如黑臉妝雖為仇恨言論，卻未達國家透過法律限制的程度，但社群媒體仍可且有必要進行處置[188]；又或於〈案例裁決 2022-007-IG-MR〉直指「Meta 的人權責任可能與國家／地區執行刑事或民事處分的責任不同」[189]。

對此，聯合國促進和保護意見和表達自由權問題特別報告員（Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression）也持肯定見解，認為社群媒體公司如欲對未達《公民與政治權利國際公約》第 20 條鼓吹仇恨主張的內容進行限制，若能說明限制理由且手段符合比例

[185] 黃昭元（2004），〈憲法權利限制的司法審查標準：美國類型化多元標準模式的比較分析〉，《國立臺灣大學法學論叢》，33 卷 3 期，頁 39-41。
[186] United Nations, Human Rights Council, *supra* note 179.
[187] 監察委員會，前揭註 16。
[188] 監察委員會，前揭註 52。
[189] 監察委員會，前揭註 160。


doi:10.6342/NTU202400262

原則，仍可認為符合言論自由保障[190]，顯示在認定「立即危害」內容的嚴謹程度上，國家行為者與非國家行為者得有一定程度之落差，而由於後者所施行手段非如刑罰侵害甚鉅，尚可以較寬鬆的規範限制之。

## 第四節　延續《拉巴特行動計劃》六項門檻

監察委員會在〈案例裁決 2021-001-FB-FBR〉引入《拉巴特行動計劃》，作為鼓吹仇恨、煽動暴力內容是否達「立即危害」的判斷標準後，由於其欲透過累積案例裁決、建立先例的運作模式，在後續可能導致現實世界發生暴力、危險的案件中，我們亦可發現六項門檻測試的適用情形。

譬如在〈案例裁決2022-006-FB-MR〉中，監察委員會審查一則由衣索比亞境內武裝勢力發佈，鼓勵國民軍反叛、政府軍投降，否則結局只有死亡的貼文，監察委員會亦以六項門檻測試，認定該內容是在當地內戰情勢升高時發佈、發文者為地方軍政部門且追蹤人數逾二十萬、意圖呼籲殺死不投降士兵等，已達導致「立即危害」的風險，即為違反《社群守則》禁止「煽動或助長嚴重暴力行為」的內容，肯認 Meta 公司予以移除的決定為妥適之舉[191]。

此外，本文亦發現，如內容發佈的時機處於戰爭、暴動或政局動盪時，監察委員會似傾向認定帶有威脅、恐嚇的言論該當《拉巴特行動計劃》的六項門檻測試，舉凡背景同為衣索比亞武裝衝突的〈案例裁決 2021-014-FB-UA〉，監察委員會甚至進一步認定「在持續發生武裝衝突且其政府機構被評估為無法履行國際法人權義務的國家／地區，Meta 得限制其原本不加以限制的表達自由」[192]；〈案例裁決 2023-001-FB-UA〉中亦可見，監察委員會雖無針對六項門檻測試逐一說明，卻對於該內容發佈時的背景、危害急迫程度詳加闡述，認為在選後否認結果、軍事政變謠言四起時機下，由軍隊將領呼籲武裝攻擊極可能導致「立即危害」[193]。

---

[190] United Nations, Human Rights Council. (2018). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/38/35, at 11. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf

[191] 監察委員會（2022），〈案例裁決 2022-006-FB-MR〉，https://oversightboard.com/decision/FB-E1154YLY/（最後瀏覽日：01/29/2024）。

[192] 監察委員會（2021），〈案例裁決 2021-014-FB-UA〉，https://oversightboard.com/decision/FB-MP4ZC4CC/（最後瀏覽日：01/29/2024）。

[193] 監察委員會（2023），〈案例裁決 2023-001-FB-UA〉，https://oversightboard.com/decision/FB-659EAWI8/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

然而，雖然同屬可能導致暴力行為的鼓吹仇恨主張，內容的呈現形式亦有可能在監察委員會適用《拉巴特行動計劃》時扮演關鍵變因。例如〈案例裁決 2022-007-IG-MR〉審查一則帶有槍枝、射殺等隱語饒舌歌曲的內容，但因監察委員會認為虛張聲勢為同類音樂的常見表現手法，而難認其有具體威脅或立即危害風險[194]；抑或〈案例裁決 2022-008-FB-UA〉中，爭議貼文列舉俄羅斯在烏克蘭的暴力行動，並引用蘇聯某詩詞的其中一句，稱「殺死法西斯主義者」，雖然內容發佈時點正值俄烏戰爭，也有可能升高緊張情勢，但因詩詞本身為展現對暴力的反思，且因一篇貼文導致衝突情況大幅加劇的效果亦有限，考量維護平台言論自由及國際人權標準，監察委員會遂裁決 Meta 公司不宜移除該內容，改採用干涉度較低的警告標語為已足[195]。

## 第五節　與「立即而危險原則」比較

關於何謂言論自由、界線如何劃定的討論，自美國於 1791 年通過憲法第一修正案[196]後從未停止，從 1798 年制定《反煽動叛亂法（the Seditious Act）》，將意圖詆毀政府、總統、國會的「不實、誹謗與惡意文字，意圖詆毀」定為聯邦犯罪，被認為是第一次對言論自由作出限制；直至 19 世紀，各級法院尚傾向容許壓制任何「具有害傾向（bad tendency）」，即可能冒犯思想健全人民的言論[197]；1919 年，聯邦最高法院大法官 Oliver Wendell Holmes 在 Schenck v. United States 案指出，言論自由不應保護「一個在劇院裡大聲謊叫失火、造成恐慌的人（a man falsely shouting fire in a theatre and causing a panic）」，但國會應該要考量言論是否具備「明顯而立即危險（clear and present danger）」[198]。

---

[194] 監察委員會，前揭註 160。

[195] 監察委員會（2023），〈案例裁決 2022-008-FB-UA〉，https://oversightboard.com/decision/FB-MBGOTVN8/（最後瀏覽日：01/29/2024）。

[196] 依據美國在台協會的中文翻譯，美國憲法第一修正案譯為「國會不得制定有關下列事項的法律：確立一種宗教或禁止信教自由；剝奪言論自由或出版自由；或剝奪人民和平集會及向政府要求伸冤的權利。（Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.）」

[197] Lewis, A.（著），林凱雄（譯）（2020），《異見的自由：美國憲法增修條文第一條與言論自由的保障》，頁 88-92，八旗文化。

[198] Schenck v. United States, 249 U.S. 47, 52, 39 S. Ct. 247, 249, 63 L. Ed. 470 (1919)



doi:10.6342/NTU202400262

逾一世紀以來，隨著「明顯而立即危險」的定義迭有演變，此原則儼然已經形成劃設言論自由界線的最重要原則之一，我國司法院大法官於 1998 年作成的釋字第 445 號解釋，也以「無明顯而立即危害之事實」作為宣告集會遊行法事前許可制度違憲的理由，將「明顯而立即危險」原則引入我國法制。本節將概略介紹此一原則的內涵、要件，並試圖與監察委員會解釋的「立即危害」判準進行比較。

## 壹、如何認定「明顯而立即危險」？

Holmes 大法官於執筆 Schenck v. United States 案判決時，對於向役男發放反對徵兵傳單、阻礙政府徵兵，涉嫌違反《間諜法（Espionage Act）》的行為，指出「國會是否有權禁止的問題關鍵，在於言論所處的客觀環境，與言論的本質是否會創造明顯而立即危險，並可能導致實際禍害發生（The question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent.）」[199]，咸認為「明顯而立即危險」原則的濫觴。

於此，吾人可知Holmes大法官於本案中認為《間諜法》可限制言論自由的界線，不同於已達成侵害結果的「實害犯」或不問危險發生與否、作出特定行為即該當的「抽象危險犯」，更相似於刑事法上的「具體危險犯」，即行為人因為其言論產生對法益侵害的危險、具備發生實害的概然性時，方能進行處罰[200]。

於其後的 Abrams v. United States 案，由於《間諜法》將處罰標的自「行動」擴張至「言論」本身[201]，Holmes 大法官在不同意見書中，進一步倡議「觀念市場（marketplace of ideas）」理論，強調「檢驗真理最好的作法，就是思想本身在市場競爭中讓自己被接受的能力（the best test of truth is the power of the thought to get itself accepted in the competition of the market）」[202]，更認為作出言論限制時，應同時考量行為人於發出言論時的意圖，是否與可能造成的危險同一，即應包括行為人亦希望透過該言論達成所謂「明顯而立即危險」的結果作為「主觀構成要

[199] *Id.*

[200] 參見謝佳穎，前揭註 57，頁 16-20。

[201] 賴祥蔚（2011），〈言論自由與真理追求——觀念市場隱喻的溯源與檢視〉，《新聞學研究》，108 期，頁 113-114。

[202] Abrams v. United States, 250 U.S. 616, 630, 40 S. Ct. 17, 22, 63 L. Ed. 1173 (1919)



doi:10.6342/NTU202400262

件」，否則恐缺乏言論與該危險間的實際關聯性[203]。

同為美國聯邦最高法院大法官 Louis Brandeis，則在 1927 年的 Whitney v. California 案的協同意見書中，植基於「明顯而立即危險」原則，提出了更為具體的三項構成要件。Brandeis 大法官認為，「若要正當化對言論自由的壓制，必須要有合理根據，懼怕嚴重禍害將會因實現言論自由而產生；必須要有合理根據，相信所擔慮的危險是迫在眉睫的；必須要有合理根據，相信所欲防免的禍害是嚴重的（To justify suppression of free speech there must be reasonable ground to fear that serious evil will result if free speech is practiced. There must be reasonable ground to believe that the danger apprehended is imminent. There must be reasonable ground to believe that the evil to be prevented is a serious one.）」[204]。

在其後逾四十年間，歷經二次世界大戰、麥卡錫主義（McCarthyism）盛行，美國政局對於共產主義的恐懼致使聯邦最高法院幾近揚棄「明顯而立即危險」原則，而傾向限制個人自由以維護國家安全[205]。直至冷戰氛圍褪去，聯邦最高法院於 1969 年的 Brandenburg v. Ohio 案，始重新採納以「明顯而立即危險」為判準，朝向原則上應保護人民言論自由，僅於例外情況才得限制的方向進行修正，此案亦成為當代美國法詮釋「明顯而立即危險」的最重要案例[206]。

Brandenburg v. Ohio 案起源，係因奉行白人至上主義的 3K 黨成員，透過電視呼籲政府不應繼續壓迫白人，並威脅將發動抗議示威，而遭俄亥俄州政府以鼓吹暴力、非法手段進行政治革命起訴獲罪。然而，聯邦最高法院認定「mere advocacy（單純鼓吹）」與「incitement to imminent lawless action（煽動立即的非法行動）」在所不同[207]，因而判決俄亥俄州處罰「鼓吹」的法律違憲，不符憲法第一修正案保障言論自由的精神。

聯邦最高法院進一步將得例外限制言論自由的情形，歸納為三項要件，分別為「主觀上故意」、「急迫性」、「可能性」[208]。「主觀上故意」要求被限制的言論應是為了「直接煽動或產生（directed to inviting or producing）」非法行為；

---

[203] 參見謝佳穎，前揭註 57，頁 26-27。

[204] Whitney v. California, 274 U.S. 357, 376, 47 S. Ct. 641, 71 L. Ed. 1095 (1927)

[205] 參見謝佳穎，前揭註 57，頁 34-47。

[206] Brandenburg v. Ohio, 395 U.S. 444, 89 S. Ct. 1827, 23 L. Ed. 2d 430 (1969)

[207] *Id* at 448-449.

[208] 林子儀、蘇永欽（1990），〈言論自由與內亂罪〉，《憲政時代》，16:1 期，頁 6-19。



doi:10.6342/NTU202400262

「急迫性」體現在該非法行為須迫在眉睫（imminent）；「可能性」則要求該言論須「可能（likely）」煽動或產生非法行為[209]。

三要件乍看明確，但如細究「急迫性」所代表的時間間隔具體為何，是數日、數週之後，還是數個月後的特定日期亦可認作迫在眉睫[210]？要多少機率，才會該當具有實現非法行為、危險的「可能性」？事實上，法院至今恐仍無足夠能力，定義「明顯而立即危險」的客觀性標準並加以適用[211]。

因而，本文參採學說見解，即所謂「可能性」宜參考其他同樣為了防免某些危害發生，而可以侵害個人自由權利的法律規定，譬如警察得基於「合理懷疑（reasonable suspicion）」的情況下，對民眾進行搜身。但在「合理懷疑」的上下，亦有可能性更低的「直覺（hunch）」及較可能的「相當理由（probable cause）」或「實質機會（substantial chance）」，直至「多半可能（more likely than not）」[212]

而本文認為，對於言論自由的限制，既屬人民受憲法保障最重要的基本權利之一，且美國聯邦最高法院對於限制言論自由的法律採取中度或嚴格的審查標準而言[213]，此處「可能性」至少應為學說、實務已有共識的「相當理由」作為門檻，以斷定危害可能會因該言論而產生，國家始得懲罰、限制為妥。

然而相較「可能性」尚可透過其他法律規範進行比較、推論，「急迫性」則顯得更難以定義。蓋因在危害尚未實際發生之前，從言論發出之後可能遭逢的變因之多而無從預測，也可能因為言論本身所欲煽動、產生的非法行為規模不同，而應有不同判斷標準。學者亦莫衷一是，有主張幾小時至幾天者[214]，也有認為危害至多須在幾個月之內發生者[215]，因而最終仍然要回歸法院的個案判斷。

---

[209] *Id* at 447.

[210] 有學者主張，在 1973 年的 Hess v. Indiana 案中，美國聯邦最高法院認為「急迫性」縱然是幾個小時的時間差也不足成立，然本文認為此標準後續既未成為通例，較不具代表性，故不於此贅述。參見謝佳穎，前揭註 57，頁 57。

[211] Feister, D. (2010). How Clear is the "Clear and Present Danger" Test. *Grove City College Law Journal*, *1*(1), 39-50, at 40.

[212] 參見謝佳穎，前揭註 57，頁 64-66。*See also* Healy, T. (2009). Brandenburg in a Time of Terror. *Notre Dame Law Review*, *84*(2), 655-732, at 713.

[213] 參見前揭註 28。

[214] Powe, L.A. (2011). Brandenburg: Then and Now. *Texas Tech Law Review*, *44*(69), 69-83.

[215] Redish, M. (1982). Advocacy of Unlawful Conduct and the First Amendment: In Defense of Clear and Present Danger, *California Law Review*, *70*(5), 1159-1200, at 1180.



doi:10.6342/NTU202400262

## 貳、「明顯而立即危險」與「立即危害」之異同

在比較美國「明顯而立即危險」原則與監察委員會的「立即危害」衡量標準前，宜先指出的是，監察委員會所援引的《拉巴特行動計劃》作為聯合國人權機構發佈、給予各會員國於制定法律規範，限制可能導致暴力、非法行為的仇恨鼓吹時，應參酌的立法建議，也就是說，美國聯邦最高法院的「明顯而立即危險」原則在欲符合國際人權標準的情況下，理應趨同於《拉巴特行動計劃》所設定的六項門檻測試。然而，本文欲討論的主題，並不在於比較美國的言論自由界線與聯合國的國際人權標準有何區別，而是監察委員會化用六項門檻測試後，所允許Meta公司在Facebook、Instagram兩平台上的內容管理處置決定，與美國憲法所允許對可能產生暴力、非法行為的言論在限制手段上的異同之處。

### 一、「主觀上故意」並非監察委員會審查重點

如同本章先前的討論，相較美國在懲罰會導致暴力、非法行為的仇恨鼓吹時，「明顯而立即危險」原則必須考慮發言者是否有「主觀上故意」，也就是其發出言論當下，須具有造成犯罪的故意[216]，發言者始具有可責性，法院將再考量其他兩項要件，以決定該言論是否值得為憲法所保護。

然而，對於監察委員會而言，縱然相較Facebook、Instagram作出內容處置決定，理應擁有更充裕的時間、資源來調查發文者的發佈動機。但從採用《拉巴特行動計劃》六項門檻測試作為「立即危害」標準的案例裁決中，本文發現，監察委員會並未特別關注發文者的主觀想法，多半是透過客觀情境，以心證、推論方式證立發文者存有導致危害的「意圖」。

再以川普案的〈案例裁決 2021-001-FB-FBR〉為例，監察委員會於裁決理由中稱「發生暴力的可能性與川普言論間的關聯顯而易見，因此監察委員會認為他可能知道或應該要知道自己的發言會對合理化或鼓勵暴力行為帶來風險」[217]。然而，就嚴謹的論理邏輯來說，此一說法恐有倒果為因之嫌，乃是因為結果的發生與言論有關，始認定發言者有造成結果的意圖。誠然，美國眾議院委員會在後續

[216] 參見本節第壹項。

[217] 監察委員會，前揭註16。



doi:10.6342/NTU202400262

的調查報告確實認定川普有煽動攻擊國會的意圖[218]，但不可諱言的是，單就川普兩則表面上呼籲和平、實質上可能給予鼓勵的內容，其欲達成的目標究竟為何，終究是政治爭論，而非事實問題。

無獨有偶，在審查柬埔寨首相洪森於 2023 年 1 月、該國國會大選前發佈，指稱將召集黨員抗議、毆打、派黑道前往政治對手家中的言論時，監察委員會雖亦提及參採《拉巴特行動計劃》的六項門檻測試，但實際上並未單獨就洪森的意圖進行分析[219]。在 Meta 公司表示「無法確定洪森在發表這些言論時的意圖」後，仍以公民社會團體、獨立專家意見，及臚列柬埔寨過往發生的政治衝突事件，來推論洪森的發言乃出於威脅的意圖[220]。

如本章第三節所述，本文固然同意，無論是基於經濟或效率上考量，將六項門檻測試以「因素」而非「要件」視之，確實是較為可行的作法[221]。此外，如要求 Facebook、Instagram 在往後移除爭議貼文前，皆須與發佈貼文用戶取得聯絡，以調查其確切意圖，方能符合國際人權標準，則未免過苛。又，相較國家刑罰具有強制性、對人民侵害程度不可謂不大，社群媒體針對單一內容進行移除，雖然仍可能會對言論自由產生影響，但是對於基本權利的干涉則尚屬輕微。因此，以內容所處的客觀情境來推論發言者的意圖，或許是目前對社群媒體來說的合理、務實之舉。

但本文也欲指出，監察委員會於〈案例裁決 2021-001-FB-FBR〉、〈案例裁決 2022-006-FB-MR〉對於《拉巴特行動計劃》的六項門檻測試，皆一一列出進行討論，以確認各指標皆傾向導致「立即危害」的結果，但至近期的〈案例裁決 2023-003-FB-MR〉的裁決理由中，已全然轉為綜合考量，不見包含「意圖」、「程度及範圍」的具體分析，是否代表監察委員會也同樣認為發文者的主觀意圖並非未來的審查重點，甚或無必要納入，則值得往後持續研析。

---

[218] Zengerle, P., Cowan, R., & Chiacu, D. (2022, July 13). Trump incited Jan. 6 attack after 'unhinged' White House meeting, panel told. *Reuters*. https://www.reuters.com/legal/government/us-capitol-riot-probe-turns-focus-trump-allies-extremist-groups-2022-07-12/

[219] 監察委員會（2023），〈案例裁決 2023-003-FB-MR〉， https://oversightboard.com/decision/FB-6OKJPNS3/（最後瀏覽日：01/29/2024）。

[220] 同前註文。

[221] 參見本章第三節第貳項。



doi:10.6342/NTU202400262

## 二、對於「急迫性」的認定標準

不同於美國對於「明顯而立即危險」原則中的「急迫性」要件相較嚴格[222]，甚有試圖釐清言論發出與危險結果間時間差的合理標準，監察委員會在不同案例裁決中，部分似有將「可能性」合為一談，來認定內容是否形成「立即危害」，或部分因內容發佈當下，即在某一事件當下或持續進行的期間中，而據此認定。

舉例而言，在與藥物、健康、管制商品相關的案例裁決，譬如本章第二節提及「宣稱的新冠肺炎療法」的〈案例裁決 2020-006-FB-FBR〉或「熬煮死藤水」的〈案例裁決 2021-013-IG-UA〉，監察委員會分別以「需要處方箋才能取得」[223]、「並未發佈使用死藤水的指示，也為提供如何取得的資訊」[224]，來認定內容與傷害風險間並無立即關聯。但事實上，監察委員會所欲表述的，毋寧說是其他用戶於接觸到爭議內容後，要實現《社群守則》希望防免的危險可能性低，譬如在此兩案例中，即為用戶不易獲得、服食可能會導致健康傷害的藥物，或不太可能在看到死藤水圖片後，逕行熬煮、飲用而致生幻覺或其他人體傷害。

而在監察委員會參採《拉巴特行動計劃》六項門檻測試進行審查的案例裁決中，或因湊巧，爭議內容發佈當下，皆處於暴力、非法行為發生期間，譬如川普案的〈案例裁決 2021-001-FB-FBR〉，兩則貼文分別為暴力示威者攻擊國會當下，以及警方已奪回國會控制權，但群眾仍未散去時[225]；〈案例裁決 2022-006-FB-MR〉中呼籲衣索比亞政府軍投降、否則將死的言論，發佈於該國緊急狀態、武裝衝突進行時[226]；巴西將軍質疑選舉合法性的〈案例裁決 2023-001-FB-UA〉，出現於該國總統宣誓就職後兩日，同時全國範圍內亦有內亂、抗議、呼籲武裝部隊推翻選舉結果[227]；〈案例裁決 2023-003-FB-MR〉則為柬埔寨國會競選期間，且影片發布者、首相洪森有長期透過暴力壓制異議人士的紀錄[228]。

也因如此，監察委員會皆得以輕易認定，內容與可能導致、加劇危害的風險存在立即關聯，因內容發佈的同時，危害亦正在進行中，而無所謂時間差的問

[222] 參見前揭註 210。
[223] 監察委員會，前揭註 50。
[224] 監察委員會，前揭註 72。
[225] 監察委員會，前揭註 16。
[226] 監察委員會，前揭註 191。
[227] 監察委員會，前揭註 193。
[228] 監察委員會，前揭註 219。


doi:10.6342/NTU202400262

題。監察委員會在〈案例裁決 2021-014-FB-UA〉主張，「就其本身而言，無法證實的謠言可能不會直接造成傷害。然而，若這類內容在衝突持續期間出現在重要且具影響力的熱門社群媒體平台上」，就有可能會直接造成立即的生命威脅[229]，似強調內容若發佈在非法行為期間，即滿足六項門檻測試中的「危害可能、急迫程度」。

本文認為，對於具有以上具有高知名度、社會影響力的帳號而言，Meta 公司實傾向寬鬆認定危害的急迫與否。例如〈案例裁決 2023-003-FB-MR〉中，雖然首相雖有威脅其政治對手，但並無到指名道姓的程度，亦未表明威脅可能發生的具體時間，此即與〈案例裁決 2022-007-IG-MR〉的情況類似，理應被認為尚不構成「立即危害」[230]。然而，前者因出於有政治壓迫歷史的非民主國家與威權領袖，加重了危險因素；反之，後者因為音樂形式呈現，減輕了危險可能性，一增一減之下，或成為〈案例裁決 2023-003-FB-MR〉跨越「立即危害」門檻的關鍵因素。

## 第六節　小結

監察委員會以「立即危害」作為 Facebook、Instagram 移除內容是否適當的標準，雖隨著案例裁決逐漸累積，已漸有自成一格的雛形，但仍有許多尚未統一、釐清之處。如同本章先前所討論的，內容須達「立即危害」始得移除，有源自於《社群守則》所明確規定者，亦有監察委員會透過詮釋《公民與政治權利國際公約》、《拉巴特行動計畫》而將「傷害」限縮解釋至「立即危害」者。

本文亦發現，除此一框架外，由於《社群守則》亦有禁止與暴力煽動未必相關的仇恨言論、讚揚恐怖主義等規範，也就是「立即危害」並非監察委員會所認為 Meta 公司在合乎國際人權標準的情況下，唯一能限制社群媒體言論自由的正當原因。事實上，在〈案例裁決 2022-001-FB-UA〉中，我們也可以發現監察委員會明確表示 Facebook、Instagram「禁止『某些歧視性的言論表達』，即使『該言論未煽動暴力或歧視行為』」仍然符合人權責任[231]。

---

[229] 監察委員會，前揭註 192。

[230] 監察委員會，前揭註 160。

[231] 監察委員會（2022），〈案例裁決 2022-001-FB-UA〉， https://oversightboard.com/decision/FB-JRQ1XP2M/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

因此，在社群媒體與國家行為者應負擔的人權義務基準可能不同的基礎認知上，本文分析監察委員會所採用聯合國人權專家所建議的《拉巴特行動計畫》，應是將原先較為嚴謹的六項「要件」，改以「因素」的方式，提供社群媒體往後遵循時，擁有較寬鬆、彈性的綜合評估機會，本文亦從數則相關的案例裁決中發現此一趨勢。

而在美國憲法第一修正案的脈絡下，聯邦最高法院對於言論自由所採取的原則保護、例外限制立場，並以「明顯而立即危險」原則為主要審查標準的架構，並非從未改變，而隨著社會氛圍有所演進，直至當代普遍接受的「主觀上故意」、「急迫性」、「可能性」三要件，卻仍尚未成功發展出能適用於所有案件、一普遍且具體的通則。

然而，對於社群媒體而言，所謂對言論自由的限制，相較國家以刑罰壓制來說，平台所能為不過移除內容、停權帳號而已，雖不啻為對使用者發表意見以干涉性手段相阻，但終究對基本權利侵害較微，人民並非無社群媒體以外之其他管道表達相同意見；一則，在內容數量、管理程序等現實可能性的考量下，社群媒體實無法以最嚴謹的方式，調查、確證內容發佈者的意圖、危害發生的可能性等。

因此，在權衡之下，或許加諸 Facebook、Instagram 等社群媒體平台《拉巴特行動計畫》六項門檻測試的責任，但在各門檻中再尋求符合平台經營效率，且達成預防效果的合適指標，並加以明確不同變因影響綜合判斷的權重，方為持續健全「立即危害」作為限制煽動暴力、非法行為鼓吹內容認定標準的主要方針。


doi:10.6342/NTU202400262

# 第四章　社群媒體言論自由界線的例外：容許差別

《社群守則》作為Meta公司管理Facebook、Instagram兩平台的最主要依據，用戶所發佈的內容如有違反情事，在政策正確執行的情況下，理應被移除、增加警告標語等不同限制。然而，社群媒體上的內容無論是數量、主題皆難以計量，可想見必然可能有某些內容，在特定的情境、背景下，雖不符合平台設下的政策原則，但考量其實際利害後，仍應例外允許而被大眾所見為宜。

## 第一節　「容許差別」起源

2016 年，Facebook 曾以違反《社群守則》有關於兒童裸露的規範為由，移除了曾獲得普立茲獎（Pulitzer Prize）的特寫新聞攝影作品《戰爭的恐怖（The Terror of War）》。該照片呈現一名顯然未成年的女孩，因在越戰時期受凝固汽油彈（napalm，或譯燒夷彈）攻擊全身著火後，撕去身上所有衣物，裸身奔跑呼救的畫面[232]。

Facebook移除該照片的處置決定，隨即引發爭議，一來是此影像在1973年獲頒普立茲獎後，早已聞名、廣傳於世界，似無理由不容於社群媒體；二來是該照片所象徵反戰、呼籲和平的涵義，早已遠超可能引發的兒童裸露、情色聯想。也因此，各界嚴厲批評 Facebook 的作法，如時任挪威首相 Erna Solberg 抨擊，「Facebook在審查這類影像時做錯了，我對這類型的審查說不」[233]。

作為回應，Facebook 不再主張原先「一張裸體兒童的相片通常會被認定違反《社群守則》，甚至在有些國家會被當作兒童色情」[234]的託詞，承諾將會「開始允許更多人們認為具有新聞價值（newsworthy）、值得注意（significant）或對公共利益重要（important to the public interest）的內容，即使它們可能違反了我們的

[232] *See* Ut, N. (1951). *The Terror of War* [Photography]. National Gallery of Art. https://www.nga.gov/collection/art-object-page.136637.html

[233] Hovland, K. & Seetharaman, D. (2016, September 9). Facebook Backs Down on Censoring 'Napalm Girl' Photo. *The Wall Street Journal*. https://www.wsj.com/articles/norway-accuses-facebook-of-censorship-over-deleted-photo-of-napalm-girl-1473428032

[234] *Id.*



doi:10.6342/NTU202400262

守則」[235]。而後，該公司進一步提出所謂「新聞價值容許差別（newsworthiness allowance）」規定，例外允許違反《社群守則》但「具報導價值，且維持顯示符合公眾利益」[236]的內容繼續發佈於 Facebook、Instagram 之上。

## 第二節　與「容許差別」相關的案例裁決

「新聞價值容許差別」既為 Meta 公司所創設，判斷爭議內容能否適用此一規定的要件理應相對容易釐清，亦即「違反《社群守則》」、「具報導價值」及「維持顯示符合公眾利益」三項。然而，如同本文先前討論，違反《社群守則》與否顯然不是非黑即白； Facebook、Instagram 作為社群媒體，如何具有能力，判斷在新聞編輯室中尚且爭論不休的報導價值；若內容可能創造原先《社群守則》希望防免的風險、危害，其同時存在的利益要到何種程度，始得被認為符合公眾利益？

以上問題，皆在在顯示出「新聞價值容許差別」實乃另一個除「立即危害」外的不確定概念。可惜的是，目前 Meta 公司並未公開地具體說明其如何適用此規則，在個案中，我們也只能看到抽象、原則性的文字說明。譬如我國時任立法委員王婉諭曾在 2022 年，於其粉絲專頁上揭露台中一托嬰中心有虐童情事，並上傳一則包含倒吊嬰兒、抹布塞入嬰兒口中的影片[237]。該影片原被 Facebook 認定違反《社群守則》「兒童性剝削、虐待和裸露」及「血腥暴力內容」的相關規範而移除，後則套用「新聞價值容許差別」規則，例外允許該影片保留於平台之上。

但是，由於 Meta 公司並不會針對已適用「新聞價值容許差別」的個案進行解釋，說明該爭議內容的報導價值、公眾利益如何估量，因此在前述影片中，用戶僅可見「這則貼文因可喚起公共意識而獲准顯示」、「我們會在公共意識的價值

---

[235] Kaplan, J. & Osofsky, J. (2016). Input From Community and Partners On Our Community Standards. *Meta*. https://about.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards/

[236] 本文以「容許差別」作為「Allowance」的翻譯，係出於 Meta 公司的繁體中文官方文件。由於 Allowance 直譯通常僅具有允許的意涵，此處譯作「容許差別」，本文認為 Meta 公司有意藉此強調，該允許與一般情形不同，尚須符合特殊條件而有區別。參見 Meta Platforms, Inc. (2023). Our approach to newsworthy content. *Transparency Center*. https://transparency.fb.com/features/approach-to-newsworthy-content/

[237] 王婉諭（2022/02/18），〈【影片畫面令人憤怒，請謹慎點入】〉，Facebook，https://www.facebook.com/wanyu.claire/posts/366228871691357/



doi:10.6342/NTU202400262

和危害風險之間進行權衡」、「我們會評估以下事項：被認為是政治程序的一環而引起爭議的觀點、公共衛生和安全的相關資訊」等語[238]。

也就是說，監察委員會案例裁決中出現多次與「容許差別」相關的案件，遂成為外界分析如何詮釋此規則最重要、關鍵的文本來源。本文欲先指出的是，在〈案例裁決 2022-014-FB-MR〉之前，所謂的「容許差別（allowance）」皆係指「新聞價值容許差別」，然而在該案中，首度出現「政策精神容許差別（“spirit of the policy” allowance）」用語[239]，但其核心概念與「新聞價值容許差別」幾無不同，皆是爭議內容雖違反《社群守則》等內容政策，但因具備公共利益而減少或不予限制，容本文後續再作比較、說明。

準此，本文就截至 2023 年 10 月前，監察委員會所作之 53 件案例裁決中，有提及「新聞價值容許差別」、「政策精神容許差別」等「容許差別」規定的案例裁決共計 16 件，列表如下。

| 裁決編號 | 裁決名稱 | 結果 | 涉及平台 | 涉及政策 | 涉及主題 |
|---|---|---|---|---|---|
| 2021-001-FB-FBR | 美國前總統<br>川普帳號遭停權 | 維持 | Facebook | 危險人物和組織 | 表達自由、政治、安全 |
| 2021-002-FB-UA | 裝扮「黑彼得」 | 維持 | Facebook | 仇恨言論 | 文化、攝影、<br>兒童／兒童權利 |
| 2021-010-FB-UA | 哥倫比亞示威抗議 | 推翻 | Facebook | 仇恨言論 | 社區組織、表達自由、<br>陳情抗議 |
| 2021-016-FB-FBR | 瑞典記者報導<br>未成年人遭受性暴力 | 推翻 | Facebook | 成人裸露和性行為 | 兒童／兒童權利、安全 |
| 2022-001-FB-UA | Knin 卡通片 | 推翻 | Facebook | 仇恨言論 | 歧視、種族和民族、<br>表達自由 |
| 2022-002-FB-MR | 蘇丹血腥暴力影片 | 維持 | Facebook | 血腥暴力內容 | 新聞事件、安全 |
| 2022-005-FB-UA | 新聞報導中提及塔利班 | 推翻 | Facebook | 危險人物和組織 | 新聞學、政治、<br>新聞事件 |
| 2022-006-FB-MR | 提格雷通訊事務局 | 維持 | Facebook | 暴力和煽動仇恨 | 政府、暴力、<br>戰爭和衝突 |

[238] 同前註文。

[239] 監察委員會（2022），〈案例裁決 2022-014-FB-MR〉， https://oversightboard.com/decision/FB-CZHY85JC/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

| | | | | | |
|---|---|---|---|---|---|
| 2022-011-IG-UA | 奈及利亞教堂<br>攻擊後的影片 | 推翻 | Instagram | 血腥暴力內容 | 虐待、安全、<br>戰爭和衝突 |
| 2022-012-IG-MR | 印度性騷擾影片 | 維持 | Instagram | 成人性剝削 | 表達自由、偏鄉社區、<br>新聞事件 |
| 2022-013-FB-UA | 伊朗抗議標語 | 推翻 | Facebook | 暴力和煽動仇恨 | 陳情抗議、政府、<br>性別和性別認同平等 |
| 2022-014-FB-MR | 斯里蘭卡處方藥物 | 維持 | Facebook | 管制商品 | 政府、健康、安全 |
| 2023-003-FB-MR | 柬埔寨首相 | 推翻 | Facebook | 暴力和煽動仇恨<br>預謀傷害和宣揚犯罪 | 選舉、政治、陳情抗議 |
| 2023-004-FB-MR | 亞美尼亞戰俘影片 | 維持 | Facebook | 預謀傷害和宣揚犯罪 | 戰爭和衝突、<br>表達自由、安全 |
| 2023-007-FB-UA<br>2023-008-FB-UA<br>2023-009-IG-UA | 土耳其大選前的<br>政治爭論 | 推翻 | Facebook<br>Instagram | 仇恨言論 | 選舉、新聞學、<br>自然災害 |
| 2023-014-IG-UA | 呼籲古巴婦女抗議 | 推翻 | Instagram | 仇恨言論 | 陳情抗議、表達自由、<br>性別和性別認同平等 |
| 2023-014-IG-UA | 呼籲古巴婦女抗議 | 推翻 | Instagram | 仇恨言論 | 陳情抗議、表達自由、<br>性別和性別認同平等 |

**表 4：與「容許差別」相關的監察委員會案例裁決列表（截至 2023 年 10 月）**

（本表為作者自製）

從上列表觀察，吾人可初步發現，與「容許差別」相關的案例裁決，除〈案例裁決 2022-014-FB-MR〉涉及健康安全、藥品，〈案例裁決 2021-016-FB-FBR〉、〈案例裁決 2022-012-IG-MR〉與性剝削相關，其餘似皆與暴力、仇恨、犯罪事件有關，而多數案件可能也與重大事件，如戰爭、選舉、抗議、武力衝突等有密切聯繫，與「容許差別」應具新聞價值的前提相呼應，社群媒體究應如何認定此一例外原則，本文將於後節進一步討論。



doi:10.6342/NTU202400262

## 第三節　監察委員會如何適用「容許差別」

### 壹、適用「容許差別」的前提與情境

「新聞價值容許差別」第一次被提及，是在前章已有討論的川普案〈案例裁決 2021-001-FB-FBR〉，蓋因美國總統在國會暴動時所發表的言論，雖違反《社群守則》「危險人物和組織」一節中，禁止「對參與違法活動的團體、領導者或個人表示支持與頌揚」的規範[240]，但其作為時任總統、抗議群眾的實質上領袖，在重大關鍵時的發言具有高度新聞價值可想而知。然而，由於川普的兩則內容皆存在引發、延續暴力及非法行為的風險，監察委員會在此案中僅斷言「在需要採取緊急措施以防範重大傷害時，不應優先考量新聞價值」[241]，屏除在此案中適用「新聞價值容許差別」的可能性。

而後的〈案例裁決 2021-002-FB-UA〉雖然也同樣沒有直接援引「新聞價值容許差別」對個案內容進行審查，但卻揭示了所謂「具備報導價值」要件的可能態樣。在該案中，有一位荷蘭用戶發佈了一則帶有影片的貼文，而在影片中，有兩位成人將臉塗黑，扮作荷蘭文化傳統的「黑彼得」[242]。然而，非黑、棕膚色以外的人種「將臉塗黑」如今已經被公認為對非裔或原住民族群具有貶低意味、歧視的行為，譬如加拿大總理杜魯道（Justin Trudeau）年輕時期的黑臉裝扮照片，就曾經在 2019 年引發加國國內政治風暴[243]，此案中的影片也因此遭到 Facebook 以違反《社群守則》中的仇恨言論規範為由移除。

監察委員會對此並無疑問，肯認如內容包含「黑臉妝」，可能會加身對弱勢族群的負面刻板印象，長遠來說，也可能會「導致黑人族群以外的民眾產生種族優越感，從而導致個人將歧視和暴力行為合理化」[244]。然而，本文認為，或許是因為先前曾有杜魯道的案例，如果在該情境下以相同理由，禁止民眾在社群媒體上傳播、討論政治領袖過往的種族歧視行動，未免矯枉過正，一方面失去再次提

[240] 關於〈案例裁決 2021-001-FB-FBR〉，請參見本文第三章第三節第壹項。
[241] 監察委員會，前揭註 16。
[242] 監察委員會，前揭註 52。
[243] Murphy, J. (2019, September 19). How damaging is blackface scandal to Trudeau? *BBC*. https://www.bbc.com/news/world-us-canada-49760160
[244] 監察委員會，前揭註 52。


doi:10.6342/NTU202400262

醒、教育社會大眾為何「黑臉妝」帶有歧視涵義的機會，另一方面更限縮了社會大眾對於公共議題的討論空間，尤其是針對如此具有政治性的攻防辯論。

因此，監察委員會主動提出，「考量到新聞價值，Facebook 若確認內容帶來的公眾利益大於可能造成的風險，便會允許用戶在平台上發佈違反規定的內容」，並舉例「如果公眾人物的黑臉妝相片或影片成為全國新聞報導主題」[245]，顯見此時「黑臉妝」的內容，應是用於譴責或呼籲社會大眾關注歧視、仇恨言論，即符合此一要件。

綜結兩則案例裁決，雖監察委員會至此尚未作成實際審查適用「容許差別」的案件，但已然揭示具備緊急、重大傷害的情境下，不適用「容許差別」；成為全國新聞報導主題的情況下，可能符合「新聞價值容許差別」的「具有報導價值」要件，而可以進入下一階段的檢驗，即權衡公共利益與可能導致的風險，來決定是否應維持該爭議內容在社群媒體上的顯示。

## 貳、考量政治環境與重大事件關鍵時刻

監察委員會第一次對「新聞價值容許差別」進行審查和闡述，出現在〈案例裁決 2021-010-FB-UA〉。該案為哥倫比亞的某新聞媒體，於其 Facebook 粉絲專頁發佈一則影片，報導發生於該國抗議政府的大型示威遊行，其中有抗議人士以西班牙語批評哥倫比亞總統為「hijo de puta（妓女之子）」、「deja de hacerte el marica en la tv（不要在電視上像個同性戀）」[246]。

Facebook 於審查時，考量「marica」一詞是針對同性戀族群的貶義、侮辱性用語，而性傾向是 Meta 公司認定仇恨言論的「受保護特徵」之一[247]，因而決定以該內容違反《社群守則》為由移除。對此，監察委員會於專家諮詢後，肯認「marica」在哥倫比亞已演變成為指涉男同性戀者為「愚蠢（stupid）」、「白痴（idiot）」的歧視性誹謗用語，《社群守則》應予禁止[248]。

---

[245] 同前註文。

[246] 此翻譯參考英文版裁決內容的「son of a bitch」、「stop being the fag on tv」。監察委員會（2021），〈案例裁決 2021-010-FB-UA〉， https://oversightboard.com/decision/FB-E5M6QZGA/（最後瀏覽日：01/29/2024）。

[247] Meta Platforms, Inc., *supra* note 64.

[248] 監察委員會，前揭註 246。



doi:10.6342/NTU202400262

然而，監察委員會主動提出，該內容應該適用「新聞價值容許差別」而保留在平台之上，理由概以「內容發佈時間正值哥倫比亞人民大規模對政府發起抗議，而影片中的口號主要也是對總統的批評」、「示威抗議的目的並不具歧視」，且該影片「是為了引起他人對抗議的關注」[249]。裁決理由進一步指出，「社群媒體承擔向所有人提供平台的重要角色，讓他們能在媒體管道受限的情況下分享抗議相關資訊」[250]，縱然具有侮辱性的詞彙本身對社會毫無正向價值，但讓用戶能在「哥倫比亞政治史上的重大時刻」表達意見，即為該內容所具備的公共利益。

監察委員會並援引聯合國人權理事會特別報告員的意見，主張網路公司在進行仇恨言論的審查、管理時，「對情景的評價可能導致作出例外的決定，其中的內容必須作為例如政治言論來保護」[251]，因此該決定符合國際人權標準。

本文認為，〈案例裁決 2021-010-FB-UA〉揭示了適用「新聞價值容許差別」的幾項因素，包含內容發佈的時間點是否為重要時刻、發佈內容的動機，甚至是當地的政治情況、言論自由程度，都可能納入考量，進而強化維持該內容於社群媒體上顯示的正當性。

此外，違反《社群守則》的內容不論來源是否為新聞媒體，或本身就是新聞內容的一部分，似不會影響判斷合乎「新聞價值容許差別」與否。然而，本件裁決中尚未明確說明的，則是在此情況下，平台應採用何種標準來衡量可能的危害風險與報導價值的公共利益。

## 參、危害風險與公共利益的衡量

在後續的〈案例裁決 2022-006-FB-MR〉[252]中，由於涉及在衣索比亞軍事衝突中，武裝力量對於政府軍的死亡威脅，監察委員會雖然同意該國的新聞自由和公民自由受到限制，社群媒體因而成為傳遞武裝衝突資訊的重要管道。然而，在該則貼文具有前章討論的「立即危害」情形下，監察委員會也同時肯認 Meta 公司對

[249] 同前註文。

[250] 同前註文。

[251] United Nations, Human Rights Council. (2019). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/74/486, at 18. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/308/13/PDF/N1930813.pdf

[252] 關於〈案例裁決 2022-006-FB-MR〉的背景脈絡，請參見本文第三章第四節。



doi:10.6342/NTU202400262

於「新聞價值容許差別」的政策解釋，即該規則並不適用於可能造成人身傷害的內容。此外，監察委員會也並未發現該則貼文「具備足以優先於傷害風險的重大公共利益價值」[253]。

同樣闡釋內容違反政策但具備新聞價值時，如何衡量風險和公益的案件，還有〈案例裁決 2022-011-IG-UA〉。該案為一奈及利亞用戶，在 Instagram 上發佈一段影片，顯示一具滿身鮮血的屍體倒臥在地板上，似為當地一處教堂遭恐怖攻擊的結果，內容發佈者並在影片說明文字中表達該事件令人悲痛，但卻使用數個與真人模擬槍戰「生存遊戲」相關的主題標籤（#hashtag），包含影射槍擊聲的字眼[254]，但裁決理由中並未提及具體用字為何。上述內容，被 Meta 公司認定提及槍擊聲、連結「生存遊戲」的文字說明，實為「對受害者及其社區所經歷的實際悲劇和現實世界的傷害輕描淡寫」[255]，且是次攻擊行動被該公司認定為「造成多人受害的暴力事件和連續謀殺」[256]，適用《社群守則》的「危險人物與組織」政策，禁止任何內容美化、支持或代表該事件，系爭影片因而遭到移除。

然而，在考量奈及利亞非法限制媒體自由的情況下，監察委員會認為，影片本身具有喚起大眾關注國家能試圖壓制的事件，雖然說明文字中同情、嘲弄並存，但未達「虐待」程度。更重要者是，雖然影片可能讓死者及其家屬的隱私權及尊嚴受到傷害，但若能對該影片加上警告畫面，使其觸及人數、互動次數被減少，但仍保留於平台上，「是一項同時尊重表達自由和他人權利且符合比例原則的措施」[257]。也就是說，監察委員會雖然同意內容若含有血腥、屍體的畫面，可能對某些相關人士造成傷害，但如果社群媒體在當地扮演資訊傳遞的關鍵管道，且已經利用警告功能進行提示，避免使用者在無心理預期的情況下接觸爭議內容時，相比民眾知的權利，允許其例外保留於平台上，是權衡之下可接受的作法。

綜上所述，本文認為，監察委員會對社群媒體管理具有新聞價值的爭議內容時，似有將前章討論的「立即危害」納入考量範圍，而其對於公共利益價值的判

[253] 監察委員會，前揭註 191。
[254] 監察委員會（2022），〈案例裁決 2022-011-IG-UA〉， https://oversightboard.com/decision/IG-OZNR5J1Z/（最後瀏覽日：01/29/2024）。
[255] 同前註文。
[256] Meta Platforms, Inc. (2023). Dangerous Organizations and Individuals. *Transparency Center*. https://transparency.fb.com/policies/community-standards/dangerous-individuals-organizations/
[257] 監察委員會，前揭註 254。


doi:10.6342/NTU202400262

斷，在於該國、地區的媒體環境，是否讓社群媒體的角色更為重要，而若能透過平台功能降低危害風險，也可以加強適用「新聞價值容許差別」的正當性。監察委員會在接續的〈案例裁決 2022-013-FB-UA〉引述 Meta 公司的立場，後者稱「若內容可能造成傷害，或可能對公共安全造成直接威脅，即使內容具有某種程度的新聞價值」[258]，仍會移除該內容，似呼應此一結論。

## 肆、目的為引起公眾關注、譴責

本文進一步發現，另一個可能影響案件適用「容許差別」的因素，在於內容本身的目的，是否是為了引起公眾關注或譴責。在〈案例裁決 2021-016-FB-FBR〉中，監察委員會審查一位來自瑞典的用戶，以露骨、寫實的用語，描述兩名不具名的未成年人遭受性暴力的過程，以及加害者被判處的刑期，藉此指摘瑞典刑事司法系統過於寬鬆，應建立性罪犯登記機制云云[259]。

由於系爭貼文具體描述了性暴力過程及對未成年受害者的後續影響，Facebook 認定違反了《社群守則》「兒童性剝削、虐待和裸露」一節中，禁止「威脅、描述、讚揚、支持、提供指示說明、陳述意圖、承認參與兒童性剝削」[260]的內容，決定將該則貼文移除。Meta 公司也在監察委員會審查的過程中主張，雖然貼文並未提及受害者姓名，但相關資訊「可能用於指認受害者的身分，而使其受到歧視待遇」。Meta 公司似乎認為，任何可能導致讓兒童遭受二度傷害的風險都極度嚴重，因此系爭內容雖然重視性暴力事件、呼籲司法改革，但其涉及的公共利益並未大於傷害風險，而無從適用「新聞價值容許差別」[261]。

然而，監察委員會卻作出不同見解，其一方面認為所謂的露骨敘述僅有兩句，篇幅尚不致兒童性剝削程度；二方面則認為「吸引大眾關注並說明此議題，或倡議法律與政策改革所帶來的公共利益，均為『言論自由』理念的核心」，且

[258] 監察委員會（2022），〈案例裁決 2022-013-FB-UA〉，https://oversightboard.com/decision/FB-ZT6AJS4X/（最後瀏覽日：01/29/2024）。

[259] 監察委員會（2021），〈案例裁決 2021-016-FB-FBR〉，https://oversightboard.com/decision/FB-P9PR9RSA/（最後瀏覽日：01/29/2024）。裁決理由中並未呈現該則爭議貼文的完整內容，惟關於露骨描述部分，裁決中有提及該貼文描述加害者稱「那個女孩『很緊』，並驕傲地炫耀他的髒手」等語，可資參照。

[260] Meta Platforms, Inc. (2023). Child Sexual Exploitation, Abuse, and Nudity. *Transparency Center*. https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/

[261] 監察委員會，前揭註 259。



doi:10.6342/NTU202400262

「不應打壓為兒童性剝削發生之倡議者和倖存者的重要性」[262]。基於上述理由，監察委員會似提供 Meta 公司在核發「新聞價值容許差別」時，除了內容涉及壓制言論自由的政治體制、發佈於重大公共時刻等因素外，另一個能強化允許爭議內容例外顯示正當性的背景因素。

其後的〈案例裁決 2022-002-FB-MR〉或可佐證此一觀點，該裁決審查一部於蘇丹發生軍事政變後，一位該國民眾遭暴力攻擊，且頭部受傷、眼睛脫落的影片，在說明文字部分，則有呼籲民眾團結、不要相信軍隊等語[263]。Meta 公司在監察委員會審查該案時表示，其參採先前〈案例裁決 2021-010-FB-UA〉意旨[264]，由於該影片「是在大規模抗議以及對蘇丹新聞自由極為擔憂的背景下所拍攝」，且民眾遭受軍隊暴力攻擊的影片可以「警告該地區用戶，讓其知道自身安全受到威脅」[265]，雖然內容違反《社群守則》禁止血腥暴力內容的規範，但此時該爭議內容的公共利益應該大於危害風險，而獲發「新聞價值容許差別」持續保留顯示於 Facebook 之上。對此監察委員會表示同意，理由略以「在人民空間和媒體自由受到國家箝制的情況下」，Facebook 及 Instagram 理應「允許出於喚起大眾關注或紀錄人權迫害行為之目的而分享人體或屍體的影片」，但仍應對此類血腥內容加上警告，避免使用者在沒有警覺的情況下接觸到令人不安的內容。

同樣因發佈目的為引起公眾關注、譴責，而讓違反《社群守則》的爭議內容得以適用「新聞價值容許差別」的案件，尚有〈案例裁決 2022-012-IG-MR〉。該案件審查標的為一則影片，顯示一位在印度種姓制度底層的「達利特」女性，被一群男子施暴的畫面，影片中雖未顯露出受暴女子臉部或身體隱私部位，但仍違反《社群守則》中禁止性暴力、性剝削內容的規範[266]。監察委員會在該案例中更明確表示，「社群媒體是紀錄此類暴力和歧視行為的重要管道，且用戶發布此案例內容的目的似乎是為了引發大眾關注。因此，該則貼文具有重大公眾利益價值並享有國際人權規範之保護」。

---

[262] 同前註文。

[263] 監察委員會（2022），〈案例裁決 2022-002-FB-MR〉，https://oversightboard.com/decision/FB-AP0NSBVC/（最後瀏覽日：01/29/2024）。

[264] 參見本節第貳項。

[265] 監察委員會，前揭註 263。

[266] 監察委員會（2022），〈案例裁決 2022-012-IG-MR〉，https://oversightboard.com/decision/IG-KFLY3526/（最後瀏覽日：01/29/2024）。



doi:10.6342/NTU202400262

## 第四節　「容許差別」的侷限

行文至此，我們似乎可以大致掌握，Meta 公司最原先為「新聞價值容許差別」設定的三項要件「違反《社群守則》」、「具報導價值」及「維持顯示符合公眾利益」，除「違反《社群守則》」一項未有變動，「具報導價值」的內涵已被監察委員會擴充解釋至需考量內容所涉及的國家政體、發佈時機，基本上可以理解為，在相對不民主、不具新聞自由的國家或地區，爭議內容可能獲得更多保護，但對於平台而言，此判斷標準仍「高度主觀」[267]；而符合公共利益與否，若爭議內容可能導致「立即危害」，即無「容許差別」的適用機會，此外也需將內容發佈目的是否為引起大眾關注與人權、社會事件納入討論。

### 壹、增設「政策精神容許差別」

不無疑問的是，是否可能有其他違反《社群守則》的內容，或許不具有該當「新聞價值容許差別」的報導價值，也與壓迫環境下爭取人權、言論自由無涉，但仍應獲平台例外保護？

本文發現〈案例裁決2022-014-FB-MR〉即創設出與「新聞價值容許差別」有所不同的「政策精神容許差別」，其所審查者，為斯里蘭卡醫療工會在其 Facebook 粉絲專頁上，呼籲大眾在該國金融危機期間捐贈處方藥給該國的貼文[268]。然而，由於《社群守則》「管制商品和服務」一節禁止用戶發佈「索求處方藥物」的內容，該貼文顯然與政策有悖，而應予移除。

當時Meta公司監控斯里蘭卡危機期間的團隊評估，禁止「索求處方藥物」的內容，原先目的在於「促進安全並遏止潛在的有害活動」[269]。易言之，就是為了防止社群媒體成為管制藥品的交易、移轉平台，一方面確保公眾在使用藥品時，皆受專業醫療監督，二方面也減少各國醫藥主管機關要求Meta公司遵守當地法規的壓力。然而，在斯里蘭卡經濟崩潰、藥物短缺之時，移除此類呼籲藥物捐贈的貼文，或許不但無從達成遏止藥物濫用、違法移轉的保障安全目的，反而可能讓某些有藥物需求的民眾無從獲得資源。因此，例外地在需要藥物的國家、地區，

[267] 監察委員會，前揭註 263。
[268] 監察委員會，前揭註 239。
[269] 同前註文。


doi:10.6342/NTU202400262

有限期地允許「索求處方藥物」的內容發佈於平台之上，應為合適作法[270]。

對此，監察委員會表示肯定，並進一步將「政策精神容許差別」定義為「當政策宗旨和 Meta 秉持的理念所要求之成果與嚴格的規則解讀不同時」，Meta 公司可以採取的政策執行手段。

弔詭的是，本文發現與「新聞價值容許差別」不同，除〈案例裁決 2022-014-FB-MR〉外，在所有公開資料中，Meta 公司並未再說明有關「政策精神容許差別」的審查標準。但根據監察委員會的解釋，吾人或許可以理解為，《社群守則》原先作為 Meta 公司對言論自由在社群媒體平台上所劃設的界線，但如果此界線的劃設，反倒使得 Facebook 四項核心價值「真實無欺」、「用戶安全」、「確保隱私」、「尊嚴無損」[271]無從實現，甚或導致相反結果，則應該在特定範圍內進行例外調整。

## 貳、「容許差別」作為極少數案例的解方

本文在耙梳與「容許差別」的案例裁決、相關資料時亦發現，雖然看似在監察委員會的審查下，無論是「新聞價值容許差別」、「政策精神容許差別」似都能被充分地詮釋、適用於不同內容，進而作出應否留存於社群媒體平台上的判定。

然而，監察委員會也多次指出，「容許差別的採用次數相當稀少」，由於 Meta 公司並未在「政策公開透明平台」揭露確切數據，本文僅能從〈案例裁決 2022-012-IG-MR〉提及的「過去一年中，Meta 於全球應用新聞價值容許差別的次數僅 68 次」[272]，和本文統計，單就 2023 年第三季，Facebook、Instagram 就移除了約十三億九千六百萬則內容[273]推知，就「新聞價值容許差別」經常適用在世界各地重大危機、關鍵政治時刻的情況下，符合「具備報導價值」、「符合公眾利

[270] 事實上，在〈案例裁決 2022-014-FB-MR〉斯里蘭卡經濟危機之前，Meta 公司表示亦曾經在古巴經濟危機藥物短缺時期、黎巴嫩藥物短缺且價格過高時期、烏克蘭因俄羅斯入侵藥物供應中斷時，分別核發三個月至九個月不等的「容許差別」，允許藥物捐贈、需求的內容出現於 Facebook、Instagram 之上。

[271] 關於「Facebook 內容政策理念」，請參見本文第二章第一節第壹項及第三節第參項。

[272] 監察委員會，前揭註 266。

[273] Meta Platforms, Inc. (2023). Community Standards Enforcement Report. *Transparency Center*. https://transparency.fb.com/reports/community-standards-enforcement/



doi:10.6342/NTU202400262

益」的爭議內容，恐怕並不能以數則乃至數十則為單位計量。

本文研究者曾實際經手本章第二節所提及，我國時任立法委員王婉諭於 Facebook 粉絲專頁發佈的虐童影像一案，其歷程或可反映為何「容許差別」可能僅為內容管理的理想作法，卻終究僅能成為極少數個案的解方。

回顧該案歷程，雖然揭露托嬰中心虐童情事的影片上傳於 2022 年 2 月，但王婉諭粉絲專頁於 2023 年 5 月才接獲通知，稱某則內容「違反我們有關兒童裸露和兒童性剝削的規則」[274]，且並未給予申訴機會。本文研究者於當日接獲王婉諭立法委員辦公室來詢，始透過內部工具搜尋、定位該則被判定違反《社群守則》的內容，並移交至負責內容審查的團隊進行調查。

由於該則影片的說明文字皆為中文，除需先由以中文為母語的 Meta 公司員工進行分析外，因涉及違反內容政策最嚴重樣態的兒童性剝削，亦須將完整內容翻譯為英文以供安全政策（Safety Policy）團隊確認。幾經來往，時隔近兩月之後，核心內容（Core Policy）團隊始確認該影片雖然「侵犯了兒童的隱私和尊嚴，……且可能使得他們面臨更多的安全風險」，但該內容「譴責虐待，呼籲對加害者更嚴厲的處罰和刑責，……且王婉諭歷有提出兒童權利議題的紀錄，她的聲音對兒童權利保護的辯論有關鍵的影響」[275]，考量台灣媒體已經報導了相同的監視器畫面，在 Facebook 刪除該則影片並不會有更多的安全利益，在為影片加上警告的前提下，應核予「新聞價值容許差別」。

礙於保密義務，研究者未經授權揭露更多與 Meta 公司內部專責兒童安全、內容政策等團隊的溝通細節。但並不難發現，若在多數情況下，一般使用者恐怕未有管道聯繫 Meta 公司人員，遑論讓有關團隊重新啟動調查，一則內容至少需動輒兩個月以上的審查時間，且尚須經在地團隊提供關於發佈內容的使用者、內容的

[274] 王婉諭（2023/05/22），〈午安，有看到這篇文的大家能舉個手嗎？〉，Facebook，https://www.facebook.com/wanyu.claire/posts/pfbid027DFCp48Y7VmJx8pziUk8gfJSmwvyVcjaEFZ4Z92kC6UTAvimtuTA1qqEii3zJJkMl

[275] 受限於研究者的保密義務，本文無法公開該案完整的「新聞價值容許差別」評估理由，但有關引句的原文為「Depictions of non-sexual child abuse infringe on the privacy and dignity of the child. … Child safety experts we have consulted with in the past have flagged that such content could also expose children to further safety risks. … However, the caption accompanying the video condemns the abuse, calls for harsher punishment and penalties against the perpetrators, and advocates for systemic reforms to child rights policies. … Wan-yu Wang has a track record of raising issues concerning child rights… Her voice carries significant importance in the debate on child rights protection.」



doi:10.6342/NTU202400262

背景脈絡，始有機會獲核予「容許差別」並恢復貼文顯示。在在突顯出「容許差別」因需高度的個案判斷，並同時考量人、事、時、地等涉及不同專業的評估人力，目前並無法透過機器系統規模化執行，終究只能成為極少數內容能例外獲得救濟的管道。

## 第五節　小結

於本章討論中，我們了解到無論 Meta 公司如何試圖完善《社群守則》，來平衡社群媒體上言論自由與其他核心價值，終究必然有在原先設想外的貼文、影音內容，雖然表面上不符合內容管理政策，但如對其進行限制，反而無助於保障公共利益，而「容許差別」機制即作為調和的工具，為 Facebook、Instagram 既有的言論自由界線創造例外空間。

本文總結前述討論，認為「新聞價值容許差別」初始的三項要件，在監察委員會所作裁決解釋下，已演變為「三段雙軌」的審查流程。而此一標準不僅可作為判斷「容許差別」的適用與否，也或可作為未來考量爭議內容是否應存在於社群媒體平台上的基本原則。

首先，自然必須先確定爭議內容違反《社群守則》或其他內容政策，才接續進入該內容有無新聞價值的討論。新聞價值的存否，一者建立在該內容涉及的事件是否發生於言論自由受到壓制的國家或地區，也就是社群媒體是否在項公眾傳遞訊息上，扮演關鍵且必要的角色，二者則在於該內容是否為揭露侵害人權的事件，且能夠有效喚起公眾注意、譴責。

若兩者皆非，則該內容將不適用「新聞價值容許差別」，而進入到是否有「政策精神容許差別」適用空間的討論，也就是維持該內容在 Facebook、Instagram 上顯示，能否促成「真實無欺」、「用戶安全」、「確保隱私」、「尊嚴無損」等四項核心價值實現。

第三階段，無論是通過「新聞價值」或是「政策精神」的門檻，最終仍要進入到近似於狹義比例原則的「合乎公眾利益」的比較，也就是內容因違反《社群守則》所創造出的危害風險，相比可能帶來的「新聞價值」或「政策精神」，是否尚屬輕微、可以接受？而若可能的危害風險已達「立即危害」程度時，則應直接認定爭議內容仍屬弊大於利，而不予核發「容許差別」。此一結論與本文第三



doi:10.6342/NTU202400262

章之研究結果，可謂殊途同歸，亦即在社群媒體平台的內容管理上，無論其目的、主題如何，一旦對導致現實世界的傷害具有急迫性且高度概然率，就觸碰了社群媒體上的言論自由紅線，而不見容於平台之上。



doi:10.6342/NTU202400262

# 第五章　結論

在社群媒體致力於開發不同互動功能，透過演算法盡可能搏取閱聽眾使用時間的同時，不容吾人忽視的，是這些跨國科技公司如何透過制定政策，來管理平台上的使用者、行為以及內容。然而，若僅從各國既有的法制觀點來分析，則將忽略這些對言論所作出的限制，其實皆發生在私人的網路平台上，因此理應從私人平台治理和企業自律的角度，來思考在社群媒體時代下言論自由如何被界定與實現為宜[276]。

作為提供網路服務的私人企業，社群媒體一方面不比國家行為者，必須在於法有據且正當性受到嚴格審查的情況下，才能對言論自由作出限制；二方面，由於絕大多數國家並不要求社群媒體為使用者所發佈的言論負責，平台是否提出有效、完善的管理措施，甚或投入資源發展規模化的內容審查執行機制，實際上往往是各家科技公司基於市場考量或社會責任所作出的商業決定。

本文因此選擇擁有世界前二大社群媒體平台 Facebook、Instagram，且政策規範相對完整、公開的 Meta 公司進行研究，透過歸整、耙梳文獻和公開資料，結合研究者自身的實務經驗，針對社群媒體如何管理「內容」提出較具系統性的理解，提出社群媒體的內容管理機制為規範、執行、監督互相建構而成，且「立即危害」及「容許差別」分別為社群媒體言論自由界線提供了原則與例外，再基於研究發現提出未來展望，以下分敘之。

## 壹、交互建構的內容管理體系

研究者認為，若要充分掌握言論自由在 Facebook、Instagram 上如何呈現，至少要從作為管理依據的內容政策、調控手段、政策修正及監督機制等三個面向綜合討論，對應的即是 Meta 公司所制定的《社群守則》及特定主題的政策規範、政策執行方式，及其獨有的監察委員會制度，才能描繪社群媒體內容管理體系的完整圖像。

[276] *See* Klonick, *supra* note 91, at 1662.



doi:10.6342/NTU202400262

本文發現，用於規管兩社群媒體平台最主要依據的《社群守則》，最初為不同禁止內容個案累積、歸納而成的原則，經歷相類於案例法精神的管理模式，才演變為分門別類、臚列限制要件及示例的成文法體系。縱然如此，隨著平台用戶不斷增加，服務擴及到不同的國家、地區，《社群守則》亦有在地化的調整，試圖避免如仇恨言論、騷擾等禁止內容因為語言差異而出現不同的審查結果，且隨著新爭議出現，Meta 公司也對禁止內容，譬如管制商品的類型，有更細緻區分、詳列的趨勢。

除此之外，本文也比較除了《社群守則》外的內容政策規範，得結論在 Facebook、Instagram 兩平台上的內容管理，會隨著形式、主題，甚或在特定期間而有不同的寬鬆、嚴謹程度。由於設下較嚴格限制或額外要求，譬如與平台有對價關係的廣告，或是涉及社會議題、選舉、政治的廣告內容，皆為過往 Meta 公司經常受公眾所關注、批評者。本文認為，此舉可能在一定程度下，協助 Facebook、Instagram 提前減除內容涉及誹謗、虛假、違反當地法令時的連帶責任。而在 2019 至 2021 年新冠肺炎於全球疫情緊張時期，Meta 公司亦加嚴有關內容的管理措施，即在尚未進行事實查核、系統判定可能為錯誤訊息時，就加上警告標籤、調降觸及範圍，藉以遏止可能對社會公眾有害的訊息。

然而，《社群守則》及其他內容政策乍看完整、近似窮舉，但社群媒體平台如何透過機器學習、演算法等科技技術執行，如同法規範縱然明確，亦要一併檢討行政機關、法院對於個案的適用程序，方能全面了解政策是否落實執行。本文呼應過往研究將政策執行概括為「被動、主動」及「事前、事後」兩種分類模式，但進一步分析主、被動的區別，在於是否由 Meta 公司的系統主動偵測作出處置，由於該公司並未揭露系統偵測指標，如「嚴重性」、「傳播性」、「違規可能性」的計算方式，事實上外界並無從得知更細節的運作模式，但主動偵測卻佔所有政策執行的絕大多數；而被動的政策執行，則可能分別由使用者檢舉、團隊接獲通報後啟動。但從監察委員會的案件裁決中，本文亦發現 Facebook、Instagram 在進行內容審查時，實會對更具知名度的帳號提供更多的人工審查資源，已降低因系統解讀導致錯誤判定的可能性，最終目的似仍為維護平台的聲譽及公眾形象。



doi:10.6342/NTU202400262

另一方面，事前、事後政策執行的分界時點，則在於貼文、影音內容是否已成功發佈，而留存於平台之上，且對特定用戶或公開顯示。本文研究者依據自身經驗及台灣發生的過往案例交互佐證，提出 Meta 公司除了會在圖片、影片上傳的同時，進行血腥、裸露的初步審查外，還有全面封鎖被外部安全系統標註為可疑網站的「黑洞」功能，以事前執行阻絕可能高度為有害內容出現於平台上。

Facebook、Instagram 與其他社群媒體最不同處，在於 Meta 公司透過捐助資金、設置信託，邀集不同專業領域、文化背景的學者專家，成立以第三方公開、獨立審查為號召的「監察委員會」，其任務在於審查被用戶「上訴」或由 Meta 公司主動移交認為有疑義的政策執行。本文認為，其審查的前置要件乃至於案例裁決的理由、效果，與美國、我國的憲法審查程序有異曲同工之處，相較於最高法院以憲法作為約束、限制政府或下級法院所為、可能侵害人民權利的法律或判決，監察委員會則是以《社群守則》本身是否明確清晰，有爭議的政策執行是否符合「Facebook 內容政策理念」，以及合乎國際人權標準與否，作為審查 Meta 公司在社群媒體上限制言論自由的舉措是否適當的主要依歸。

Meta 公司及旗下二平台既非國家行為者，理當並不能與後者等量齊觀，而課予相同的人權保障義務，畢竟科技巨擘縱再龐大，終無法適格成為國際條約的簽署成員。但在 Meta 公司自願配合下，監察委員會遂援引聯合國《工商企業與人權指導原則》，將《公民與政治權利國際公約》及意見書、人權調查報告等標準，加諸於社群媒體平台之上，要求 Facebook、Instagram 於限制用戶發佈的內容時，必須明確、可得預見且公開可近用，始具「合法性」；亦須符合《公民與政治權利國際公約》第 19、20 條以維護國家安全、公共秩序、公共健康、禁止仇恨等「正當目的」；並在最後通過「必要性與比例原則」的檢驗，始能避免監察委員會推翻原先移除、封鎖使用者帳號或內容的處置決定。

## 貳、社群媒體言論自由的原則與例外

然而，本文也進一步發現，在截至 2023 年 10 月的 53 則案例裁決中，分別有十餘則裁決理由與「立即危害」、「容許差別」相關，但兩者卻在《社群守則》中，並無具體案例或明確定義。亦即，雖然監察委員會援引兩項原則作為移除內容妥適與否的判斷標準，卻實際上需透過個案的審查，加以定義、詮釋此二不確


doi:10.6342/NTU202400262

定概念的內涵，進而作為 Meta 公司往後審查案件的指引。

依據監察委員會多則裁決意旨，本文歸結出，社群媒體如要判定內容可能導致「立即危害」，單就內容本身的文意、所顯示影像已非足夠，而必須考量內容發佈的背景，譬如對於產生潛在人體危害的療法宣傳，如可能接觸該貼文的使用者無從獲得聲稱具療效的藥物，即無從具備危害實現的可能性而不該當「立即危害」[277]，基於保障言論自由立場，Facebook、Instagram 亦應予以寬容；又或對廣泛、未指定對象和時間的暴力威脅，雖然確實仍有導致現實世界危害的可能性，但仍未達「立即危害」的標準[278]。

至時任美國總統川普於 2020 年國會暴動時，透過 Facebook、Instagram 發佈涉嫌煽動抗議群眾影片，促使監察委員會所發佈的〈案例裁決 2021-001-FB-FBR〉，遂成為監察委員會、Meta 公司後續進行審查內容是否達「立即危害」時的指標案例。於該案中，監察委員會結合社群媒體的特性，操作原先提供給國家立法者於制定限制煽動暴力仇恨言論的人權保障建議，即《拉巴特行動計劃》，從「背景」、「發言者」、「意圖」、「內容和形式」、「程度及範圍」和「危害可能、急迫程度」六個向度對內容可能導致「立即危害」與否進行評估。本文注意到，監察委員會將原先《拉巴特行動計劃》設定的此六項「要件」，轉作為可綜合考量的「因素」，留予平台在未來移除潛在危害內容時，得以綜合考量的彈性與空間，且減少了發言者「意圖」影響判定結果的重要性，作為平衡人權責任與社群媒體有限審查資源的權宜之舉。

本文也透過回顧已在美國成為言論自由限制合憲與否最重要標準的「立即而危險原則」，指出聯邦最高法院在 Brandenburg v. Ohio 案提出的「主觀上故意」、「急迫性」、「可能性」三項要件仍各自在學說、實務界中迭有爭論，尤其所謂「急迫性」是否能有通用、一致的明確時間差，尚且眾說紛紜。本文認為，或許為考量社群媒體在審查內容時的客觀可能性及合理流程，監察委員會所解釋的「立即危害」標準並未如「立即而危險原則」重視「主觀上故意」，且對於急迫與否的認定，似無意作出預測，而以內容發佈於平台當下，與危害風險相關的事件是否已發生或持續進行中為已足。

[277] 參見本文第三章第二節第壹項。
[278] 參見本文第三章第二節第貳項。


doi:10.6342/NTU202400262

如論釐清「立即危害」得以讓社群媒體的言論自由界線更為清晰，甚或是在更明確的解釋下，加以限縮平台管理的任意空間；「容許差別」則為 Meta 公司所自創，由監察委員會逐步補充使其更為具體，得以為《社群守則》保留例外空間，同時維護公共利益的特別條款。「容許差別」原先係為保護少數具有新聞價值，卻違反《社群守則》的個案，例外地允許其繼續留存於平台之上，但社群媒體如何判斷內容具備新聞價值，以及如何衡量可能的公共利益與風險孰重孰輕，則不無疑問。

本文認為，監察委員會結合「立即危害」原則，認定一旦內容該當「立即危害」，則無「容許差別」的適用空間；且若該則內容的圖文素材，已經成為當地新聞報導的主題時，則可視為該當「具報導價值」的要件，而得以進入後階段狹義比例原則的審查。本文亦綜結出「容許差別」的三項討論重點，包含若內容發佈在當地發生重大公眾事件，如集會遊行、經濟危機時，又或所指涉的國家、地區為專制極權、打壓新聞自由的政治環境，監察委員會皆更傾向核予爭議內容「容許差別」，作為協助該資訊有效獲得傳遞的通融之舉；而若該內容涉及揭發某些侵害人權的事件，且發佈目的為喚起公眾注意，監察委員會也較易肯認其具備公共利益，研究者也舉證台灣曾經發生過的實際案例，說明此一原則在監察委員會裁決後，確實為 Meta 公司所採納，作為內部評估爭議內容同時具備公益和弊害時的重要衡量標準。

然而，本文也提出「容許差別」仍有其侷限性，包含監察委員會所發現，並非每則違反《社群守則》但仍值得留於平台之上的內容皆具有新聞價值，因而創設出「政策精神容許差別」，導致在「容許差別」的審查流程上，已演變至本文歸結的「三段雙軌」制，亦即首先要判定內容確實存在違反政策情事，再透過調查內容所涉及的社會環境、目的，判定有無新聞價值，如無，則進入維持該內容顯示是否符合「Facebook 內容政策理念」的審查，以確定可能適用的「容許差別」屬於「新聞價值容許差別」或「政策精神容許差別」，但最終仍然要釐清爭議內容是否可能導致立即危害或其他重大弊害，再與潛在的公共利益價值進行比較。



doi:10.6342/NTU202400262

## 參、對於社群媒體內容管理的未來展望

行至文末，研究者希望再次強調，呼應本文引言與研究動機，吾人應當認知到，當代可能對言論自由作出不當限制的已經不僅僅是國家、政府，如今絕大多數人日常實踐言論自由的場域，也早已非傳統曾受管制的廣播電視、報章雜誌，而社群媒體恰恰是兩者的交會之處，平台如何限制我們的言論，我們又該如何約束平台，仍然會是值得往後研究繼續探索、辯證的重要課題。

同時，本文在從業、研究過程中，也了解到在監察委員會為 Facebook、Instagram 提出建言的同時，Meta 公司似乎也逐漸意識到是項投資，並未為該公司帶來預期「擦脂抹粉」的效果，反而屢屢被監察委員會推翻原先內容處置的案例裁決，突顯出社群媒體平台在內容管理上的不足，而失去持續提供資源支持監察委員會的意願。

然而，監察委員會作為一形式上獨立組織，也似正在尋求能永續、擴大影響力的解決方案。本文設想，雖然各社群媒體有技術、功能上的細微差別，但監察委員會透過個案累積所揭示出「內容應予保留或移除在平台上」的標準，卻可能為不同公司所採，再透過跨平台的合作、協議，將保障言論自由的精神擴及到資源、用戶數量相對較小的社群媒體。然而，此構想的最終挑戰將會是，為何各大科技巨頭會願意服膺一外部監督機制，以及在維運資源來自不同平台的情形下，監察委員會的獨立性是否可以有效維持，皆不無疑問。

再者，社群媒體環境日新月異，本文撰寫當下，Meta 公司旗下的新平台 Threads 甫於台灣獲得廣泛關注，尤其被政黨點名為能與青年族群溝通、替代 TikTok 的重要管道；歐盟《數位服務法》、《數位市場法（Digital Market Act）》相繼上路，明確要求數位平台為有害內容負起管理責任[279]，此一立法舉措會否讓美國跟進，進而根本性地改變原先平台以自律、企業經營角度出發的內容管理模式，發展動見觀瞻。

[279] Shankar, P. (2023, August 25). What impact will the EU's Digital Services Act have? *Deutsche Welle*. https://www.dw.com/en/what-impact-will-the-eus-digital-services-act-have/a-66631337



doi:10.6342/NTU202400262

台灣由於市場有限，近年來各類試圖規管平台的法規，到實際要求外國科技公司負擔責任時，無不是被以錯誤實體（wrong entity）抗辯[280]，要期待社群媒體平台往後全然配合我國立法，恐非務實。然若欲作為亞洲網路治理的先驅，甚或進一步聯合區域各國向科技巨頭施加壓力，則不必妄自菲薄，但對於有志之士而言，最重要的仍然是先掌握平台如何運作、實際管理情形與期待的落差，也是研究者提出本文最重要的目標。

[280] 理由可參照我國高等法院 110 年度上易字第 276 號民事判決，台灣臉書有限公司稱：「臉書社群網站係由依美國德拉瓦州法設立，主要營業地點在美國加州之美國臉書公司所經營。伊與美國臉書公司為個別法人，伊之業務僅有諮詢或業務推廣等臺灣地區之廣告、行銷服務，且無權移除臉書社群網站上之系爭圖文與評論，美國臉書公司刊載系爭圖文與評論與伊無涉。」



doi:10.6342/NTU202400262

# 參考文獻

## 中文部分

### 專書

Frenkel, S. & Kang, C.（著），陳柔含、謝維玲（譯）（2022），《獲利至上：你的一舉一動，都是他們的賺錢工具！Meta 集團（Facebook, Instagram, WhatsApp）稱霸全球的經營黑幕》，野人文化。

Kaiser, B.（著），楊理然、盧靜（譯）（2020），《操弄【劍橋分析事件大揭祕】：幫川普當選、讓英國脫歐，看大數據、Facebook 如何洩露你的個資來操弄你的選擇？》，野人文化。

Levy, S.（著），許恬寧（譯）（2022），《後臉書時代：完整解讀社群霸主從起步、成長、爭議到轉型，每一步的選擇與思考》，天下雜誌。

Lewis, A.（著），林凱雄（譯）（2020），《異見的自由：美國憲法增修條文第一條與言論自由的保障》，八旗文化。

李惠宗（2020），《行政法要義》，元照。

李惠宗（2020），《憲法要義》，元照。

吳信華（2022），《憲法訴訟基礎十講》，元照。

林子儀（2002），〈言論自由導論〉，李鴻禧（等著），《台灣憲法之縱剖橫切》，頁 103-179，元照。

陳澍、陳奕儒、Aidan Hoy（2020），《捍衛台灣選舉公正》，Facebook。



doi:10.6342/NTU202400262

湯德宗（2009），〈違憲審查基準體系建構初探——「階層式比例原則」構想〉，廖福特（主編），《憲法解釋之理論與實務》，頁 1-38，中央研究院法律學研究所籌備處。

## 期刊論文

林子儀、蘇永欽（1990），〈言論自由與內亂罪〉，《憲政時代》，16:1 期，頁 6-19。

陳陽升（2023），〈憲法法庭審查大法庭裁定之裁判效力 〉，《憲政時代》，第 47 卷第 2 期，頁 185-225。

陳憶寧（2021），〈社群平台的治理：從 Facebook 成立監察委員會兼談如何處理虛假資訊〉，《中華傳播學刊》，第 39 期，頁 129-143。

黃昭元（2004），〈憲法權利限制的司法審查標準：美國類型化多元標準模式的比較分析〉，《國立臺灣大學法學論叢》，33 卷 3 期，頁 45-148。

黃銘輝（2019），〈假新聞、社群媒體與網路時代的言論自由〉，《月旦法學雜誌》，292 期，頁 5-29。

劉靜怡（2012），〈社群網路時代的隱私困境:以 Facebook 為討論對象〉，《臺大法學論叢》，41 卷 1 期，頁 1-70。

盧建誌（2023），〈假訊息管制與言論自由的平衡：美國網路中介責任的邊界探察與反思〉，《中華傳播學刊》，第 43 期，頁 153-195。

賴祥蔚（2011），〈言論自由與真理追求——觀念市場隱喻的溯源與檢視〉，《新聞學研究》，108 期，頁 103-140。



doi:10.6342/NTU202400262

## 研討會及學位論文

林彥廷（2022），〈疫情中的社群平台——以 Facebook 的內容政策為例〉，發表於《新興科技應用之法治對策期末研討會》，國立臺灣大學法律學院（主辦），臺北。

張智瑋（2023），《社群媒體假帳號之管制研究》，國立中央大學法律與政府研究所碩士論文。

楊劭楷（2021），《社群媒體假訊息管制之言論自由分析》，國立臺灣大學法律學研究所碩士論文。

謝佳穎（2014），《明顯而立即危險原則之回顧與前瞻》，國立臺灣大學法律學研究所碩士論文。

蘇彥圖（2019），〈美國聯邦最高法院審理流程概論〉，發表於《司法院大法官一〇八年度學術研討會議——憲法訴訟的新時代》，司法院（主辦），臺北。

## 網路資料

中央社（08/19/2022），〈中介法業界歧見大 NCC 暫緩辦理 25 日公聽會〉，https://www.cna.com.tw/news/afe/202208190225.aspx

中央社（12/21/2022），〈媒體數位平台首輪對話落幕 學者籲新聞業統整訴求〉，https://www.cna.com.tw/news/afe/202212210150.aspx

中央社（05/30/2023），〈黃仁勳遭冒名詐騙 陳建仁：感謝公私協力迅速下架〉，https://www.cna.com.tw/news/aipl/202305300111.aspx

中央社（06/06/2023），〈Risu 臉書社團洩私密影像 Meta：已移除違反政策社團〉，https://www.cna.com.tw/news/asoc/202306060325.aspx



doi:10.6342/NTU202400262

中央社（08/29/2023），〈Facebook 抗中國大外宣垃圾訊息戰 刪除逾 7700 個帳號〉，https://www.cna.com.tw/news/aopl/202308290429.aspx

新聞民主與平台議價論壇（2022），〈《新聞媒體與數位平臺強制議價法草案》總說明〉，https://www.twjour.org/post/新聞媒體與數位平臺強制議價法草案-總說明

遠見雜誌（05/09/2021），〈你今天「被祖」了嗎？臉書文字獄引眾怒，揭曉背後審查模式〉，https://www.gvm.com.tw/article/79423

鏡週刊（01/18/2023），〈【色毒滿推特】裸拍、援交、毒趴全不設防　推特成情色犯罪溫床〉，https://www.mirrormedia.mg/story/20230115soc013

鏡週刊（11/28/2023），〈陳柏惟臉書被抄家「200 多篇文消失」　疑被判定色情相關！帳號火速歸隊〉，https://www.mnews.tw/story/20231128nm018

### 判決

臺灣高等法院 110 年度上易字第 276 號民事判決

## 外文部分

### 專書

Knox, J. H. (2011). The Ruggie Rules: Applying Human Rights Law to Corporations. In Mares, R. (Ed.), *The UN Guiding Principles on Business and Human Rights: Foundations and Implementation*. Martinus Nijhoff Publishers.

Lewis, A. (2007). *Freedom for the thought that we hate: a biography of the First Amendment*. Basic Books.



doi:10.6342/NTU202400262

## 期刊論文

Arun, C. (2021). Facebook's Faces. *Harvard Law Review Forum*, *135*, 236-264. https://harvardlawreview.org/forum

Chander, A. (2022). Section 230 and the International Law of Facebook. *Yale Journal of Law and Technology*, *24*, 393-420.

Cowls, J., Darius, P., Santistevan, D., & Schramm, M. (2022). Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance. *New Media & Society*, *0*(0), 1-25.

Douek, E. (2020). "What Kind of Oversight Board Have You Given Us?". *University of Chicago Law Review Online*, 2020, 1-11. https://lawreviewblog.uchicago.edu

Feister, D. (2010). How Clear is the "Clear and Present Danger" Test. *Grove City College Law Journal*, *1*(1), 39-50.

Healy, T. (2009). Brandenburg in a Time of Terror. *Notre Dame Law Review*, *84*(2), 655-732.

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, 131(6), 1598-1670.

Klonick, K. (2020). The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, 129(8), 2418-2499.

Powe, L.A. (2011). Brandenburg: Then and Now. *Texas Tech Law Review*, *44*(69), 69-83.

Redish, M. (1982). Advocacy of Unlawful Conduct and the First Amendment: In Defense of Clear and Present Danger, *California Law Review*, *70*(5), 1159-1200



doi:10.6342/NTU202400262

Reidenberg, J. R., Debelak, J., Kovnot, J., & Miao, T. (2012, April 25). Section 230 of the Communications Decency Act: A survey of the legal literature and reform proposals. *Fordham Law School- Center on Law and Information Policy, Fordham Law Legal Studies Research Paper*, No. 2046230.

Thépot, F. (2013). Market Power in Online Search and Social Networking: A Matter of Two-Sided Markets. *World Competition*, *36*(2), 195-221.

Wong, D. & Floridi, L. (2023). Meta's Oversight Board: A Review and Critical Assessment. *Minds and Machines*, *33*, 261-284.

## 網路資料

BBC. (2018, March 13), UN: Facebook has turned into a beast in Myanmar. https://www.bbc.com/news/technology-43385677

BBC. (2019, November 2). General Election 2019: Facebook takes down 'political' advertising. *BBC News*. https://www.bbc.com/news/uk-50270727

BBC. (2023, November 9). Meta requires political advertisers to mark when deepfakes used. *BBC News*. https://www.bbc.com/news/technology-67366311

Bubna, V. (2022, April 10). How Instagram has played a role in accelerating the Russia-Ukraine war. *Modern Diplomacy*. https://moderndiplomacy.eu/2022/04/10/how-instagram-has-played-a-role-in-accelerating-the-russia-ukraine-war

Carville, O. (2022, November 11). TikTok's Viral Challenges Keep Luring Young Kids to Their Deaths. *Bloomberg*. https://www.bloomberg.com/news/features/2022-11-30/is-tiktok-responsible-if-kids-die-doing-dangerous-viral-challenges

Clegg, N. (2023). How Meta Is Planning for Elections in 2024. *Meta*. https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024/



doi:10.6342/NTU202400262

Clegg, N. (2023). Ending Suspension of Trump's Accounts With New Guardrails to Deter Repeat Offenses. *Meta*. https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/

CNN. (2021, January 7). Trump tells rioters to 'go home' while repeating election lies. *YouTube*. https://youtu.be/3_JxN9CwIMU

Delkic, M. (2022, May 10). Trump's banishment from Facebook and Twitter: A timeline. *The New York Times*. https://www.nytimes.com/2022/05/10/technology/trump-social-media-ban-timeline.html

Dixon, S. (2023, October 11). Most popular social networks worldwide as of July 2023, ranked by number of monthly active users. *Statista*. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Fung, B. (2020, October 22). Facebook's Oversight Board is finally hearing cases, two years after it was first announced. *CNN*. https://www.cnn.com/2020/10/22/tech/facebook-oversight-board/index.html

Gilbert, B. (2020, February 16). Facebook is spending $130 million to create a 'Supreme Court' that can overrule Mark Zuckerberg — here's everything we know about it. *Insiders*. https://www.businessinsider.com/facebook-moderation-independent-oversight-board-supreme-court-mark-zuckerberg-explained-2020-2

Haggin, P. (2022, December 12). Elon Musk's Twitter Disbands Trust and Safety Council. *The Wall Street Journal*. https://www.wsj.com/articles/elon-musks-twitter-disbands-trust-and-safety-council-11670898329

Hovland, K. & Seetharaman, D. (2016, September 9). Facebook Backs Down on Censoring 'Napalm Girl' Photo. *The Wall Street Journal*. https://www.wsj.com/articles/norway-accuses-facebook-of-censorship-over-deleted-photo-of-napalm-girl-1473428032



doi:10.6342/NTU202400262

Kaplan, J. & Osofsky, J. (2016). Input From Community and Partners On Our Community Standards. *Meta*. https://about.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards/

Kelly, S. (2021, October 28). Facebook changes its company name to Meta. *CNN*. https://www.cnn.com/2021/10/28/tech/facebook-mark-zuckerberg-keynote-announcements/

Klonick, K. (2021, February 12). Inside the Making of Facebook's Supreme Court. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court

Murphy, J. (2019, September 19). How damaging is blackface scandal to Trudeau? *BBC*. https://www.bbc.com/news/world-us-canada-49760160

Myers, S. L., Thompson, S. A., and Hsu, T. (2023, October 28). The Consequences of Elon Musk's Ownership of X. *The New York Times*. https://www.nytimes.com/interactive/2023/10/27/technology/twitter-x-elon-musk-anniversary.html

Paul, K. (2023, November 6). Meta bars political advertisers from using generative AI ads tools. *Reuters*. https://www.reuters.com/technology/meta-bar-political-advertisers-using-generative-ai-ads-tools-2023-11-06/

Shankar, P. (2023, August 25). What impact will the EU's Digital Services Act have? *Deutsche Welle*. https://www.dw.com/en/what-impact-will-the-eus-digital-services-act-have/a-66631337

Stecklow, S. (2018, August 15). Why Facebook is losing the war on hate speech in Myanmar. *Reuters*. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

The Facebook Files. (2021, September). *The Wall Street Journal*. https://www.wsj.com/articles/the-facebook-files-11631713039



doi:10.6342/NTU202400262

Tibken, S. (2018, April 11). Questions to Mark Zuckerberg show many senators don't get Facebook. *CNET*. https://www.cnet.com/news/politics/some-senators-in-congress-capitol-hill-just-dont-get-facebook-and-mark-zuckerberg/

United Nations Human Rights Committee. (2011). *General comment no. 34, Article 19, Freedoms of opinion and expression*. CCPR/C/GC/34. https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

United Nations, Human Rights Council. (2013). *Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred*. A/HRC/22/17/Add.4. https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf

United Nations, Human Rights Council. (2018). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/38/35. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf

United Nations, Human Rights Council. (2019). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/74/486. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/308/13/PDF/N1930813.pdf

United Nations, Human Rights Council. (2020). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/44/49/Add.2. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G20/190/89/PDF/G2019089.pdf

United Nations, Office of the High Commissioner for Human Rights. (2011). *Guiding Principles on Business and Human Rights*. HR/PUB/11/04, at 13. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf



doi:10.6342/NTU202400262

Vargas, J. A. (2012, February 17). Spring Awakening. *The New York Times*. https://www.nytimes.com/2012/02/19/books/review/how-an-egyptian-revolution-began-on-facebook.html

Wells, G., Horwitz, J., & Seetharaman, D. (2021, September 14). Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

Wong, B. (2023, May 18). Top Social Media Statistics And Trends Of 2023. *Forbes*. https://www.forbes.com/advisor/business/social-media-statistics/

World Health Organization. (2023, May 5). Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic

Zengerle, P., Cowan, R., & Chiacu, D. (2022, July 13). Trump incited Jan. 6 attack after 'unhinged' White House meeting, panel told. *Reuters*. https://www.reuters.com/legal/government/us-capitol-riot-probe-turns-focus-trump-allies-extremist-groups-2022-07-12/

## 判決

Abrams v. United States, 250 U.S. 616, 630, 40 S. Ct. 17, 22, 63 L. Ed. 1173 (1919)

Biden v. Knight First Amend. Inst. At Columbia Univ., 209 L. Ed. 2d 519, 141 S. Ct. 1220, 1227 (2021)

Brandenburg v. Ohio, 395 U.S. 444, 89 S. Ct. 1827, 23 L. Ed. 2d 430 (1969)

Chaplinsky v. State of New Hampshire, 315 U.S. 568, 62 S. Ct. 766, 86 L. Ed. 1031 (1942)



doi:10.6342/NTU202400262

Cornelius v. NAACP Legal Def. & Educ. Fund, Inc., 473 U.S. 788, 802 (1985)

Knight First Amend. Inst. at Columbia Univ. v. Trump, 928 F.3d 226, 232 (2d Cir. 2019)

Schenck v. United States, 249 U.S. 47, 52, 39 S. Ct. 247, 249, 63 L. Ed. 470 (1919)

Vargas v. Facebook, Inc., No. 21-16499, 2023 WL 6784359 (9th Cir. Oct. 13, 2023)

Whitney v. California, 274 U.S. 357, 47 S. Ct. 641, 71 L. Ed. 1095 (1927)



doi:10.6342/NTU202400262