
Engrave Condition

Apple's Political Censorship Leaves Taiwan, Remains in Hong Kong

By Jeffrey Knockel and Lotus Ruan

MARCH 22, 2022

RESEARCH REPORT #152

Copyright

© Citizen Lab



Licensed under the Creative Commons BY-SA 4.0 (Attribution-ShareAlike licence). Electronic version first published in 2022 by the Citizen Lab. This work can be accessed through <https://citizenlab.ca/2022/03/engrave-condition-apples-political-censorship-leaves-taiwan-remains-in-hong-kong/>.

Document Version: 1.0

The Creative Commons Attribution-ShareAlike 4.0 license under which this report is licensed lets you freely copy, distribute, remix, transform, and build on it, as long as you:

- give appropriate credit;
- indicate whether you made changes; and
- use and link to the same CC BY-SA 4.0 licence.

However, any rights in excerpts reproduced in this report remain with their respective authors; and any rights in brand and product names and associated logos remain with their respective owners. Uses of these that are protected by copyright or trademark rights require the rightsholder's prior written agreement.

Suggested Citation

Jeffrey Knockel and Lotus Ruan. “Engrave Condition: Apple’s Political Censorship Leaves Taiwan, Remains in Hong Kong,” Citizen Lab Report No. 152, University of Toronto, March 2022.

Acknowledgements

We would like to thank Lokman Tsui for valuable feedback. Funding for this research was provided by foundations [listed on the Citizen Lab's website](#). Research for this project was supervised by Masashi Crete-Nishihata and Professor Ron Deibert.

About the Citizen Lab, Munk School of Global Affairs & Public Policy, University of Toronto

The Citizen Lab is an interdisciplinary laboratory based at the Munk School of Global Affairs & Public Policy, University of Toronto, focusing on research, development, and high-level strategic policy and legal engagement at the intersection of information and communication technologies, human rights, and global security.

We use a “mixed methods” approach to research that combines methods from political science, law, computer science, and area studies. Our research includes investigating digital espionage against civil society, documenting Internet filtering and other technologies and practices that impact freedom of expression online, analyzing privacy, security, and information controls of popular applications, and examining transparency and accountability mechanisms relevant to the relationship between corporations and state agencies regarding personal data and other surveillance activities.

Contents

Key findings	1
Introduction	1
Experiment	2
Results	2
Discussion	5
Is Apple required by law to proactively censor political content in Hong Kong using broad, keyword-based filtering?	6
Is Apple attempting to protect its users in Hong Kong from legal repercussions?	7
Is Apple negligent in understanding its own political censorship?	7
Is Apple attempting to appease the Chinese government?	8
Questions for Apple	8
Availability	9
Appendix	9

Read the report's key findings in [Simplified](#) and [Traditional Chinese](#)

Key findings

- › Six months after our [August 2021 study](#) of how Apple filters its engraving services, we reanalyze the filtering system across six different world regions.
- › Since our initial report, we find that Apple has eliminated their Chinese political censorship in Taiwan. However, Apple continues to perform broad, keyword-based political censorship outside of mainland China in Hong Kong, despite human rights groups' recommendations for American companies to resist blocking content.
- › As other tech companies do not perform similar levels of political censorship in Hong Kong, we assess possible motivations Apple may have for performing it, including appeasement of the Chinese government.

Introduction

In previous [work](#), we found that Apple moderates content over its engraving services in each of the six regions that we analyzed: the United States, Canada, Japan, Taiwan, Hong Kong, and mainland China. Across these regions, we found that Apple's content moderation practices pertaining to derogatory, racist, or sexual content are inconsistently applied and that Apple's public-facing documents failed to explain how it derives their filtering rules. Most notably, we found that Apple applied censorship targeting mainland Chinese political sensitivity not only in mainland China but also in Hong Kong and Taiwan. Following our report, Apple [responded](#) that its censorship rules are largely manually curated and that "no third parties or government agencies have been involved [in] the process." Moreover, Apple indicated that their rules depend on each region's local laws and regulations.

In this report, six months later, we perform a similar experiment to see what changes Apple has made to their engraving filtering since our previous study. Notably, we find that Apple has eliminated their Chinese political censorship in Taiwan. However, Apple continues to proactively apply broad, keyword-based political censorship in Hong Kong, despite human rights groups' [recommendations](#) for American Internet companies to resist censorship pressures and law enforcement requests to block content. As other major U.S-based tech companies such as Netflix, Microsoft, Facebook, Google, and Twitter do not apply, proactively or otherwise, similar levels of political censorship in Hong Kong, we conclude our report by assessing possible motivations Apple may have for performing it, including appeasement of the Chinese government.

Experiment

During the week of February 21–27, 2022, we performed the following experiment. We formed a test set by taking the combined set of all keywords that we had previously tested in the United States, Canada, Japan, Taiwan, Hong Kong, and mainland China. We then added to this test set a previously untested set of profane keywords that we found included with iOS 14.8 used to replace profane words with asterisks in captions of dialog with Apple’s automated assistant, Siri (see Appendix for details). Using our previous methodology, we then tested this test set against the same six regions, newly discovering keyword filtering rules as well as which keyword filtering rules Apple removed from their filtering system.

Results

Mainland China, Hong Kong, and Taiwan were the only regions where we observed keywords being removed from triggering filtering, and we observed no major changes to Apple’s filtering in any region except for Taiwan, which removed the following wildcard keyword filtering rules:

Keyword filtering rule	Translation
FALUNDAFA	–
上訪	petition
人民代表大会	People’s Congress
人民代表大會	People’s Congress
修憲	Amend the Constitution
克強	[Li] Keqiang
党和国家	Party and country
共产党	Communist Party
劉鶴	Liu He
国务院	State Council
國務院	State Council
外交部	Ministry of Foreign Affairs
央	Central
孫春蘭	Sun Chunlan
屎	feces
岐山	[Wang] Qishan
最高領導人	Supreme leader
毛主席	Chairman Mao
法輪功	Falun Gong
習書記	Secretary Xi
習皇	Emperor Xi
胡春華	Hu Chunhua
胡錦濤	Hu Jintao

Keyword filtering rule	Translation
胡锦涛	Hu Jintao
解放军	People's Liberation Army
解放军	People's Liberation Army
軍委	Military Commission
近平	[Xi] Jinping
部队	military
部隊	military
黨和國家	Party and country

Table 1: Keyword filtering rules no longer applied in Taiwan.

As in our previous report, we use asterisks (*) in keyword filtering rules to document observed keyword wildcard matching behaviour. For instance, *POO would match “SHAMPOO” but not “POOL”, POO* would match “POOL” but not “SHAMPOO”, and *POO* would match “SHAMPOO”, “POOL”, and “SPOON”. See our [previous report](#) for further elaboration of Apple’s wildcard matching system.

We found that nearly all of these keywords referenced political issues considered sensitive by the Chinese government and that the remaining keywords that we had found filtered in Taiwan were not political in nature. Two possible exceptions are *央* (Central) and *屎* (feces), which may have been removed for being overly broad. Although *屎* is no longer filtered, we newly observed the rules *吃屎* (eat shit) and *屎窟* (shit hole). Using our methodology, we are unable to determine if these rules were newly added or if they always existed but only newly detectable due to being previously subsumed by the broader *屎* rule.

Despite Apple eliminating their Chinese political censorship in Taiwan, we did not observe any reduction to Apple’s political censorship in Hong Kong. However, like with Taiwan, we did see the removal of the two rules *央* (Central) and *屎* (feces). Like with Taiwan, we also newly observed *吃屎* (eat shit) and *屎窟* (shit hole). Thus, it is again likely that that *央* and *屎* were removed from Hong Kong’s filtering for being overly broad.

In mainland China, we continued to observe *屎* (feces) being filtered. However, like in Taiwan and Hong Kong, *央* (Central) was removed. However, we newly observed two more specific keyword filtering rules: *从央视春节晚会衰落* (decline from CCTV Spring Festival Gala) and *裆中央* (Crotch Central Committee, homonym of 党中央 – “Party Central Committee”). Additionally, in the opposite direction, in mainland China we could no longer detect the filtering rules *做愛* (have sex), *性愛* (sex), and *網愛* (Internet love) as they had been subsumed by the broader *愛* (love) rule.

We did not see keyword filtering rules removed from any of the other regions we tested.

CA	CN	HK	JP	TW	US
	AV12电影				
	PORN				
	南京小姐				
	在线看AV				
	复国				
	大奶小姐姐				
	天堂AV				
	广州小姐				
PORN	上门				
	復國	PORN	PORN	PORN	PORN
	愛				
	武汉学生妹				
	海南小姐				
	深圳小姐				
	珠海小姐				
	约妹				
	色妹妹导航				
	长沙学生妹				

Table 2: Keyword filtering rules that we discovered added by Apple.

In each of the regions we tested, we discovered new keyword filtering rules. Some newly discovered keyword filtering rules we can establish that Apple has introduced since our last report, since in that study we had performed tests that would have detected them had those rules already existed. We found that every region tested introduced the keyword filtering rule PORN. However, we only discovered a large number of added keyword filtering rules in the mainland China region, where they appeared to be mostly non-politically motivated (see Table 2 for details).

A	CN	HK	JP	TW	US
BITES	CHINAMEN	CHINAMEN	CHINAMEN	CHINAMEN	CHINAMEN
BRANLER	FAGGOTS	FAGGOTS	FAGGOTS	FAGGOTS	FAGGOTS
CHINAMEN	GOBSHITE	GOBSHITE	GOBSHITE	GOBSHITE	GOBSHITE
COUILLE	SUCKMYDICK*	SUCKMYDICK*	SUCKMYDICK*	SUCKMYDICK*	SUCKMYDICK*
GOBSHITE	*クソ*	*仆你個臭街*	*土人*	*仆你個臭街*	
NIBARDS	*仆你个臭街*	*屎窟*		*屎窟*	
POUFIASSE	*仆你个街*	*扑野*		*扑野*	
PÉTASSE	*仆你個臭街*				
SUCER	*扑野*				
SUCKMYDICK*	*빌어먹을*				

Table 3: Keyword filtering rules that we discovered that we had not previously tested.

Because of our expanded test set, we also discovered many keyword filtering rules that we had not previously tested for. While these may have been added since our last study, they are keyword filtering rules that we detected using new samples in our test set. Since these new samples were derived from a profanity filter applied to Siri captions, these rules are generally related to profanity (see Table 3). These discoveries underscore the need for a comprehensive test set to better understand Apple's censorship practices as applied to engravings. Since the keywords Apple applies to filtering Siri captions, which Apple identifies as a profanity filter, do not include political content, this suggests that, by including political censorship in their filtering of engravings, Apple is including filtering beyond the scope of a simple profanity filter.

CN	HK	TW
从央视春节晚会衰落 (previously *央*)	*吃屎* (previously *屎*)	*吃屎* (previously *屎*)
裆中央 (previously *央*)	*裆中央* (previously *央*)	

Table 4: Keyword filtering rules that we discovered after broader rules were removed.

Finally, we discovered some keyword filtering rules that were revealed after Apple removed a broader rule. In this case, using our methodology we are not able to definitively determine whether Apple introduced these rules after removing the broader rule versus if they had already existed but were undetectable due to being subsumed by a broader rule (see Table 4). However, in our previous report, we had discovered that Apple's mainland China censorship rules appeared largely formed by reappropriating keyword lists from other sources. In Table 4, both the filtering rules *从央视春节晚会衰落* and *吃屎* appear on lists that we previously identified as being keyword lists similar to those that Apple likely copied from when creating their own. In fact, *从央视春节晚会衰落* appears to originally be from a [2003 news article](#), making it especially unlikely that Apple chose to recently add it and may have copied it from a list originally curated over a decade ago. Thus, it is likely that these two filtering rules had always been present but were previously undetectable due to the earlier presence of broader filtering rules.

Discussion

Apple removed its political censorship in Taiwan but continues to proactively apply broad, keyword-based political censorship to its engraving service in Hong Kong. Apple's approach to users' rights in Hong Kong is in stark contrast to its approach in North America, where Apple bills itself as a company [advocating](#) for human rights and [opposing](#) North American law enforcement requests when they are unjust. In the remainder of this section, we assess different hypotheses concerning why Apple chose to cease politically censoring in Taiwan but not in Hong Kong.

Is Apple required by law to proactively censor political content in Hong Kong using broad, keyword-based filtering?

One hypothesis for why Apple ceased political censorship in Taiwan but not in Hong Kong is that Apple is required by law to proactively censor political content in Hong Kong using broad, keyword-based filtering. While such filtering is a requirement for Internet operators operating in mainland China, 2021 Freedom House testimony [argues](#) that Hong Kong's media and Internet space remains different from that of mainland China. There have been numerous instances of content removal attributed to Hong Kong's National Security Law, but they have largely been taken down by authors of the content themselves [without a legal request](#).

Moreover, while [Article 43](#) of Hong Kong's National Security Law authorizes law enforcement to issue deletion requests to Internet operators for content "endangering national security or is likely to cause the occurrence of an offence endangering national security," even authorizing law enforcement to seize electronic equipment if companies do not immediately respond, there exists no language in this Article concerning obligations to proactively apply censorship based on broad filtering rules. Further to this, we are aware of no other major U.S.-based tech company applying automated political censorship to users in Hong Kong. In response to Hong Kong's National Security Law, social media companies [Microsoft, Facebook, Google, and Twitter](#) stopped responding to Hong Kong data access requests related to the law, and [Netflix](#), while acknowledging that the company would acquiesce to deletion requests under the law, stated that they would not engage in proactive review of content. Civil rights organization Freedom House [recommends](#) for American companies to resist law enforcement requests to block content and to not proactively block content in absence of such requests if the content is politically sensitive to the Chinese government.

While the National Security Law is [considered](#) to have given police forces far-reaching power to target content based on vaguely and [broadly-defined national security concerns](#), Apple's censorship system seemingly exceeds even the most liberal of possible interpretations of the law's breadth. In Hong Kong, Apple's censorship often pertains to political and human rights, including keyword filtering rules such as [*信仰自由*](#) (freedom of religion), [*新聞自由*](#) (freedom of the press), and [*真普選*](#) (true universal suffrage). While the National Security Law could be argued to outlaw some uses of these phrases, Apple's system of keyword filtering censors all content containing them. The breadth of Apple's censorship system also encompasses the censorship of any engraving containing [“宗教”](#) (religion), [“共党”](#) (Communist Party), and Party leaders such as [“韩正”](#) (Han Zheng), the Senior Vice Premier of the State Council of the Communist Party of China. While one might imagine that one could make a critical engraving concerning Han Zheng, one could also imagine content that is positive and politically supportive of him, and yet surely no

interpretation of the National Security Law is so broad so as to interpret it as outlawing even the utterance of the names of Communist Party leaders.

Is Apple attempting to protect its users in Hong Kong from legal repercussions?

Another hypothesis for why Apple ceased political censorship in Taiwan but not in Hong Kong is that Apple chooses to proactively perform political censorship in Hong Kong to protect its Hong Kong users from endangering themselves, as certain political expressions in Hong Kong may put users in legal peril or otherwise in danger of government reprisal. Hong Kong's National Security Law authorizes police to target and arrest individuals based on [broadly-defined national security concerns](#). While there is an [uptick](#) in [arrests](#) and police-ordered [censorship](#) under the National Security Law, we are aware of no public policies describing Apple enabling political censorship to protect its users from legal peril or government reprisal. Moreover, we are unaware of any other region in which Apple performs political censorship to protect its users from politically expressing themselves in ways that may put themselves in legal peril or in danger of government reprisal. Further to this point, we are aware of no other U.S.-based tech company applying automated political censorship to users in Hong Kong for this purpose or otherwise.

Is Apple negligent in understanding its own political censorship?

A third hypothesis for why Apple ceased political censorship in Taiwan but not in Hong Kong is that the company is negligent in understanding the content that they censor. Apple has demonstrated that they have a poor understanding of what content they censor and the process for which they determine what they censor. For instance, Apple [claims](#) that they rely on no third parties for information on what to censor and that their methods rely largely on manual curation. However, in our [previous work](#), we present strong evidence that Apple thoughtlessly reappropriated their censored keywords, effectively copying and pasting ranges of keywords from other sources, and we identify lists used in other companies' products which are or are related to those from which Apple copied. Many of the terms that Apple censors seem originally intended to censor in-game chat, others outdated and related to news articles written over a decade before the advent of Apple's engraving service, and others yet have no clear censorship motivation whatsoever except that we found them in a range of keywords in third party lists from which Apple appears to have carelessly copied.

While Apple reappropriating ranges of censored keywords from other sources is an already problematic process for how Apple decides what to censor in mainland China, our previous work presented strong evidence that Apple subsequently formed Hong Kong and Taiwan's keyword filtering rules by pruning rules from keywords derived in

the above fashion. Together, this problematic process would explain how Chinese political censorship and other content which Apple poorly understands could have slipped into both Taiwan and Hong Kong's lists, and our finding that Apple no longer politically censors in Taiwan would appear to be a tacit acknowledgement of Apple that their political censorship in Taiwan was negligent. However, despite Apple being equally informed of their political censorship in Hong Kong as that in Taiwan, our findings in this report show that they have not similarly abandoned it.

Is Apple attempting to appease the Chinese government?

Finally, a fourth hypothesis for why Apple ceased political censorship in Taiwan but not in Hong Kong is that the company is attempting to appease the Chinese government so that Apple can retain mainland Chinese market access. In the same fashion that Apple proactively censors engravings using automated methods, previous reporting has [revealed](#) that Apple, to appease the Chinese government, also proactively censors its App Store using automated methods. Furthermore, other reporting shows that Apple, to appease the Chinese government, engaged with the government in a [five year deal](#), in which Apple pledged to invest over \$275 billion (USD) in the Chinese economy, including research labs, retail stores, and Chinese tech companies as well as agreeing to use more Chinese suppliers and software. In this deal, Apple also pledged to "strictly abide by Chinese laws and regulations." It is unclear, however, how Apple interprets or complies with Chinese laws and regulations especially when those regulations are vaguely and broadly defined. To fully assess the appeasement hypothesis, we would require clear explanations and transparent records of Apple's deals with the Chinese government, as well as answers concerning which laws or regulations Apple might believe to be obligating their automated political censorship system. However, in lieu of that information, we find the hypothesis that Apple proactively applies political censorship to Hong Kong users solely to appease the Chinese government plausible.

Questions for Apple

On March 11, 2022, we sent a letter to Apple with questions about Apple's censorship policies concerning their product engraving service in Hong Kong, committing to publishing their response in full. [Read the letter here.](#)

As of March 22, 2022, Apple had not responded to our outreach.

Availability

The keyword filtering rules we discovered for each of the six regions analyzed are available [here](#).

Appendix

In this report, we included keywords in our test set extracted from the list of keywords Apple developed to filter captions of dialog with Apple’s automated assistant, Siri. In this appendix, we further detail the function of these keywords and how we extracted them.

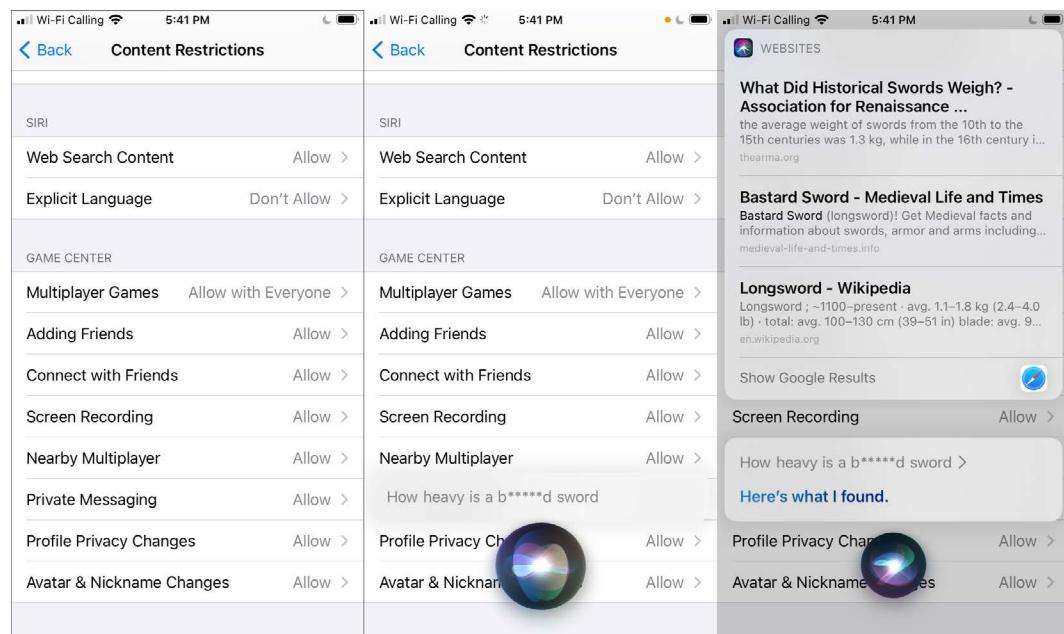


Figure 1: Left, enabling explicit language filtering for Siri; center, asking Siri a question containing a filtered keyword; right, viewing Siri’s response.

Analyzing iOS 14.8, we found that Apple includes functionality to partially replace explicit language in Siri dialog with asterisks. This functionality can be enabled by setting “Explicit Language” to “Don’t Allow” in the “Content Restrictions” settings (see Figure 1). By using reverse engineering methods, we discovered that this filtering is implemented in Apple’s DialogEngine framework in the function

```
siri::dialogengine::GetProfanityFilter(const std::string &).
```

We found that this function takes an ISO language code as an argument and that each language has its own list of filtered keywords. Furthermore, we found that each filtered keyword is associated with another replacement string, which we found to be some

version of the filtered keyword where some or all characters are replaced with some number of asterisks. We extracted the filtered keywords and their replacements for each language using a Frida script we developed. We make available [here](#) the Frida script and the extracted keywords.

Like Apple’s filtering of engravings, we found that Apple’s Siri dialog filtering applied different lists to different regions depending on those regions’ languages and customs. Unlike Apple’s filtering of engravings, we did not find Apple’s Siri dialog filtering to include Chinese political censorship. This finding is unsurprising, as, unlike engravings, Siri dialog cannot be used to communicate with other people and, unlike Apple’s filtering of engravings, Apple’s filtering of Siri dialog is optional and disabled by default. However, these filtered words, which Apple identifies as “explicit language” and “profanity,” serve as a contrast to Apple’s treatment of engravings, as they demonstrate how Apple’s political censorship of engravings goes beyond simply controlling explicit language and profanity.

