A News Outlet Sample

A.1 Sample

To construct the sample of 145 Egyptian news websites, I relied on *ABYZ Web Links*,[1] a community-based list gathered by the *Citizenlab*,[2] and outlets collected by the Egyptian NGO *Association of Freedom of Thought and Expression* (AFTE 2021). I did not consider 16 websites that were created after May 2017 since they display missing values in my outcome variable until their establishment and I cannot estimate their pre-May 2017 size as well as identify the exact creation date. I gathered this information by searching for domain information using the Internet Corporation for Assigned Names and Numbers and website availability relying on `archive.org`, also known as the Waybackmachine. Finally, as mentioned in the main text, I focused on websites that were identified to have a primarily Egyptian audience and scope. Tables A.1 and A.2 show the sample of Egyptian outlets including all collected variables. In Appendix H, I also present the 78 regional and international outlets that were mentioned by the used sources.

A.2 Variables

Apart from the identifying variable **domain name**, the dataset of news outlets includes the following variables that were collected by visiting archived versions of the websites,[3] current versions, or relying on secondary sources such as the outlets' social media channels or publicly available information.

Alternative name lists alternative domain names if the original domain name was changed during the period of study.

Stance judges from the website's content, including self-description, and funding structure to ascertain the stance of the outlet towards the Sisi regime. This variable is either

[1] `http://www.abyznewslinks.com/`
[2] `https://github.com/citizenlab/test-lists`
[3] For this task, the research assistant and I relied again on the Waybackmachine.

coded as pro-government (including outlets uncritical of the government), Islamist opposition, independent or the residual categories sports/entertainment or "other sectarian." The research assistant coded an outlet as independent (default category) if it is not clearly related to the Islamist opposition or promoted pro-government messages.

Funding source lists the funding organization if this information was available on the website or otherwise publicly available. The variable is either coded as state-owned/backed, party- or group-run, or private if there was no indication of support through a state or party or group.

Institution of funding contains the organization or country of funding if the funding source is other than private.

HQ location contains the country of the outlet's main editorial board location if revealed on the website or otherwise publicly available (not the location of the outlet's server).

Main audience/scope was judged by analyzing the website's content or relying on information on the main audience in the "about the newspaper" section, if available. Of particular interest was whether the outlet predominantly reports on Egypt-related news to create the main sample. The variable is either coded as Egypt, regional, international, or another Middle Eastern country.

Media type describes the type of website and is coded as digital media, TV/Radio, former periodical newspaper or magazine, periodical newspaper or magazine, or news agency.

Blocked includes information about whether the website was blocked during the period of study. The variable can have the following values: not blocked, blocked, or potentially unblocked. The latter value was chosen if the website was listed in the blocking list by

AFTE but the website could be accessed in Spring 2021 from Egypt.[4]

Status describes the status of the website as checked by the research assistant (from within Egypt) and the researcher (from outside Egypt) by visiting the website in March 2021 and relying on information from AFTE. The variable can have the following values: working, closed.

Date of blocking includes the date of the first blocking as determined by AFTE.

Active social media is coded as "yes" if the website has an active social media account either on Facebook, Twitter, or both, in early November 2021 or "no" if this is not the case. A social media channel is defined as active when the last social media post is not older than two months (i.e., early September 2021).

Average pre-May 2017 traffic is calculated as the average reach per million before the first censorship wave on May 24, 2017, relying on the Alexa traffic data (see Appendix C for details on the traffic data).

A.3 Ethical considerations

Since the checking of the website could be potentially risky for the Egyptian research assistant, we undertook several steps to reduce this risk: First, before the coder checked for the availability and blocking of websites, I went through the list of domain names to check whether websites (if found) contain phishing or malicious content. While some indeed ended up showing "advertisement" content, no malicious or suspicions content was downloaded. Second, the coder should only very briefly stay on the website to check whether it is reachable. The classification of websites was mostly done relying on the Waybackmachine that contains archived versions of websites and ensures encrypted communication.

[4]I double-checked these cases by examining recent measurements by OONI (when available) for the respective website. Websites that became inactive and were on the blocking list are assumed to having been blocked until they ceased activity. Unfortunately, given the sparse censorship data, it was not possible to determine the exact blocking duration.

Overall, and after speaking with the research assistant about potential risks, the coding was deemed relatively safe. Reassuringly, OONI, which also relies on worldwide volunteers to measure internet censorship, is not aware of any incident of a volunteer being prosecuted when using their tool to check whether websites are blocked.