Subtle Censorship via Adversarial Fakeness in Kyrgyzstan

Christopher Schwartz
KU Leuven
Centre for Ethics, Social and Political Philosophy
christopher.schwartz@kuleuven.be

Rebekah Overdorf
École Polytechnique Fédérale de Lausanne
Distributed Information Systems Laboratory
rebekah.overdorf@epfl.ch

1. SUMMARY

With the shift of public discourse to social media, we see simultaneously an expansion of civic engagement as the bar to enter the conversation is lowered, and the reaction by governments and others to silence these voices. Since traditional forms of censorship struggle in such a situation to enforce the preferred narrative of those in power, adversaries of free speech have developed new methods for controlling the conversation that use the social media platform itself. Using the Central Asian republic of Kyrgyzstan as a main case study, this talk explores how this new form of “subtle” censorship relies upon pretense and imitation, and why interdisciplinary methods of research are needed to grapple it. We examine the role that “fakeness” in the form of fake news and profiles are used by as methods of subtle censorship.

2. ADVERSARIAL FAKENESS

We examine fakeness at two levels: a) Fake news is a deliberate presentation of either misleading or untrue claims as journalistic reporting. b) Fake profiles are digital entities in which either i) the persons with the alluded identities did not really create and/or control the profile, or ii) the identities are utterly fabricated and hence cannot be attached to any real person. In both cases, what characterizes “fake” is adversarialness: the active intention to falsify or mislead for purposes beneficial to the agent and harmful to the target.

Recently, fakeness has been associated with “hybrid warfare”, but this framing is too military-oriented to encompass the phenomenon. Civilian activities such as election and marketing campaigns are increasingly deploying fakeness in order to sway public opinion and consumer interest against rival politicians and brands. Hence, fake news and profiles would be better conceptualized in broader terms of “adversarial fakeness”, i.e., when malicious deception occurs as a direct action by an agent seeking to harm a target. This framework enables a fuller vision of the threat the world is facing. As a form of hybrid warfare, adversarial fakeness inflames inter-communal grievances and mobilizes populations against each other. As a form of disinformation, it undermines the integrity of journalism, elections, and markets. As a form of censorship, it drowns out moderate and alternative voices, and targets journalists, civil society organizations, opposition politicians and human rights activists.

3. SUBTLE CENSORSHIP

Adversarial fakeness itself can be conceptualized as “subtle censorship.” Insofar that censorship can be considered at root as any attempt to silence information, scholars traditionally taxonomize the phenomenon in terms of withholding information, destroying information, altering or using selective information, and self-censorship, as well as being explicit or implicit, direct or indirect. Subtle censorship is different in that it relies primarily upon pretense and imitation to silence information. As a result, it obfuscates whether censorship is even happening.

Traditional forms of censorship are attractive for scholars and activists because they imply their own solutions. In philosophical terms: if there is a censor censoring, then Q counter-censor by circumventing (re. withholding information), archiving (re. destroying or manipulating information) or whistle-blowing or reforming (re. self-censorship). In contrast, subtle censorship eludes measurement, does not imply its own solution, and has three unusual forms:

Silencing through attacks.

Fake accounts are actively used to directly attack journalists and activists through social media posts and comments, the primary purpose of which is to intimidate individuals or groups into silence and submission. The example of the reporter Nedim Turfent is highly illustrative. In 2016, while on assignment in Turkey’s Kurdish region for the pro-Kurdish Dicle News Agency, Turfent published video of soldiers threateningly standing over villagers, who were face down with their hands bound. Soon afterward, Twitter accounts that were later discovered to be linked to Turkish counter-terrorism units began taunting locals with the question, “Where is Nedim Turfent?” In the words of Bloomberg magazine, “The threat was clear: Give him up, or you’re the next target.” Within days, Turfent was in military custody and charged with membership in a terrorist organization. The Twitter-based manhunt abruptly ceased when one of the accounts tweeted a photograph of Turfent in custody. He remains in jail.

Silencing through noise.

An effective method of censoring opposing viewpoints is to simply drown them out. For example, automated and semi-automated fake accounts are frequently deployed to flood social networks with fake news and inflate content views, to the point that opposing perspectives appear to be in the minority or disappear down the newsfeed. As there is often an asymmetric distribution of resources favoring those

1Separate from the political views of Turfent himself, what is relevant about this case is how a journalist has been targeted by fake accounts.
already in power, this type of censorship can be quite effective and difficult to counter. “Troll armies” are the example 
per excellence of this form of censorship. However, more sophisticated tactics than just sheer volume exist. For ex-
ample, “hashtag poisoning” attacks, in which bots or fake accounts tweet a trending opposition hashtag repeatedly in
order to trick Twitter into considering the hashtag as spam, have been used in Mexico and elsewhere to effectively censor
anti-government discussions [2].

Silencing through misinformation.

Similar to silencing through noise, fake accounts are also used for making audiences unreceptive toward real news al-
together. For example, Kyrgyzstan was destabilized in June 2010 when interethnic clashes between Kyrgyz and Uzbeks
were sparked by fake news about Uzbek gangs sexually assaul-
ting Kyrgyz women and Uzbek separatists attempting to
join neighboring Uzbekistan. In the years since, fake news
has claimed Uzbeks are undermining the “Kyrgyz character”
of the state and are spreading radical Islam in the country.
Fake news also promotes the idea that Uzbeks and the in-
ternational community are engaged in an “information war”
against the Kyrgyz nation [4]. This engenders a defensive
mindset throughout Kyrgyz audiences, rendering it impos-
sible to discuss alternative views about the 2010 events.

4. BACKGROUND AND CONTEXT

This talk focuses on Kyrgyzstan, which is on the front-line of adversarial fakeness and subtle censorship. In a region
notorious for authoritarianism, Kyrgyzstan boasts a gener-
ally free market and parliamentary and presidential elections
held according to a schedule and rules, not the whims of rul-
ing elites as happens among its neighbors. The quality of
elections is debatable due to widespread vote-buying and -
 extortion, but the 2017 presidential election was considered
genuinely unpredictable by international observers. In 2018,
the Economist Intelligence Unit assessed Kyrgyzstan as a
“hybrid regime” on a trajectory toward becoming a “flawed
democracy” (the same category as the United States).

As such, Kyrgyzstan stands out among its neighbors as
having a genuinely free mediascape, with very little “ex-
 plicit” censorship. The capital, Bishkek, acts as a haven
for journalists from across Central Asia, and the Interna-
tional Exchange and Review Board has assessed Kyrgyzs-
tan’s domestic scene as “near sustainable” in its 2019 Media
Sustainability Index. Yet, self-censorship is a rampant prac-
tice among local journalists [3], partially due to a history of
periodic violence against the profession, and partially due
to ruling elites exploiting the weak judicial system to target
the press with inflated slander lawsuits.

Throughout Central Asia, state control of ISPs is the
norm, whether by overt monopolization or covert means.
Kyrgyzstan not only lacks such control, but it also has “no
functional legal and technical provisions exist for shared use
of existing infrastructure by ISPs, forcing them to build re-
dundant and expensive infrastructure” [1]. Although the
state-owned KyrgyzTelecom remains the largest ISP within
the Kyrgyz market, this situation nonetheless poses quite a
hurdle to potential state surveillance and censorship.

With respect to adversarial fakeness, geostrategic issues
crucial to Kyrgyzstan’s sovereignty are the target of cam-
paigns involving fake news and profiles. In 2015, fake news
about an imminent nuclear war between the United States
and Iran, as well as about American support for “gay prop-
ganda” and Uzbek separatism, rallied public support for
the pro-Russian Kyrgyz government to shut down an Amer-
ican military base and cancel an important 1991 bilateral
treaty. Since 2018, fake news about Chinese “imperialism”
has fuelled violent anti-Chinese demonstrations throughout
the country, risking another crucial relationship. Demonstr-
ators have called for bans on Chinese citizens from doing
business in Kyrgyzstan and marrying Kyrgyz women.

5. NEW CENSORSHIP, NEW METHODS

As a problem to be combated, social media platforms are
actively seeking solutions to adversarial fakeness and subtle
censorship. For example, Facebook has publicized the fake
accounts they have shut down, a majority of which are shut
down at the time of the account creation. The effective-
ness of their methods, however, can only really be evaluated
internally and these platforms lack incentives to enforce con-
tent moderation relating to fake news.

Although proposed solutions rely on machine learning meth-
ods, there are two problematic trendlines: a) No unifying
framework: Researchers focus on detecting fake news or fake
profiles as interconnected yet still distinct phenomena, i.e.,
using one to detect the other, but not really conceptualizing
them as two sides of the same coin (adversarial fakeness),
and as occurring within a larger practice (i.e., subtle cen-
 sorship)b) No human focus: Related to the lack of a uniting
framework, researchers also tend to focus on the fake
news propagation role of fake profiles, rather than on the
re-propagation role of audiences.

Our project in Kyrgyzstan broadens this research by com-
bining machine learning, philosophy, and journalism. With
our partners the Civic Initiative for Internet Policy, a Kyrgyz
civil society organization, we have assembled a team of jour-
nalists and researchers in Bishkek who are assisting us in the
following. i) Establishing ground truth by manually identi-
fying a) fake accounts that seed fake news, and b) human
beings that serve as re-propagators. ii) After crawlers de-
rive social graphs on both populations, deriving feature sets,
then training classifiers upon these features. iii) Validating
results journalistically by a) utilizing context knowledge to
critically analyze the fake news samples and b) endeavor-
 ing to confirm claims made by various sources regarding the
origins of the examined fake profiles.

6. REFERENCES

[1] 2016 ICT Sector Overview for Kyrgyzstan. DR
Chaos Communication Camp.
K. Winneg. “The Spiral Of Silence on Social Media:
Cultures of Self-Censorship Online and Offline in
Kyrgyzstan”. Internet Policy Observatory, 2018.
Narrative: Kyrgyz and Uzbek Narratives in
Kyrgyzstan”. Master’s thesis, Central European
University, 2015.