Twitter Thread by Giovanni Pagano

Giovanni Pagano
@giovanni_pag

What are the implications of using hacked data for research?

A short thread inspired by the fact that, before AWs took it down, #Parler was extensively hacked and user data was leaked.

The #Parler dataset seems crazy interesting for doing research, and my first reaction after the breach was to shre it with other #CompSocSci ppl.

However, I started having second thoughts, so what follows is to organize ideas and have it somewhere I can look back to.

2/n

Generally speaking, as far as the ethics of research goes a good advice would be to handle hacked data with caution.

First of all, there's an issue of quality. Data might be altered or incomplete, and the source cannot be considered accountable (assuming src is anonymous).

3/n

Secondly and more importantly, a researcher using the data would probably be violating users' consent and acting against the data collector's will.

Finally, users' privacy is at stake, since researchers could see material that users didn't agree for other people to see.

4/n

Sharing private information without consent might put people at risk of harm.

This is all the more true in cases such as the #ParlerHack, where the leaked information is of particularly sensitive nature, and there's a high risk of unintended consequences.

5/n

However, it can be argued that in many cases the milk is already spilled.

After all the data is out there, users are already exposed, and using the leaked information for rsrch (with some precautions) might not cause any additional harm.

Does this mean free for all then?

6/n

Short answer, I am not sure.

On practical grounds, there might be legal boundaries in place (depending on the context).

But more generally, from a deontology perspective I think that (as long as the resercher is not responsible for the hack) the picture is blurred.

7/n

Sure, the issue of privacy when data is out in the open becomes secondary. Plus, data can be anonymized by the researcher, so that private information is not furtherly disseminated.

On the other hand, I think the problem of users' consent should not be bypassed as easily.

8/n

There's also another issue.

In fact it can be argued that using illegally obtained data for research purposes might legitimize (or even encourage) illegal or unethical behavior.

9/n

Ultimately, the fact that data is publicly available data it doesn't mean neacessarily that it is available for research, and some of the arguments against its use are hard to dismiss.

Do you know of any explicit guidelines in poli /soc sciences that address this issue?

n/n

cc @therriaultphd @ylelkes @conjugateprior @cjw_phd

https://t.co/IvcTXARoga

> If I were a grad student in poli-sci /social psych / etc right now, I'd be getting my advisor to let me change my dissertation topic right now. The potential of this dataset is insane. https://t.co/0F5o6tUdYu
>
> — Andrew Therriault (@therriaultphd) January 11, 2021