# Some Observations on Fact Checking Work

Rob Procter

Warwick University

rob.procter@warwick.ac.uk

# Previous work

- Verification is not a 'one stop shop' but a concern that becomes relevant again and again throughout process of writing a story.
- Tools need to support verification as ongoing work, including making visible what verification work has already been undertaken.


https://www.pheme.eu/

swissinfo.ch

Tolmie, P., Procter, R., Randall, D. W., Rouncefield, M., Burger, C., Wong Sak Hoi, G., ... & Liakata, M. (2017). Supporting the use of user generated content in journalistic practice. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3632-3644).

# Methodology

- Series of semi-structured interviews with fact checkers.
- Mainstream news organisation
- FullFact, fact checking organisation:
  - Fact checker
  - Data scientist
  - Editor

# Surfacing

"Lionel Shriver wrote an article in The Spectator in which she said that the vaccines can't stop Corona virus spreading, and you could sort of interpret that in two different ways, you could say: 'Well, not every single person who gets vaccinated is perfectly protected from catching or spreading it so vaccination doesn't completely stop all examples', but maybe that isn't what she meant. Maybe like 'it can stop it because there are loads of times in an individual case when it does stop it from spreading, someone doesn't catch it in the first place'. So which is the meaning of that sentence that we would be checking? In the context of that article, I think it was clear she was saying the second one, she was saying that it doesn't make any difference at all to reducing the spread, and that is untrue because we can prove that's untrue, whereas the first possible meaning would be true, but it's not really what she was saying and not really what anyone says. So, we have a lot of detailed conversations like that."

# Surfacing

*"What we try to do is develop tools that help surface claims they can then decide are worth checking…* It roughly groups by category, things like health, education, crime rates. If a story's picked up by several journalists it will put that on top of our list. The fact checker is going to look and say 'are there any major stories in the area I'm interested in I might have missed?'… *It takes few minutes to read, which is important…* In the last few months they have published two or three fact checks based on things found using that tool. "

# Selection

"In the morning fact checkers will choose what to check by discussion amongst themselves. It does vary a lot. Every claim is very different, but often will include things like working out exactly who said what, what exactly is the claim, what's the audience, what's the context? In more technical cases it may also have involved speaking to an expert. *A lot of these pandemic related stories, they have spoken to a doctor or medical researcher 'how do you know MRNA vaccines are actually safe?'... Get expert opinion, or explanation*. In a lot of cases it may involve reading published reports."

# Selection

"*Providing an accessible summary of the evidence around the claim, including sources and, experts, trusted opinions or further sources of information in such a way that fact checkers can very quickly decide this one is worth digging into a bit more*. They still do the work themselves, which still requires this kind of expertise, you know, research and storytelling. But if you can give them enough evidence at the outset so that something that may look unpromising in itself, with a bit of data around it, may think, actually, that is worth looking into a bit more. *That kind of tool, I think, is something that actually could have an impact*."

# Selection

"The other one I wrote this week is about Lorraine Kelly and her TV show, on which she and Doctor Hilary Jones the claim that 90% of people in hospital with COVID are unvaccinated. That's a good example of something I already knew wasn't true because I've written about this subject fairly recently. The real number is about 36%, so you could certainly say that number is substantially misleading. What they were basically saying is you should get vaccinated because it seriously reduces your chance to go into hospital and then they use this 90% numbers as substantiation for that. *So, their basic point is right, you should get vaccinated because it really does substantially reduce the chance of going to hospital. But the number they used to substantiate it was wrong*. So that would be an example of something that was important to check because it was you know, on national TV on a widely seen show. It's been made numerous times, but the potential for harm is probably lower. *The Lionel Shriver case might be the other way around, where it's probably not going to be seen by as many people as a column in the Spectator, but the potential for harm is much higher because she was carefully constructing an argument based on data where she explained to people that being vaccinated didn't make any difference on spread, and if people believe that and take it seriously it could actually, you know, make it. It's untrue and it could make a really harmful effect on the world*."

# Find & review evidence

"Sometimes, even if you know what sort of broad topic area, there are still so many sources of information on that topic to sift through. *Something that would simplify that process, or at least pull out a list of what you might want to look at would definitely make that process easier*. And similarly, the experts to talk to, here's some people you might want to consider, or some organisations, that kind of thing… *Primarily the first thing that we'll check for is have they been fact checked previously? Has this person shared misinformation? Is that something that we need to aware of? And what sort of expertise that they have in an area, that kind of thing.*"

# Find & review evidence

"You could get your algorithm to learn this has been used in articles 10 times in the last month as a credible source, that would be helpful because journalism's the polar opposite of academia. We do things at pace and hope for the best… Sometimes you're going for tried and trusted sources so, *anything that can give you a shortcut to doing would be hugely beneficial*."

# Find experts

*"We would always seek to speak to an expert…* you can get the data if you wanted to find out if somebody said this number of people have died from COVID in in the US. You could go to the Johns Hopkins stats or whatever. But if we were doing something that was in any way more detailed… *we would always look to speak to somebody who is an expert… 'we've spoken to X expert or Y expert and this is what they said'*."

# Find experts

*"but that notion that rather than having to sort of search yourself, you could have a kind of a triage would mean that we're not search searching blindly particularly when something comes up that we're not that knowledgeable about*… we're all researchers, we're all journalists by trade. We know how to do this sort of stuff, but particularly when something comes up that isn't necessarily your area of expertise. When the 5G stuff started… you'd go Googling, trying to work out who the national group is for 5G operators… *so you are spending a lot of time trying to find the right people*."

# Write up

"You could have somebody who's going to deny climate change, but do you give them equal airtime to somebody who is saying climate change is real? And probably if you give that sort of false equivalence, the public comes away thinking, nobody really knows, but actually the evidence is overwhelming. I think the challenge for a tool to do this kind of thing is not just find both sides of the story and present them. *But to show where the balance of evidence is pointing, which fact checkers are skilled at and it'll be interesting to see how you can automate that*."

# Write up

"A lot of the work is writing up what's been found. The writing of an article, which has to be very neutral, very clear, easy for anyone to read, and presenting all of the evidence in a kind of non-judgmental way so readers can then decide themselves whether not they believe this claim. *And that process of writing and editing and publishing is quite collaborative*. *I think three fact checkers will read a piece before it gets published to make sure that it is neutral, there's no claims there that should be verified more carefully*."

# Write up

"*The editing process, the reviewing stage, then it's very collaborative*. The reviewer virtually does the same amount of work as the writer because they're repeating the entire process of building up the evidence base… which can be a very, very complex, sometimes drawn out process… *Collaboration is really, really important. Mainly to check the writer has got it all right but also it allows more than one person to understand the subject*."

# Write up

"Our audience… don't want someone saying well this is true, take our word for it. They want us to do the research for them… then let make their own mind up. *Can you get a machine to find evidence that a fact checker can then filter and present it to the public?* That's a feasible goal. *But automation of whether something is true or false depends so much on real-world understanding. It's a very, very hard thing for any AI to do*."

# Automation

"There are certain types of claims which can be verified by machine… claims that are quantitative and which have a clear, a trustworthy ground source that we can verify against. If someone says that retail price inflation is down 5%, either that's right or wrong, it's the kind of claim which potentially can be checked by machine... It's challenging to figure out what's actually being said... *Even if it worked perfectly, it would only cover a tiny range of claims that are being made*."

# Automation

*"There is the risk that the algorithms are black box that nobody is going to really trust without a lot of experience. Explainability for fact checkers would tend to be extracts, quotes and snippets and links to other sources*. Primary sources. Things like links to the university homepage, saying yes, this person is employed by this university, here's their homepage is kind of evidence. If a person has been talked about by other journalists, links and quotes from their articles and they're quoted in the New York Times and in these 20 odd newspapers you know they're widely spoken about in the media. You want to show the evidence, but you don't want to reduce it down to a number or list, you want to say here is how they were quoted."

# Requirements…

- *Providing an accessible summary of evidence, including sources and, experts, trusted opinions or further sources of information in a way to quickly decide this one is worth digging into:*
  - *The first thing that we'll check for is have they been fact checked previously? Has this person shared misinformation? Is that something that we need to aware of? And what sort of expertise that they have in an area.*
  - *We would always look to speak to somebody who is an expert.*
  - *Sometimes you're going for tried and trusted sources so anything a shortcut to doing would be hugely beneficial.*
  - *A triage so that we're not search searching blindly particularly when something comes up that we're not that knowledgeable about.*
- *Show where the balance of evidence is pointing, which fact checkers are skilled at and it'll be interesting to see how you can automate that.*
- *I think three fact checkers will read a piece before it gets published to make sure that it is neutral, there's no claims there that should be verified more carefully:*
  - *Collaboration is really, really important. Mainly to check the writer has got it all right but also it allows more than one person to understand the subject*.
- *There is the risk that the algorithms are black box that nobody is going to really trust without a lot of experience.*

# Use case

1. Claim entered.
2. Matching documents returned & summarised:
   - Type, organisation, author, relevance score, stance.
3. Organisations & authors examined:
   - Evidence that organisation and/or author are credible?
     - Document type – scientific, news article, etc.
     - Have they written on this or similar topic before?
     - Has organisation and/or author been used before?
4. Summarise evidence:
   - Filter function highlights +ve & -ve stance.
   - Authors approached:
     - Most relevant sentences in document help decide questions.
5. Write fact check

# Summary

- Claims can be difficult to interpret.
- Prioritising what to check is important.
- The fact checking process:
  - Finding evidence
  - Finding sources
- Importance of (organisational) memory:
  - Has claim or a variant of it appeared before?
  - Has source been used before, if so, are they considered reliable?
- Collaboration is important for robustness.
- Practices vary between organisations.

# Next steps..?

- Ethnographic studies:
  - Dedicated to observing in detail everyday practices.
  - Attentive to how work actually ‘gets done’, cooperative activities.
- Social context and how this influences the work.
- How to embed new tools such that practices are enhanced.