The State of Automated Factchecking

How to make factchecking
dramatically more effective
with technology we have now.

Full Fact

With warmest thanks to

BYTEMARK who generously approached us to offer free hosting back in 2012 and who have supported us ever since, now by hosting our automated factchecking platform.

flax who have been generous in sharing their knowledge, skills, and expertise in open source search and media monitoring.

The organisers of the 2016 Tech & Check conference: the International Factchecking Network based at the Poynter Institute, and Duke University's Reporters' Lab.

Gabriel Pogrund and to Google, whose News Lab Fellowship at Full Fact supported his contribution to this paper.

The researchers and open source developers whose extraordinary work over decades has created these opportunities.

All our funders and supporters, listed at fullfact.org/funding, and the trustees and other volunteers of Full Fact.

You can follow our progress at fullfact.org/automated.

Mevan Babakar and Will Moy

Full Fact, August 2016

Contents

Summary

- We can scale up and speed up factchecking dramatically using technology that exists *now.*

- We are months—and relatively small amounts of money—away from putting practical automated tools in factcheckers' and journalists' hands. This is not the horizon of artificial intelligence; it is simply the application of existing technology to factchecking.

- Automated factchecking projects are taking place across the world, but they are fragmented. This means factcheckers and researchers are wasting time and money reinventing the wheel.

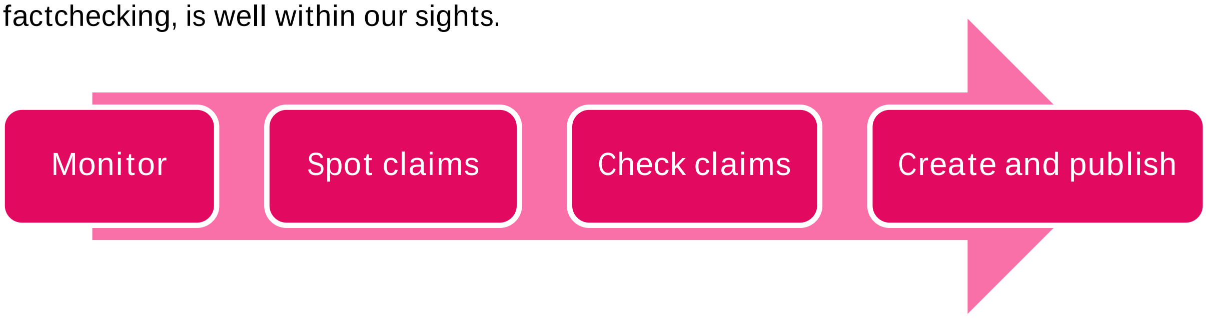
- We propose open standards. Automated factchecking will come to fruition in a more coherent and efficient way if key players think in terms of similar questions and design principles, learn from existing language processing tasks, and build shared infrastructure.

- International collaboration is vital so that the system works in several languages and countries.

- Research into machine learning must continue, but we can make serious progress harnessing other technologies in the meantime.

Terms of reference

Factchecking is the same four stage process whether it's done by humans or machines. Technology which integrates these four stages, achieving automated "end to end" factchecking, is well within our sights.



Much of today's political debate depends on repetition. The proliferation of media across many channels, less airtime, and smaller sound bites together demand that campaign managers doggedly stay on message above all else.

This repetition means automated factchecking can have real impact, but the proliferation of different channels is a challenge for factcheckers too: as campaigns get their messages out in ever more targeted ways, factcheckers will have to move quickly to adapt our monitoring and automated checking to keep up.

This report offers an overview of automated factchecking and the challenges which stand in the way of its development, as well as several immediate and easily achievable solutions. It comes in two parts.

Part One: A roadmap for automated factchecking

In this section, we argue for shared international standards and systems for building automated factchecking tools, and we set out Full Fact's roadmap for getting fully automated factchecking running as soon as possible.

If we build the tools right, and create the right conditions, we will make it as simple as possible to track more sources, check more claims, and build different products with the results.

Part Two: What we can do now, and what remains to be done

Here, we define the challenges for factcheckers, computer scientists, and others in developing or extending practical factchecking tools. We examine existing efforts at automated factchecking and identify interesting and potentially fruitful areas for further research.

At the core is the distinction between two approaches. Some tools automate existing human factchecking processes, which are often repetitive and ripe for automation. Some tools use machine learning and artificial intelligence techniques.

We cannot suddenly achieve the artificially intelligent holy grail of a machine that can replace everything human factcheckers do. That is a very long way off. But there is still a lot we can do.

The state of automated factchecking right now

The table below summarises the state of automated factchecking at the time of publication.

It includes the challenges which researchers have already solved and for which working technologies exist (at the top) and goes down to those which researchers are still working on (at the bottom).

	1	2	3	4
	Monitor	Spot claims	Check claims	Create and publish
READY NOW “Off the shelf” tools	Monitoring Twitter, emails, and websites Monitoring TV subtitles Monitoring online adverts and Facebook	Claim recognition for known claims Claim ID: spotting known forms of claims	Check subject specific claims, depending on precise wording and structured data.	Create simple human-readable content from templates Real time tools.
		Claim ID: recognising new claims generally	Check wider set of claims without specialist methods but with structured data	Create persuasive human-friendly content and tools
Ongoing research problems: NOT YET POSSIBLE	Monitoring radio or TV via speech recognition	Handling paraphrases	Check any claim better than a human with the internet	

This table shows we are capable of creating an “end to end” factchecking system using technology already available to us.

Of course, taking it further will require new tools, and breakthroughs in computer science and natural language processing. Nevertheless we have the right tools to make a strong start.

With well-targeted investment, practical automated factchecking is months, not years, away.

PART ONE

A roadmap for automated factchecking

Five principles for international collaboration

Automated factchecking research projects are taking place at universities across the world. Full Fact is among the handful of factchecking organisations exploring how such research can be integrated into their operations. The others currently include Chequeado (Argentina), Les Decodeurs at Le Monde (France), and Politifact (US).

Improved collaboration between key players will help proper automated factchecking further and faster along the road to everyday usage. We propose five principles:

1. Standard data formats, so that any new automated factchecking tool can work with any known source.
2. Shared monitoring systems, so we do not duplicate work unnecessarily.
3. Open and shared evaluation, so we know what works and what it works for.
4. Published roadmaps, to attract volunteers, researchers, partners and funders to work with us.
5. Think global, so that where possible new automated factchecking tools are designed with the aim of being able to work for many languages and countries.

1. Standard data formats

We propose an open standard format for recording content, so that any new automated factchecking tool works with any known source.

This would see monitoring tools providing output and factchecking tools receiving input all using the same format.

The format would also store contextual information such as the date, location, speaker, and so on, as open data. This is discussed further in Part Two of the report.

There is also work going on with schema.org to provide a standard format for publishing the results of factchecks.

2. Shared monitoring systems

Factcheckers across the world all want to monitor a wide range of sources 24/7, including the media, social media, the internet, government sources, and advertising.

It takes a lot of effort to create a monitoring tool that operates on any or all of these sources—and so it makes sense to avoid duplication of that effort where possible.

We therefore propose that factcheckers and researchers coordinate and where possible deploy the same or compatible monitoring tools.

We have identified a series of 14 sources that factcheckers in different countries monitor, and in Part Two we give examples of existing tools that could be used to provide automated monitoring of each source where this is possible.

3. Open and shared evaluation

We propose that factcheckers together set out shared success criteria for automated factchecking tools. This would give researchers clarity as to our overall needs and also allow them to compare the efficacy of different approaches.

A recent paper commented: "there is no data for this task to evaluate, thus making development difficult".[1] This is a shared problem that is best solved together.

Any criteria should prioritise functionality over perfection. In practice, perfect accuracy is not achievable with existing technology. Most automated factchecking tools depend on splitting text into individual sentences, a task which can only be done with 90–95% accuracy.

In terms of developing specific criteria, Full Fact's position is that a successful automated factchecking system is one that saves factcheckers and journalists time, and makes factchecking more effective at limiting the spread of unsubstantiated claims.

We therefore prefer accurate tools to comprehensive ones. If an automated tool can monitor and check 10% of claims accurately, that is 10% more claims than we can handle at the moment without automated factchecking tools. It is all a benefit. On the other hand, if the tool checks 20% of claims but only half of results are accurate, then it will cause us work instead of saving us work.

4. Published roadmaps

We invite other research, factchecking, and media organisations to publish their roadmaps for automated factchecking, as we have done below.

We hope this will inspire funders to recognise the importance of this moment in this area, promote collaboration among us all, and attract volunteers, researchers, and partners to this work. Our roadmap will continue to be updated at https://fullfact.org/automated.

5. Think global

We should aim to design tools to be adaptable to different languages and countries, just as independent factcheckers around the world aim to collaborate and support each other's work in other areas.

This needs to be planned in from the start, and it requires extra effort and expense, but it would be less effort than different organisations in different countries reinventing the same thing from scratch.

It will not always be possible to make tools that work globally. Some approaches are closely tied to particular languages, or toolsets that are only available for particular languages. It is still useful to consider, and thinking about it early will help us make tools that are as robust as possible.

[1] Vlachos and Riedel. Identification and Verification of Simple Claims about Statistical Properties.

Longer term infrastructure

As a technical note, we expect that in the longer run mature automated factchecking tools will adopt standards and tool chains that have been developed for other language processing applications such as UIMA (Unstructured Information Management Architecture) and GATE (General Architecture for Text Engineering, used by Pheme [see part 2, stage 3]).

We nevertheless believe taking such an approach now would be premature. This kind of infrastructure is excessive for present needs and, unless dramatically increased levels of funding and staff becoming available, it will slow down the adoption of practical automated factchecking. The five principles proposed move us towards such frameworks, and the standardisation and componentisation will make it easier to adopt frameworks in the future.

Full Fact's roadmap

Overview

We aim to put the first fully automated factchecking system into use within the next year. Our current system is made up of:

Hawk, which is our monitoring and claim recognition engine. It monitors public debate and spots claims we have previously factchecked. It is in internal beta.

Stats, which is our tool for automatically checking statistical claims. It helps us factcheck more quickly and efficiently. It is a proof of concept, moving into prototype stage.

Trends, which is our monitoring product. It shows how common a claim is, where it is being made, and who is making it, using the results from Hawk. It helps us scale, target, and evaluate our work. It is in internal beta.

Robocheck, which is our real time product. It will provide subtitles of live TV, and add verdicts to claims in real time using the results from Stats and Hawk. It will help journalists and others hold public figures to account. It is in design phase.

The four components together will monitor, spot statistical claims, check them, prepare responses, and publish without human intervention, and also spot and respond to all kinds of claims that Full Fact has previously factchecked.

	1	2	3	4
	Monitor	Spot claims	Check claims	Create and publish
Hawk	Yes	Yes		
Stats		Yes	Yes	
Trends				Yes
Robocheck				Yes

Underneath all four components is a real time infrastructure. This includes server setup, documentation, logging and monitoring, and everything else that goes into making a 24/7 production system.

Each component is designed to be able to work quickly, and become much more capable over time.

In accordance with the design principles set out below, we want to make simple and scalable tools which will be useful across the world straightaway. We will work with our Argentinian friends and colleagues at Chequeado to make sure we achieve this.

Where are we now?

Full Fact's automated factchecking system is in internal beta and could soon be ready for external use. We track what is said in Parliament and in major media outlets in the UK. Our system can identifies we have previously factchecked. It can also identify and automatically check the accuracy of some statistical claims. The results are provided to our factcheckers on a password-protected section of our website.

We have two main objectives for our own development now:

1. To upgrade our existing monitoring system to provide real time results, including covering major TV news in real time.
2. To develop the first ever end-to-end fully automated factchecking system, from real time monitoring through to putting factchecks into the hands of journalists in major newsrooms.

With appropriate funding, we can deliver this within 12 months.

We are building these tools on top of a robust, scalable, production-ready infrastructure. They are intended to run 24/7 and each component is intended to make the others stronger. Having this infrastructure makes the development of new tools far easier.

Each component can be made to work quickly, and then be extended significantly. New sources—from TV subtitles to Twitter—can be added as modular components. New checking modules can be added as well to check different kinds of claims. And new products can be built with these tools depending on user needs.

Design principles

Our goal is to build tools that do something useful right now, and that can grow to cover the great volume of claims and sources that we or factchecking organisations around the world want to check. This means building a system that is flexible, modular, extensible and robust. We will –

1. Make simple tools that work and make factchecking more effective, then make them work better

2. Make each simple tool do one thing well, and work well with others

3. Make tools that will scale

4. Design tools to be adaptable to languages other than English

5. Focus on taking existing technology and applying it to making factchecking more effective, not on pushing the boundaries of computer science research

PART TWO

What we can do now and what remains to be done

Stage 1: Monitor

Shared tools for shared needs

In order to take material out of the newspaper, web, radio or TV and put it into a factchecking tool, you need to read, listen to or watch the content in question.

This major task is one for computers. A computer can 'read' the entire content of a newspaper or transcript of a debate in less time than it takes a person to read this sentence. If you want to know every time somebody has claimed that 'crime is rising' in the past year, again, it's a job for a computer. Computers do fall down when it comes to pictures, speech, and video, but they are getting stronger in these areas.

Most of the things factcheckers might want to monitor can be monitored, largely with existing tools. But turning these kinds of tools into a system that can monitor all these sources consistently and stably for a year is still a difficult engineering challenge. Scaling and fault tolerance take far longer than initial proofs of concept.

Shared tools for the 14 sources we need to track

At the moment the creators of automated factchecking tools have the added burden of extracting text information from the original source, such as news articles or TV. When different projects duplicate the same work on this, it is wasteful.

For some sources there are mature toolsets for this kind of work that it would be natural to standardise on. In other areas, we still urge that we will benefit from trying to use shared tools where possible.

> Question for factcheckers: which are the most important sources to you?
>
> Question for factcheckers: what are the other sources you might want to track? E.g. Periscope, Snapchat, how important are these?

> Research problem: More reliable speech recognition for TV, radio, and online sources

Here is a survey of some of what's available –

	Source	Status	International?	Tools available for extracting text information
Media	TV	●	Yes	Vlc, speech recognition APIs such as MAVIS
	TV subtitles	●	Yes	CCExtractor (unstable), TextGrabber (US only?)
	Radio	●	Yes	Speech recognition APIs such as MAVIS
	Newspapers	●	No	Custom feeds (usually costly) or custom scrapers
Government	Legislative debates	●	Yes, via SayIt	SayIt, a Poplus component

Online	Press releases	●	Yes	Will be covered by websites, emails, or RSS
	Websites	●	Yes	Heretrix (archive.org), Nutch
	Blogs (RSS)	●	Yes	API for most languages; tools such as LogStash
	Emails	●	Yes	IMAP APIs for most languages
Advertising	TV adverts	●	No	In the US, Political TV Ad Archive
	Online adverts	●	Yes	Ad verification services such as CompetiTrack
Social media	Twitter	●	Yes	API for most languages; tools such as LogStash
	Facebook	●	Yes	None. Only Facebook can access users' posts.
	YouTube	●	Yes	None? YouTube captions are often unreliable.

Key: The status indicates whether Full Fact is monitoring a source.

- ● *Green for active*
- ● *Orange for development*
- ● *Red for technical barriers to monitoring.*

Public debate as open data

We want to ensure that people working on different automated factchecking projects don't have to duplicate the effort of getting information out of raw sources. Any new automated factchecking tool should be able to work with any known source.

An open standard will make it possible for different automated factchecking tools to benefit from one monitoring tool, and for up to 12 of our 14 major sources it will mean that monitoring tools can be built that would be useful in many countries.

It will make it quicker, and easier and less expensive to get working automated factchecking in many places.

Who, where, and when: understanding what claims are talking about

Who makes a claim, when they say it, where they say it, and who they say it to, can all affect the conclusion a factcheck could reach. Whether it's true to say "unemployment is rising" depends on what country or which part of a country a speaker is referring to, and when the speaker makes the claim.

An open format for recording public debate should support metadata, including at least the time, the place, the venue or publication, and the speaker.

Actually obtaining that metadata is another challenge. It's easy for some kinds of metadata and some sources. Newspaper articles usually have dates and bylines, for example. But it's hard to know what place is being talked about in many cases. That may have to be assumed or inferred with less than perfect accuracy.

Similarly, it can be hard to identify the speaker in some contexts, including on television. To solve that problem, the BBC are experimenting with using voice “fingerprinting” to identify the same speaker in different recordings.

We need a data format that can cope with uncertain metadata as well.

> Research problem: How to infer the place being talked about in various texts? For example, a good project might be to improve the accuracy of these inferences for parliamentary debates using the titles of debates, the constituencies of members, key phrases like ‘my constituency’, and other clues.

Stage 2: Spot claims

The open source software that can track a million claims in real time

There are at least four different technical tasks in this stage of factchecking:

1. Monitoring claims that have been factchecked before in new text
2. Identifying new factual claims that have not been factchecked before in new text
3. Making editorial judgements about the priority of different claims
4. Dealing with different phrasing for the same or similar claims

Monitoring claims that have been factchecked before

Media monitoring companies use technology to work at very large scale, helping brands keep track of their press, for example. Their task and the task of monitoring for previously factchecked claims are very similar, and the same open source tools can be used for both.

Two leading open source search engines are Apache Solr and Elastic Search. Both are based on the same underlying system, called Apache Lucene. They provide the ability to search millions of documents in milliseconds.

Checking many new statements in real time needs a slightly different tool. Fortunately these search engines also have Structured Query Engines which are geared towards providing real time results for many searches across an incoming stream of text. They are: Percolator, which runs off Elastic Search, and Luwak which is built directly on Lucene.

Full Fact has been actively prototyping these kinds of tools since late 2014, and they now underpin our alpha-stage internal monitoring system, which we have designed to help us scale, target, and evaluate our work. We are pleased with the useful results we are able to get but we have learned a number of lessons in the process.

First, making these standard tools perform well for this task is highly specialist. Search engines are designed to return the ten most relevant results for some search, not to return all documents that match a very tightly-defined claim. Search engines are also highly configurable. That tuning matters a great deal to the accuracy and usefulness of their results, and it requires specialist skills.

Secondly, writing effective search terms requires both skill and some understanding of the monitoring system. Not all claims can be tracked in this way. It depends on constructing search terms that catch the right sentences but don't accidentally catch too many others.

Finally, sometimes it is better simply to choose to search less material than to use more sophisticated search techniques. Although we can search all the sports news, tackling the extra complexity of distinguishing deficits in football scores from deficits in the national budget is not currently the right focus for us.

> Research problem: Is it possible to suggest high quality search terms given instances of claims? What about the next step, with machine-derived search terms?

Identifying factual **claims we don't yet know about**

Finding claims that have previously been factchecked is already practical. The next step is to get computers to identify claims we have not yet checked.

There are two underlying approaches to this: the simpler one is recognising predictable forms of claims such as 'X is rising', while the more complex one is using machine learning to teach a computer what claims look like.

Either approach has to distinguish factual claims from rhetoric, which is not always easy. Sweeping statements like "this government has been an economic disaster" are part of political speech, and they are not always intended to be understood as factual claims. It is possible but not easy to get a high level of agreement among human factcheckers on this boundary. You have to choose where to draw the line, not simply learn to spot it. This means getting consistent results from computers will be difficult.

The simplest approach to identifying claims is to look for forms of claims. For example, 'so-and-so voted for such-and-such' is a predictable form of claim that many factcheckers see. It is fairly easy for a computer to spot every sentence in that form, and so to find every instance of a certain way of making a certain kind of claim.

If you want to identify every claim in a text and ultimately reduce the burden on humans for monitoring, you need something more. This is a problem researchers are tackling with machine learning, training computers to spot patterns in what counts as a claim from human examples.

The most advanced generalised automatic claim spotting is by a tool called ClaimBuster developed by Chengkai Li at the University of Texas at Arlington and others, which uses machine learning to determine the probability that a sentence contains a 'check-worthy claim' based on manually coded examples from past US presidential debates.[2] In other words, the software looks at known check-worthy sentences, identifies features they have in common, and looks for these features in new sentences. Although it is focused on debates at the moment, this approach seems easiest to generalise out to media other than presidential debates.

Outside the realm of factchecking there have been efforts to classify sentences into several different types, which can include identifying factual claims.

For example, for scientific papers, Simone Teufel, then at the University of Edinburgh, developed an approach which she called "argumentative zoning".[3] Scientific papers tend to have predictable components such as stating an aim, giving background, and comparing the paper with other research. Her work aimed to assign every sentence in a paper to one of these categories, again with machine learning. This kind of classification is

[2] Naeemul Hassan, Bill Adair, James T. Hamilton, Chengkai Li, Mark Tremayne, Jun Yang, and Cong Yu. The quest to automate fact-checking. http://cj2015.brown.columbia.edu/papers/automate-fact-checking.pdf

[3] http://www.cl.cam.ac.uk/~sht25/az.html

useful beyond factchecking – such as in automatic summarisation and in improving search.

Speeches in parliamentary debates often have similarly predictable components. Full Fact is involved in a project of the UK Parliamentary Office for Science and Technology to see how this approach might be applied to the records of House of Commons debates.[4] This may have benefits for factchecking in due course.

Machine learning is usually dependent on the quality and volume of training data available, so getting these approaches to work at their best is likely to involve investing human time in providing example data.

Research on making editorial judgements to prioritise claims

When human factcheckers choose what claims to check, we look at two things.

First, we look at the content of the sentence: is it a claim at all? Secondly, we at the context of the claim: who is saying it, who is repeating it, and so on. The two put together tell us something about how interesting, important, and influential the claim is.

The content approach is harder for computers. ClaimBuster, mentioned above, seeks not only to spot claims but also to suggest an order of priority for tackling them. It does this through machine learning about the content of claims. Sentences which share most features with sentences previously marked as check-worthy get higher probabilities of being check-worthy.

The contextual approach is routine for computers. Identifying important content based on reach and engagement is a central concern in much online marketing, for example, so tracing claims and judging the social influence of those making them is something computers can do well.

Combining the content and context approaches—just as human factcheckers do—is the most helpful approach for automated factchecking.

It is essential that researchers and factcheckers collaborate on this particular kind of work, as with Politifact's and factchecker-turned-Professor Bill Adair's contributions to the development of ClaimBuster. Many of the tools factcheckers need depend on new research, and that research can only succeed if factcheckers take the time to engage with researchers. Equally, many of the machine learning approaches need quality training data that factcheckers are well-placed to provide—just as Politifact did for ClaimBuster.

Dealing with paraphrasing

The same claim can be made in many different ways and that causes problems for spotting claims. A fully automated factchecking system would be able to identify different phrases which make the same claim, and distinguish very similar phrases that make different claims, as humans can.

[4] https://fullfact.org/blog/2015/jul/automatically-spotting-interesting-sentences-parliamentary-debates/

An associated problem is claims made over more than one sentence. Most automated factchecking tools start by breaking texts up into individual sentences, and simple sentences are by and large the building blocks of political debate. But that move can sacrifice a whole set of interrelated claims. Finding ways of tackling claims over more than one sentence will become more important when the basic underlying technology improves.

Useful automated factchecking tools can be and have been built without solving this problem, not least because public debate depends on repetition so much anyway. But this problem is firmly on the roadmap for fully automated factchecking, and it is one of the problems that still need to be solved by researchers, not just by the application of existing tools.[5]

Dealing with paraphrasing in factchecking is likely to be harder than in most natural processing applications because precise wording can matter so much to a factcheck's conclusion. Recognising that claims are very similar will not be enough for some factchecking applications.

[5] Dan Schultz's thesis *Truth Goggles: Automatic Incorporation of Context and Primary Source for a Critical Media Experience* contains a useful discussion of handling paraphrasing at pp74-80

Stage 3: Check claims

What factcheckers are learning from driverless cars

In 2004, the US Defense Advanced Research Projects Agency ran a challenge for driverless cars. It did not go well. Not a single car finished its route and no winner was declared. Most cars failed to complete the route the following year and those who did were very slow: the winner's average speed was 19 miles per hour. Yet just a few years later, Google had driverless cars running hundreds of thousands of miles on public roads.

What changed?

Peter Norvig, Google's Chief Scientist, explained that: "We didn't have better algorithms, we just had better data." If you can change the problem from identifying a traffic light in a picture, to knowing that the traffic light is there and simply needing to find out whether it is red or green, you can make the job much easier for a computer.

Better data is also making automated factchecking possible. Algorithms are still nowhere near capable of doing what human factcheckers consider basic, but they can crunch data a million times faster than us.

So, although factcheckers around the world are still struggling to get reliable data in the first place, one of the most important priorities for automated factchecking is to make sure the kinds of sources factcheckers can rely on are available as structured data that computers can use.

Three working approaches to automated checking

Take a very simple claim, such as "Victoria and David Beckham are married". Three different approaches to checking it might look something like this:

1. Reference approaches: Look up their names in the register of marriages, with a tool that knows about these. This is what you do if you want to be absolutely sure and there is only one way to do it. ("The official record says they are married.")
2. Machine learning approaches: Make a mathematical model of things we know, or of the nature of known things. Test whether the claim, or the contradiction of the claim, is more probable. This puts all the weight on how you define knowledge and how you test it. ("The computer says they are probably married, so they probably are.")
3. Contextual approaches: Look at how claims that they are married spread. This does not directly tell you whether the claim is true, but is quite a practical way of finding out whether it's likely. It's the ideal of the marketplace of ideas. ("Claims that they're married survive longer in open discussion than contradictory claims, so they probably are.").

We are glossing over common challenges, such as knowing that married is equivalent in this case to husband and wife, or knowing that Victoria and David Beckham are the same people as Posh and Becks.

Clearly, all three approaches could be used together, but they are quite different in underlying technology. It is too early to make strong predictions, but fully automated factchecking when achieved seems likely to use a combination of these approaches. In the meantime, reference approaches and, in a slightly different domain, contextual approaches seem closest to delivering real products now, while the forefront of research seems to be more in machine learning and statistical approaches.

> Research problem: There is an urgent need for a thorough literature review of work on automated checking, including work outside academia.

Reference approaches: Reasoning with known sources and techniques

These approaches are closest to trying to automate the same processes that human factcheckers go through. They therefore look at whether claims correspond to reference data or a reference model of how they should be tested. They have built-in knowledge of how to check particular types of claims.

They may be relatively narrow in their scope, but they are relatively simple to get to an effective stage because they put more emphasis on understanding the structure of the claims or data than on ground-breaking algorithms.

For example, 'Statcheck', created by two Dutch researchers, automatically tests whether the findings in psychology papers are statistically significant.[6]

Another example is a commercial product called jEugene Cross-References[7], which claims to automatically verify the existence of sections cited (e.g., "Section 1.01") in legal documents.

Instant answers from search engines do a similar job, albeit on more than just factchecking.

Machine learning approaches: Reasoning with probability and mathematical models

These approaches try to get the same quality of results as human factcheckers but not the same way as human factcheckers get them. This approach is closer to the frontiers of research and puts more emphasis on sophisticated algorithms than topic knowledge.

When computers translate text, they don't look things up in bilingual dictionaries, they use masses of statistical information and inference instead. This isn't understanding of the topic in the way we would normally think about that in people or reference books, but that isn't always necessary to get accurate results, whether you are translating or factchecking.

For example, Giovanni Ciampaglia and his collaborators took a database of facts from Wikipedia, restructured it into a mathematical representation called a knowledge graph, which is a network of points and joins where each point is something (like one of the

[6] Nuijten, M.B., Hartgerink, C.H.J., Assen, M.A.L.M. van, Epskamp, S., & Wicherts, J.M. (2015). The prevalence of statistical reporting errors in psychology (1985-2013). Behavior Research Methods at https://mbnuijten.files.wordpress.com/2013/01/nuijteneta_2015_reportingerrorspsychology1.pdf
[7] https://jeugene.com/

Beckhams) and each join is a relationship (like marriage). The factchecking works on the theory that true claims are likely to be closer together in that structure than false ones and returns a truth score, a numerical probability that the claim is accurate.[8]

A challenge here is that the algorithms by nature do not always provide transparent reasoning that readers can judge them for themselves.

Contextual approaches: Reasoning about the social and other context of claims

These tools look at what reaction claims get when they are made. Do they spread? Do people contradict them? Do the opposing claims spread too?

This is the sideways look at the problem. Useful factchecking can be done without understanding anything about the claim itself. Computers can decide who to trust in the same way most of us do most of the time—not by doing all the research we could, but by placing trust in people. Or they can use another shortcut of seeing how contested a claim is. Using extrinsic evidence to infer accuracy cannot be certain, but we all rely on it when we choose who to trust rather than looking things up for ourselves.

For example, Twitter Trails assesses the accuracy of stories based on expressions of scepticism in tweets and patterns of dissemination. It relies on the 'wisdom of the crowd' assumption that, on one hand, users will react with scepticism towards the most dubious claims, and that, on the other, users won't re-tweet those claims which they believe to be inaccurate.[9]

Another example is part of the Pheme project, which uses epidemiological models seeded with information from human factchecking to understand the spread of rumours online and try to identify emerging rumours.[10]

A new tool for checking statistical claims

Statistical knowledge is largely out there on the internet somewhere. Nonetheless, making a tool capable of checking all statistical claims in any form would be very hard. Even searching for simple statistics on the internet can be frustrating. But providing authoritative data sources can help us make a tool that can check many important statistical claims automatically.

Take claims like 'X is rising' or 'Y has doubled in the last ten years'. Human factcheckers look at what the sentence is saying has changed (employment, crime, etc.), get the data on that topic, and test whether the claim is true using subject expertise about the data.

Computers can do all that, but they need a lot of help from structured data to do it. Like human factcheckers, they need to know first where to find the appropriate statistics, then they need the statistics themselves, and finally they need caveats and context to

[8] Ciampaglia, Giovanni Luca, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. "Computational Fact Checking from Knowledge Networks". In: PLoS ONE 10.6 (June 2015), e0128193. http://dx.doi.org/10.1371/journal.pone.0128193.
[9] http://twittertrails.wellesley.edu/~trails/
[10] http://www.derczynski.com/sheffield/papers/synergy-workshop.pdf

understand those statistics. As all human factcheckers know, changes in methodology and external events can completely change the meaning of statistics, so just having the numbers isn't enough.

Since we first demoed a basic automated statistical factchecking proof of concept in 2013, Full Fact has been working on securing the three sources of structured data needed to power a practical product, and making sure they have APIs.

1. A database of sources of statistics with an API for finding the right source for the right topic.
2. Getting important statistics into standard open formats and APIs (such as the new API-driven Office for National Statistics website in the UK).
3. A standard format for statistical caveats and reuse information, which we call Open Statistics.

Now we are ready to integrate these into a working product that will interpret statistical claims and use this data to reach conclusions about them. This is complex, but not on the same scale as building a tool that can reason generally about unstructured statistical claims.

The development work here is complex, but it is important to highlight that at least as much work has gone into advocacy towards and working with information publishers, particularly the UK's Office for National Statistics.[11] By adding value to the data, we are helping to create assets which are useful far beyond automated factchecking applications, and that helps to get traction.

It is plausible similar approaches will work for other topics. Useful results in automated checking may well come at first topic by topic, and may well owe more to specialist knowledge and structured data than to artificial intelligence.

The automated checking landscape

Automated checking projects vary in what kinds of sources they deal with, what kinds of claims they deal with, and what topics they deal with.

It general terms, the narrower the scope, the more likely the project is to provide practical tools for factcheckers. The more ambitious the scope, the closer it is likely to be to pure research.

So, on the one extreme there is a practical tool for cross-checking references in legal documents, and on the other there are academic projects researching how to check everything on social media.

[11] Some more detail can be found at https://fullfact.org/news/make-official-figures-more-useful-and-accessible-say-mps/ and https://fullfact.org/blog/2015/aug/typology-caveats/

Here are the automated checking projects we've spotted. There are further details of these in Appendix 1.

Name	Sources	Claim types	Topics
jEugene	Legal documents	Legal cross-references	Legal citations
Pheme	User generated content on social media	Rumours on social media and the wider web	Any
Statcheck	Academic papers	p-value calculations	Psychology
Trooclick	Online news articles	Reported speech	News
Truth Googles	Any webpage	Misinformation across the web	Any
Truth Teller	US political speeches	Misinformation	Politics and government
Twitter Trail	User generated content on Twitter	Manually identified rumours	Any

Stage 4: Create and publish

Computers can do facts but can they do flair?

Existing automated factchecking tools mostly rely on content prepared by humans. Alternatively, they tend to provide something that looks like the output of a computer programme, such as a number giving the probability of a claim being true.

There is a great deal of work being done on creating editorial content automatically. The Tow Centre for Digital Journalism's Guide to Automated Journalism[12] is a helpful overview.

Automated journalism is in a sense the reverse of the process of automated factchecking. Both tend to rely on high-quality structured data to be useful, and both tend to be most useful when they are working within a well-defined topic area. Automated journalism takes data and tries to make stories, while automated factchecking takes stories and tries to reverse-engineer the data or sources.

People working on automated journalism have tended to focus on things that are easy to communicate: "routine news stories for repetitive topics", as the Tow Centre Guide puts it.

The challenge for factcheckers is that communicating our conclusions is crucial to our work: what good does it do knowing a claim is wrong if you and your editor are the only people who read the article? One good graphic can be worth a thousand words. Eight words can be worth more than a thousand. A lot of skill goes into the research, but perhaps even more goes into simplifying it. That simplification has to be compelling, it has to be authoritative, and it has to be recognisably fair for people on different sides. So this stage of factchecking is a job for a human being.

There may be potential for developing a library of compelling but standard ways of expressing certain kinds of conclusions of factchecks that we see regularly. For example, "he never said that" or "alternative data gives a different picture". If so, the first step will be developing a pathology of factchecking results that categorises the kinds of inaccurate or unsubstantiated claims we see. This will make it easier to develop standard responses and to automate their production. It is also useful for many other purposes, so Full Fact has been working on this anyway.

> Question for factcheckers: what kinds of conclusions do you see regularly from your factchecks? Do you have a 'pathology' of regular problems? Do you have standard responses (phrasing, graphics, examples...) to certain kinds of errors?

Common standards are already emerging for presenting factchecks through work done with schema.org.[13] This is about encoding factchecks as structured data so they can be presented in different places, from shareable widgets to search results. Combined with effective real time monitoring for previously-checked claims, it has the potential to get factchecking both well presented and presented on the right channels.

[12] http://towcenter.org/research/guide-to-automated-journalism/
[13] See https://github.com/schemaorg/schemaorg/issues/1061 for details and links to examples

Getting the right information to the right people at the right time

There are many ideas for how factchecking and automated factchecking can be presented.

Much of this relates to the psychology of the readership. Dan Schultz, creator of Truth Goggles, wrote the most charming and insightful line on automated factchecking so far: “I learned about the millions and millions of reasons why my idea could never work.” He said this because many people in fact end up believing something *more* strongly when you tell them it’s wrong. How you communicate factchecks with people matters, then, for human factcheckers and for machines. There is plenty of work to do to make sure that the way we publish has the right effect.

> Question for factcheckers: what kind of tools work, and what do users really want?

Two different kinds of approaches to presenting the results give a sense of just how versatile and useful automated factchecking could be.

Real time pop ups for audiences

Most factcheckers will have been asked why we can’t have factchecks appearing in real time on TV. Some of us occasionally get the opportunity to provide that service. The Washington Post’s Truth Teller demo went a step further, providing automated factchecking annotations for video clips based on previous human factchecks.[14]

For a web based example, Truth Goggles[15] was a tool that augmented web pages with factchecks in three ways. Highlight mode simply highlighted factchecked claims and invited the user to see factchecks. Safe mode covered the claim, requiring the user to click to read the factcheck before seeing the claim. Goggles mode obscured text after a claim, requiring the user to read the factcheck to continue through the page.

Claim tracker for factcheckers and journalists

At the other end of the scale is building tools to help make factcheckers and others more effective in their accountability work.

Full Fact Trends, which is in internal beta, is intended to help us target, scale, and evaluate our work. It provides a graph of how frequently claims have appeared over time, the details of where the claim has appeared, both of which move us towards rigorous evaluation of the before and after effects of our interventions. The list of claims is also useful for targeting our interventions, such as correction requests, by showing where they are most needed. Finally, an indicator of how common the claim is relative to others helps us to prioritise.

---

[14] https://www.washingtonpost.com/news/ask-the-post/wp/2013/09/25/announcing-truth-teller-beta-a-better-way-to-watch-political-speech/
[15] https://slifty.com/2012/05/achievement-unlocked-thesis/

Next steps

A global discussion is beginning around automated factchecking, thanks to the Tech & Check conference recently organised by the International Factchecking Network based at the Poynter Institute and Duke University's Reporters' Lab.

People are attacking the problem from two sides: factcheckers trying to get practical tools into actual use, and computer scientists trying to solve deep underlying research problems.

We hope that this paper has contributed to that conversation by giving a sense of an overview of automated factchecking projects, showing how research and practical projects can support one another, and sharing more about where Full Fact's plans fit in.

Please get in touch, and especially please let us know of any omissions or additions. Mevan Babakar can be contacted on mevan@fullfact.org. We will update this paper periodically, and information will be at https://fullfact.org/automated.

We would like to participate in open collaborative standards-based development of automated factchecking tools that are useful for factcheckers around the world.

We would like to thank the generations of researchers whose work has made possible the tools we are now putting to use, and the researchers who continue to push the frontier forward.

But most of all we are convinced that automated factchecking tools can and should be making a difference right now and we hope others share our urgency.

APPENDIX

Known automated factchecking projects

Known automated factchecking projects

This survey is probably not yet comprehensive. We hope it is the first step towards a much better map of activity in this area, and therefore towards better collaboration. We would also gratefully accept any correction or clarification of the descriptions we have provided. We apologise for any errors. Please get in touch on mevan@fullfact.org.

No factchecking products have so far been widely used by factcheckers or journalists.

Most projects specialise in one or two components of the factchecking process, such as detecting check-worthy claims and cross-referencing claims against structured databases of previous checks. Many are restricted to certain subject domains.

To keep the task manageable, we have not included projects which seek to make crowdsourcing of verification more efficient, and focused on tools that verify claims in text and speech, rather than checking the veracity of images, for example. Each project comes with a graphic showing which of the four stages of factchecking it engages in.

This list will be kept up to date at fullfact.org/automated.

Stage 1: Monitor

Trooclick (2012–present) http://trooclick.com/about

1

Uses natural language processing to extract reported speech in online news and convert it into data. Speech on a given news item is gathered from across the web and encoded with associated metadata, such as the name of the person who behind it and the organisation they work for. Each statement is catalogued into a structured database. The database forms an API allowing users to analyse the news alongside individuals and organisations.

Stage 2: Claim recognition and claim ID

Claimbuster (November 2015–present)

1 2

http://cj2015.brown.columbia.edu/papers/automate-fact-checking.pdf

Designed to help journalists factcheck this year's US presidential debates. It scans content in live streams, websites and social media, finding what its developers call the most 'check-worthy' sentences. This is based on natural language processing and machine learning software, which is fed thousands of manually annotated sentences from past presidential debates. Every sentence is ranked 0 to 1, so factcheckers are notified of the most check-worthy claims.

Simple Numerical Fact Checker (2014–present) 2

http://andreasvlachos.github.io/activities/

A machine learning approach to statistical factchecking, starting with claim identification. Their first paper discussed the main challenges, namely the open domain nature of the task and the importance of context: temporal, geographical, conversational. They followed this up with a distantly supervised machine learning approach for factchecking simple claims about statistical properties, i.e. what you might call 'is claims' such as "Lesotho has a population of nearly 2 million", based on looking for patterns of entity (Lesotho), property (population), and value (2 million). They achieved 60% accuracy overall but wide variation between patterns.

Emergent.info (2014– 2015) http://www.emergent.info/ 1 2

Emergent.info was a real-time rumour tracker specialising in unverified information and the spread of rumours in the media. It found unconfirmed reports early on in their life cycle, gave them a 'true', 'false' or 'unverified' rating, and tracked their spread online. The website would include information on the source of the rumour, the amount it had been shared socially, and which media organisations had reported on it. The Tow Center for Digital Journalism project required a lot of manual input; the team members found rumours using RSS feeds, Twitter, Google Alerts and user tip offs, and manually tested their accuracy using Google News article. The primarily automated aspect was tracking rumours and their spread online. Emergent.info used parsers to detect changes to the headline and body of text of articles on the story, highlighting the addition of new information and showing how rumours evolved across time. It also automatically tracked shared on social media and displayed the effect that debunking/confirmation had on sharing figures. The project was discontinued last year.

LazyTruth (2013) http://www.lazytruth.com/ 1 2

Provided 'verified rebuttal' to viral rumours shared on email. It firstly searched emails for unique phrases which appear in widespread viral emails. It then used its 'misinformation database' - a catalogue of tracked claims and corresponding factchecks by Politifact.com, Factcheck.org and other factchecking and urban rumour websites - to provide a pre-prepared 'rebuttal' to the rumour in question. The original source of the misinformation was included as well. LazyTruth was a Google Chrome browser extension, so this all took place within a user's email inbox, ensuring they needn't have searched elsewhere. The largely experimental project, overseen by developers drawn from several different organisations, was discontinued in the year it was launched.

ContentCheck (June 2015–present) 1 2

https://team.inria.fr/cedar/contentcheck/

A claim detection and analysis tool designed by Le Monde newspaper and a French academic research team. The technology, which is being developed in time for next year's

French presidential elections, automatically recognises claims made by politicians and then provides a 'suite of tools'—infographics, charts, graphs, explainers—on the facts in question. Its aim is not to directly analyse the claims or provide verdicts on their accuracy. As Xavier Tannier has said, 'These machines do not have the ability to reason or analyze, but will provide the desired material by the media.' Rather, its intelligence lies in the fact it can recognise certain types of claim and readily provide relevant information on the topic. Content Check's aim is equipping journalists to analyse and fact-check statements more effectively.

Stage 3: Reference approaches

Full Fact (2013–present) https://fullfact.org/automated 1 2 3 4

A monitoring platform, with industry standard media monitoring software used for claim recognition, with statistical claim ID and factchecking based on pattern recognition and structured data on top of that. The human factchecking and the automated factchecking will appear in 'Robocheck', which annotates TV subtitles with live factchecking. Full Fact also has a claim monitoring platform called Trends.

Stage 3: Machine learning approaches

Computational Fact Checking from Knowledge Networks (2015) 3

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128193

The complexities of human fact checking can be approximated quite well by finding the appropriately-defined shortest path between nodes on a mathematical representation called a knowledge graph. They took a database of facts from Wikipedia, restructured it into a knowledge graph, which is a network of points and joins where each point is something (like one of the Beckhams) and each join is a relationship (like marriage). The tool returns a truth score, a numerical probability that the claim is accurate.

Stage 3: Contextual approaches

Pheme (January 2014–January 2017) http://pheme.eu/ 1 2 3

A research project of various EU universities, aims to detect four kinds of conversation on social networks and online media - speculation, controversy, misinformation and disinformation—and model their diffusion across the internet. It uses natural language processing to scour user-generated content and identify widespread claims which fit the four categories mentioned above. It then checks those claims against trusted data sources, such as PubMed and Linked Open data, and attempts to assess their veracity. Pheme then examines the links between different posts to determine how claims have diffused across social networks: who started the claim; who has sent or shared it onwards; who has doubted or denied the claims, and so on.

Twitter Trails (March 2014–present) 1 2 3 http://twittertrails.wellesley.edu/~trails/

Designed by Trails of Propagation—a Wellesley University project—which assesses the accuracy of stories based on expressions of scepticism in tweets and patterns of dissemination. It relies on the 'wisdom of the crowd' assumption that, on one hand, users will react with scepticism towards the most dubious claims, and that, on the other, users won't re-tweet those claims which they believe to be inaccurate. The inverse is that claims are likely to be true if they attract relatively little scepticism and are widely shared.

Stage 5: Publishing tools

Notably, all these tools only present factchecks prepared by human factcheckers—for now.

Truth Teller (January 2013–June 2014) 1 2 4

https://www.washingtonpost.com/news/ask-the-post/wp/2013/09/25/announcing-truth-teller-beta-a-better-way-to-watch-political-speech/

A discontinued Washington Post factchecking product, attempted to factcheck political speeches in real-time. It converted audio to text, extracted factual claims from that text, and cross-referenced the claims in question against a database of previous factchecks, in order to realtime verdicts on the accuracy of the claims in question. However, the audio to text and natural language processing worked poorly, and the Washington Post's database of factchecks (drawn from its own in-house fact checkers as well as Politifact.com and Factcheck.org) was incomprehensive. The tool was discontinued shortly after being rolled out in early 2013.

See also a comparable manual tool developed by ClearerThinking.org and used for a project used Factchecking 2.0 (October 2015). http://www.lse.ac.uk/philosophy/blog/2015/10/26/stefan-schubert-and-clearerthinkings-fact-checking-2-0/

Truth Goggles (2011–2014) https://slifty.com/project/truth-goggles/ 1 2 4

A browser plugin which aimed to automatically detect and highlight dubious claims on a webpage. An early incarnation of the 'fact check the web' approach, users were supposed to click on highlighted text and then find a written factcheck in a floating sidebar on the right-hand of the screen. The problem with the tool was that its natural language processing was ineffective and its database of factchecks, drawn from US sites including Politifact.com and Factcheck.org, was not comprehensive enough. It eventually evolved into an annotation tool, the aim of which enabling journalists and other users to highlight text on webpages which they perceived to be inaccurate. The plugin is no longer in development, although it was an influential and early experiment in automated factchecking.

Niche factchecking

Statcheck (paper submitted 2015) https://mbnuijten.com/statcheck/

1 2 4

A software package which aims to increase the accuracy of null hypothesis testing in psychology. It automatically extracts statistics from articles, recomputes their p values and highlights potential errors. Inconsistent p-values can make a statistically non-significant result seem significant and vice versa, but recent research suggests half of all published empirical work in the field contains at least one inconsistent p-value. When such errors match up with the researcher's expectations, they can be included in the paper's results and allowed to influence its conclusions. Statcheck therefore mitigates against the possibility of systematic bias bleeding into research.

jEugene (January 2015–present) https://jeugene.com/

1 2 4

Legal software that combs through drafts of legal documents and detects potential errors. Finding these errors is immensely important because they can undermine the legality of a whole document, and yet they are often overlooked by the most skilled legal eyes. Like ClaimBuster, jEugene is based on machine learning systems. It applies similar techniques to large quantities of legal texts and highlights potential errors. Its aim is to save time and money by proofreading legal documents faster and more effectively than human.

Full Fact

fullfact.org

t: FullFact
f: FullFact.org