



How to stop people misusing statistics

Automatic verification of statistical claims

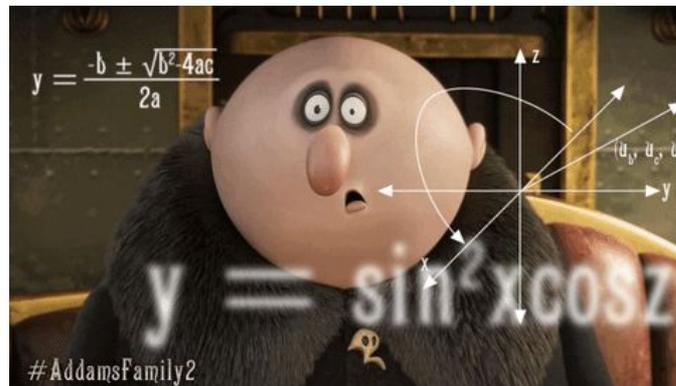
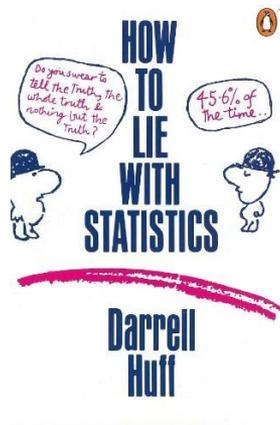
Bad information ruins lives

Automated Fact Checking

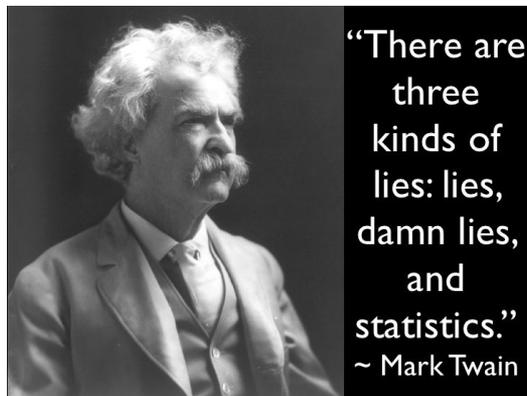
Claim detection

Claim matching

Claim checking



FULL
FACT



Office for
National Statistics

Consumer Prices Indices Technical Manual, 2019

This technical manual is a reference tool for anyone wanting to understand how measures of consumer price inflation and associated indices are compiled.

Page 1 of 116



"We now have 420,000 more people in employment than there were before the pandemic began."

Boris Johnson, 18 January 2022.



OUR VERDICT

False. The total number of people in employment, including the self-employed, is 600,000 lower than before the pandemic began.

THE ECONOMY / JOBS AND WORK

Boris Johnson makes false employment claim for ninth time in Parliament

21 APRIL 2022

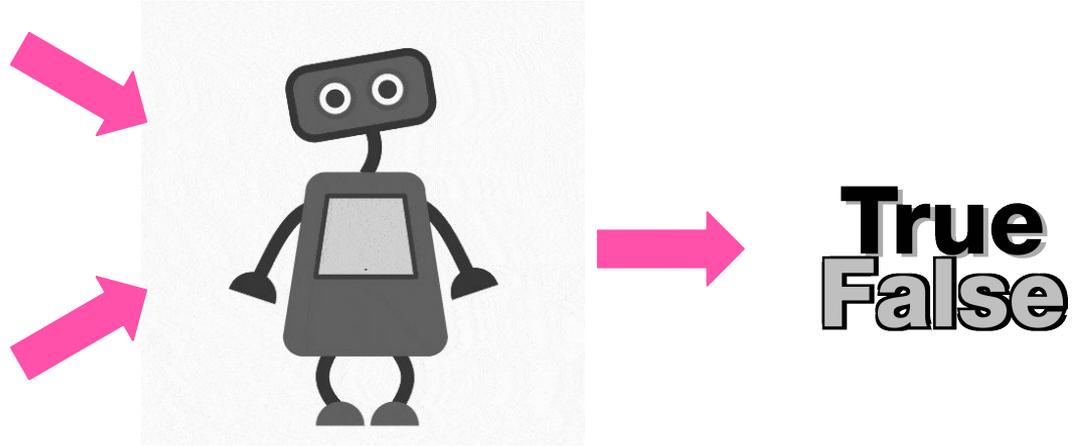
WHAT WAS CLAIMED

There are more people in work than there were before the pandemic.

OUR VERDICT

False. There are 600,000 fewer people in work than there were before the pandemic.

The dream of automated verification



		CPIH Index (UK, 2015 = 100)	CPIH 12-month rate
2021	Apr	110.4	1.6
	May	111.0	2.1
	Jun	111.4	2.4
	Jul	111.4	2.1
	Aug	112.1	3.0
	Sep	112.4	2.9

Office for National Statistics



Government of India
Ministry of Statistics and Programme Implementation

Full Fact's Stats Checker

Analyses claims from UK newspapers, Hansard & social media

Uses a trained BERT model to extract information from each claim

Compares that to official ONS statistics



Inflation reached 5.4 percent in December as the British public found they were forced to find more money for the weekly food shop, fuel and energy



Linguistic logic (BERT)

“inflation”: topic

“5.4%”: number

“5.4%” refers to “inflation”

“December” relates to “inflation”

Business logic (Python)

“inflation” means “CPI” (or RPI, CPIH...)

Inflation has a checkable source:

<https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7g7/mm23>

“December” means December 2021, a point in the CPI time series

Examples of success

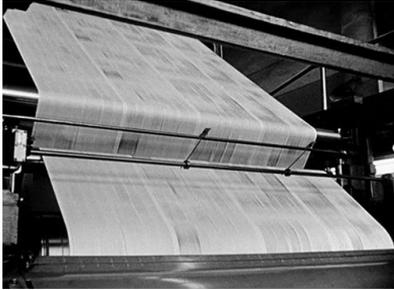
Claim	Stats Checker output
<p>Consumer prices inflation, according to the ONS, has now hit 9 per cent and is forecast to go higher still.</p>	<p>True. Latest rate for ons_inflation_cpi_rates is: 9.0</p>
<p>In a further headache for ministers, the RPI measure of inflation has rocketed even higher to 11.1 per cent in April - with unions threatening strikes unless that is used as the basis for pay rises in the public sector.</p>	<p>True. The rate for ons_inflation_rpi_rates 2022-4 is: 11.1</p>
<p>New figures out this morning, as we have heard, show that today, the UK's unemployment rate is less than half that, at 3.7%, the lowest in almost half a century.</p>	<p>True. Latest rate for ons_unemployed_monthly_rate is: 3.7</p>
<p>The Consumer Price Index (a measure of inflation) hit 6.7 percent in the UK in April.</p> <p>fullfact.org</p>	<p>False. The rate for ons_inflation_cpih_rates 2022-4 is: 7.8. The rate for ons_inflation_cpi_rates 2022-4 is: 9.0. The rate for ons_inflation_rpi_rates 2022-4 is: 11.1</p>

Examples of failure

Claim	Reason for failure
<p>For example, if something that cost £1 a year ago now costs £1.02 today, that rate of inflation would be 2%.</p>	<p>Hypothetical example (<i>not claiming to be true today</i>)</p>
<p>Policymakers at the Bank have raised interest rates to 1 per cent, with hikes at each of its past four meetings, to try to cool rampant inflation.</p>	<p>Different topic (<i>inflation only mentioned incidentally</i>)</p>
<p>Albanese has said wages should keep pace with inflation - currently at 5.1 per cent - and has called for a \$1 rise in the minimum wage of \$20.33.</p>	<p>Different country (<i>Anthony Albanese is the Prime Minister of Australia</i>)</p>

What's next?

Can we make the **claims** better?

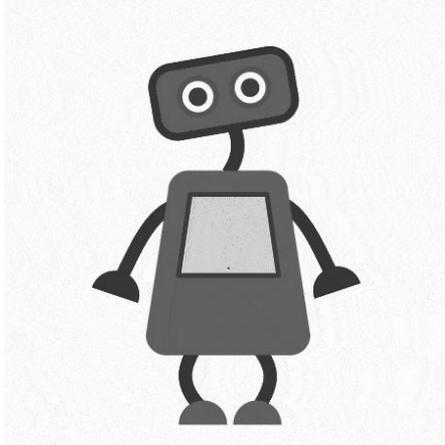


Language is always ambiguous and communication relies on implied, shared context

But we can do better at identifying ‘checkable’ claims (and not try and check the rest)

Extract more information from the wider context
Speaker, audience, shared understanding

Can we make the algorithms better?



There is a lot of hype about AI and large language models

We are 20-30 years away from “artificial general intelligence”

and have been for 70 years...

But BERT was a huge breakthrough making text analytics a lot more powerful and accessible

So with more training data we can extract information more reliably from more sentences

We can extend the business logic to cover more cases

Can we make the data better?



All NSOs publish statistical data about their country

Different countries have different resources and needs

Best practices can guide publication policies:

- Make data discoverable
- Make datasets legible
- Data should be useful for everyone
 - Different users have different needs
- Present data with context

		CPIH Index (UK, 2015 = 100)	CPIH 12-month rate
2021	Apr	110.4	1.6
	May	111.0	2.1
	Jun	111.4	2.4
	Jul	111.4	2.1
	Aug	112.1	3.0
	Sep	112.4	2.9

Data Publishing Guide

- “A living document with guidance for national statistics institutes on how to publish data so that fact checkers can do their jobs better.”
- Report based on interviews with fact checkers, data scientists and others who use NSO data
- Guidance on communications, file formats, fact checking, APIs etc.

Can we stop people misusing statistics?

Probably not. But let's make misuse harder and correct use easier!

If you take away one thing: **Talk to each other!**

NSOs are often trying to do the same thing

You don't need to solve every problem from scratch

Further reading

<p>A Data Publishing Guide by Full Fact, Africa Check, Chequeado and the ODI</p>	<p>https://fullfact.github.io/data-publishing-guide/docs/about/</p>
<p>Code to fetch and reformat data from several NSO website by ODI and Full Fact</p>	<p>https://github.com/FullFact/nso-stats-fetcher</p>
<p>More about Full Fact's tools for automated fact checking</p>	<p>https://fullfact.org/blog/2021/jul/how-does-automated-fact-checking-work/</p>
<p>Tidy data - a simple format for sharing data by Hadley Wickham</p>	<p>https://vita.had.co.nz/papers/tidy-data.html</p>
<p>Some open data publishing principles by Leigh Dodds</p>	<p>https://digitalblog.ons.gov.uk/2017/01/06/some-open-data-publishing-principles/</p>

Full Fact

17 Oval Way
London
SE11 5RR



fullfact.org/contact



+44 (0)20 3397 5140



@FullFact

fullfact.org

Thanks for listening!

david.corney@fullfact.org

FULL

FACT