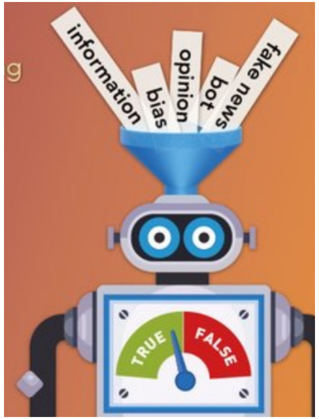
The Magnitude of Truth: On Using Magnitude Estimation for Truthfulness Assessment in Fact-Checking

Stefano Mizzaro

Truth is in the Eyes of the Machines
Amsterdam, 9/5/2025



UNIVERSITÀ DEGLI STUDI DI UDINE
HIC SUNT FUTURA

https://tinyurl.com/5ykcd389

Ack

- This research has been partially supported by project PRIN 2022 "MoT – The Measure of Truth: An Evaluation-Centered Machine-Human Hybrid Framework for Assessing Information Truthfulness" - Codice n. 20227F2ZN3 CUP n. G53D23002800006 "Finanziato dall'Unione Europea – Next-Generation EU – PNRR M4 C2 I1.1" RS Mizzaro

Outline

- Intro
- Crowdsourcing for fact-checking
- Truthfulness Scales and Magnitude estimation

Outline

- Intro
- Crowdsourcing for fact-checking
- Truthfulness Scales and Magnitude estimation

Aims

- Our overall approach → Fact-checking, we use crowdsourcing
- A bit of the story → Many people, many papers
- Focus on truth scales → Magnitude Estimation

The team

- Isabelle Augenstein
- Francesco Bombassei De Bona
- Davide Ceolin
- Alessandro Checco
- Vincenzo Della Mea
- Gianluca Demartini
- Massimiliano De Luise
- Tim Draws
- Shaoyang Fan
- David La Barbera
- Joel Mackenzie
- Eddy Maddalena
- Beatrice Portelli
- Yunke Qu
- Kevin Roitero
- Giuseppe Serra
- Michael Soprano
- Damiano Spina
- Denis Eduard Tapu
- Dustin Wright
- Arkaitz Zubiaga

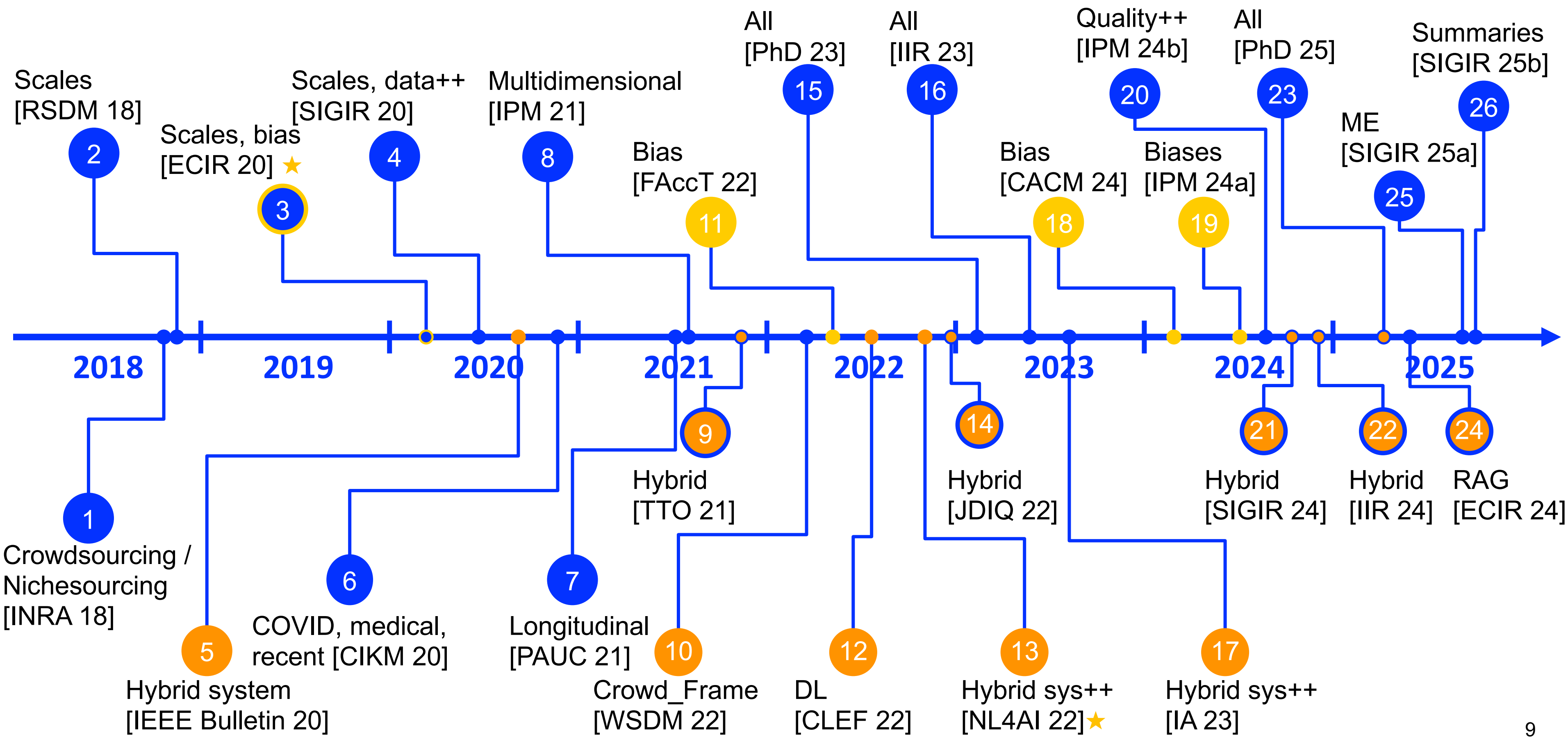
The team

- Isabelle Augenstein
- Francesco Bombassei De Bona
- **Davide Ceolin**
- Alessandro Checco
- Vincenzo Della Mea
- **Gianluca Demartini**
- Massimiliano De Luise
- Tim Draws
- Shaoyang Fan
- **David La Barbera**
- Joel Mackenzie
- **Eddy Maddalena**
- Beatrice Portelli
- Yunke Qu
- **Kevin Roitero**
- Giuseppe Serra
- **Michael Soprano**
- **Damiano Spina**
- Denis Eduard Tapu
- Dustin Wright
- Arkaitz Zubiaga

Many people, many papers

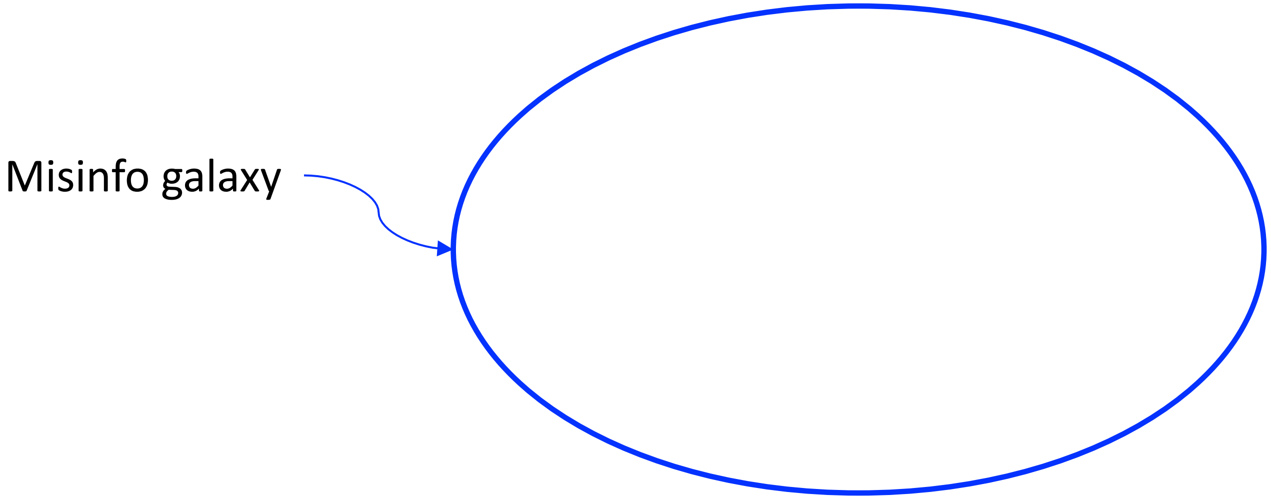
- The gang has been working!
- 7 – 8 years, 22 people → 26 papers
  - (at least)
- How to present them?

Timeline


1 Crowdsourcing / Nichesourcing [INRA 18]
2 Scales [RSDM 18]
3 Scales, bias [ECIR 20] ★
4 Scales, data++ [SIGIR 20]
5 Hybrid system [IEEE Bulletin 20]
6 COVID, medical, recent [CIKM 20]
7 Longitudinal [PAUC 21]
8 Multidimensional [IPM 21]
9 Hybrid [TTO 21]
10 Crowd_Frame [WSDM 22]
11 Bias [FAccT 22]
12 DL [CLEF 22]
13 Hybrid sys++ [NL4AI 22]★
14 Hybrid [JDIQ 22]
15 All [PhD 23]
16 All [IIR 23]
17 Hybrid sys++ [IA 23]
18 Bias [CACM 24]
19 Biases [IPM 24a]
20 Quality++ [IPM 24b]
21 Hybrid [SIGIR 24]
22 Hybrid [IIR 24]
23 All [PhD 25]
24 RAG [ECIR 24]
25 ME [SIGIR 25a]
26 Summaries [SIGIR 25b]
2018
2019
2020
2021
2022
2023
2024
2025

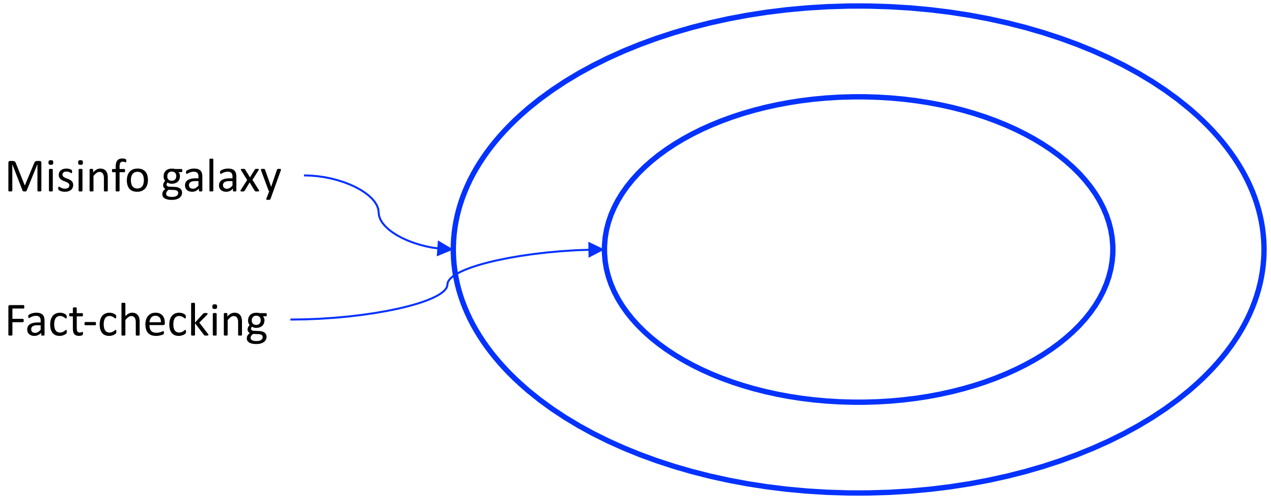
Misinformation

- Misinformation & Disinformation are spreading
- Need for researchers and society to find countermeasures



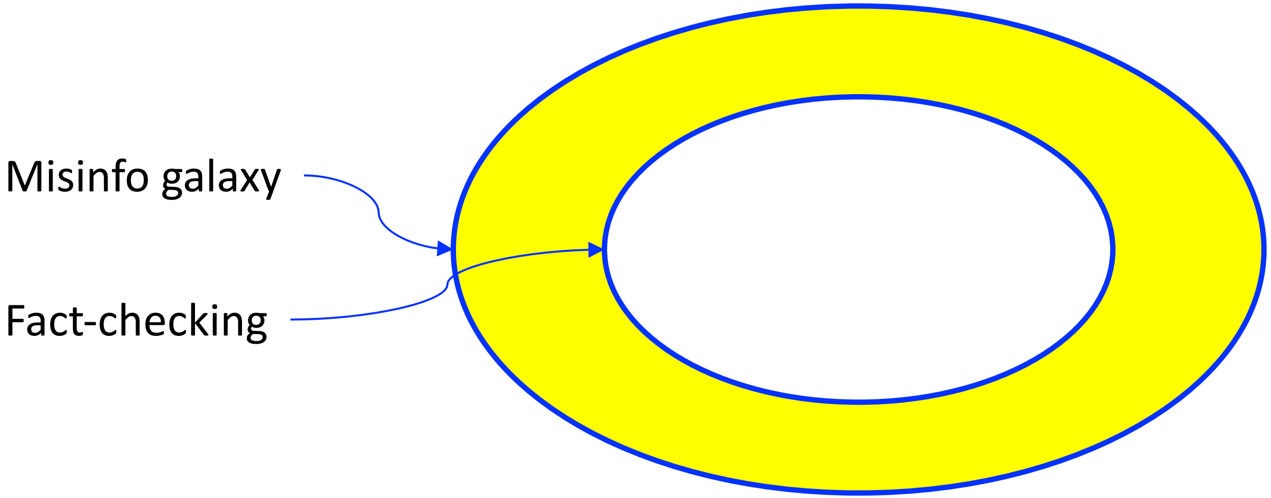
Fact-checking

- The process aimed at determining if a claim/info/news/etc. is true or not
- Rather complex process
  - Experts, follow a protocol, seek evidence, discuss, reach a verdict



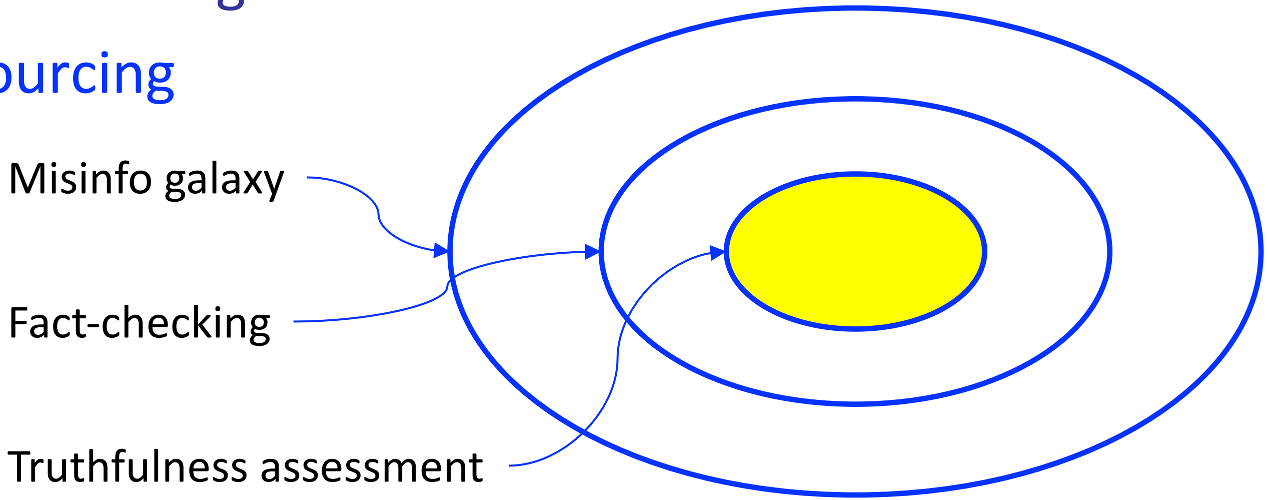
Beyond Fact-checking

- Of course there's much more
  - E.g., how the fake news spread when compared to the true ones?
  - Echo chambers
  - Filter bubbles
  - Etc.



Truthfulness assessment

- We focus on "Is this true or not?"
- And on how to answer that question
- Usually done by experts
- A lot of research on using AI
- We use crowdsourcing



Outline

- Intro
- Crowdsourcing for fact-checking
- Truthfulness Scales and Magnitude estimation

Crowdsourcing?

Crowdsourcing

文A 45 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

"Crowd work" redirects here. For the performing arts term, see audience participation.

This article **is written like a personal reflection, personal essay, or argumentative essay** that states a Wikipedia editor's personal feelings or presents an original argument about a topic. Please help improve it by rewriting it in an encyclopedic style. *(September 2022)* (*Learn how and when to remove this message*)

Crowdsourcing involves a large group of dispersed participants contributing or producing goods or services—including ideas, votes, micro-tasks, and finances—for payment or as volunteers. Contemporary crowdsourcing often involves digital platforms to attract and divide work between participants to achieve a cumulative result. Crowdsourcing is not limited to online activity, however, and there are various historical examples of crowdsourcing. The word crowdsourcing is a portmanteau of "crowd" and "outsourcing".[1][2][3] In contrast to outsourcing, crowdsourcing usually involves less specific and more public groups of participants.[4][5][6]

Crowdsourcing – Definition

- "taking a task traditionally performed by an employee or contractor, and outsourcing it to an undefined, generally large group of people or community in the form of an open call"
- [http://en.wikipedia.org/wiki/Crowdsourcing]

Crowdsourcing – Definition

- "The practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers"
- [Merriam-Webster]

Crowdsourcing – Definition

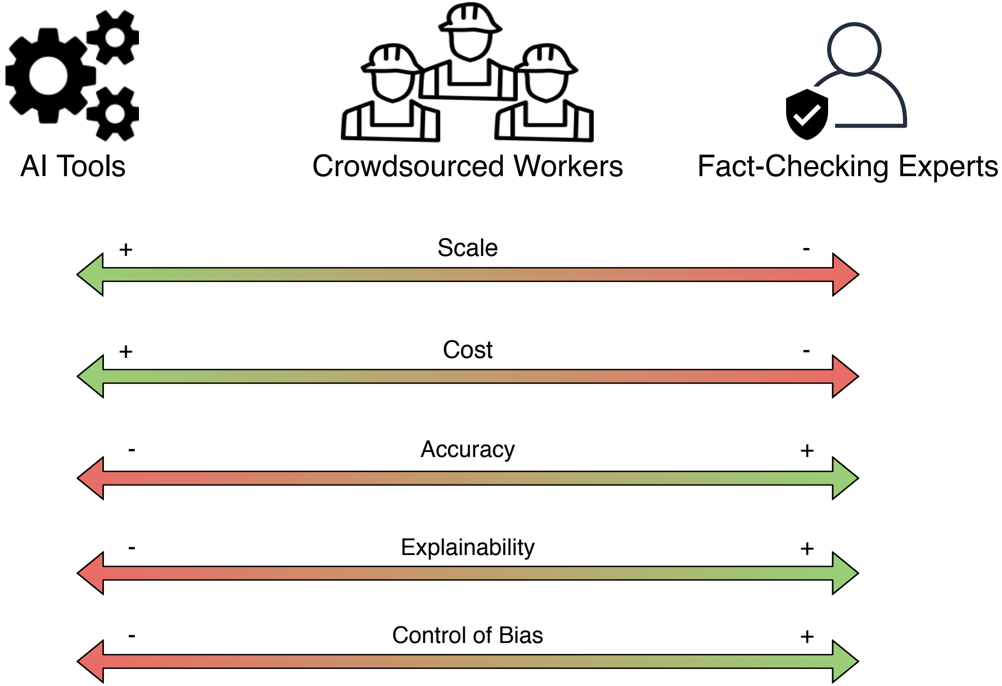
- "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an **open call**. This can take the form of peer-production (when the job is performed **collaboratively**), but is also open undertaken by sole **individuals**. The crucial prerequisite is the use of the open call format and the **large network of potential laborers**."
- [Howe, 2006]

Motivation / Justification

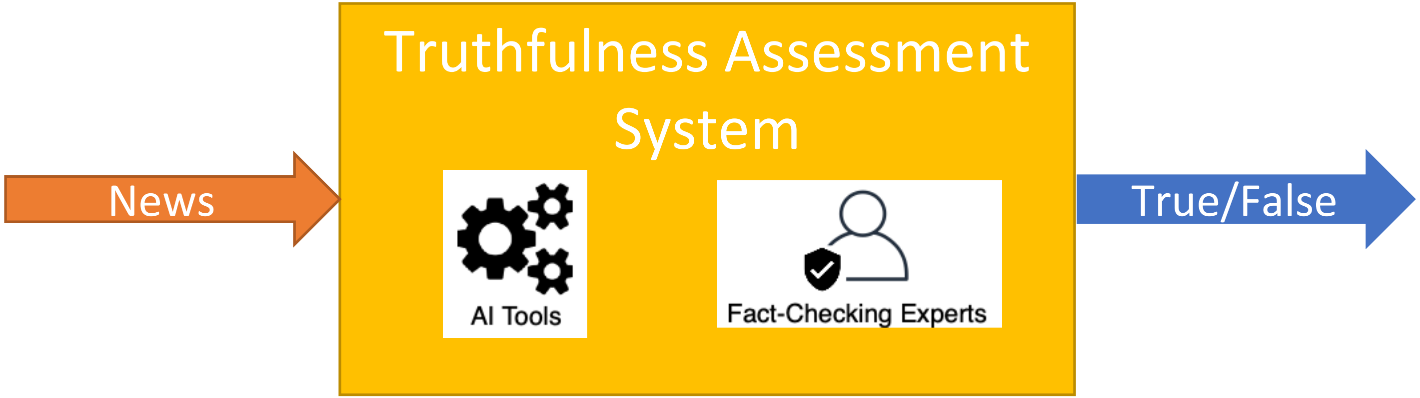
- So, we propose to crowdsource truthfulness assessment
- Does it make sense, and why?
  - In principle

AI vs. Experts

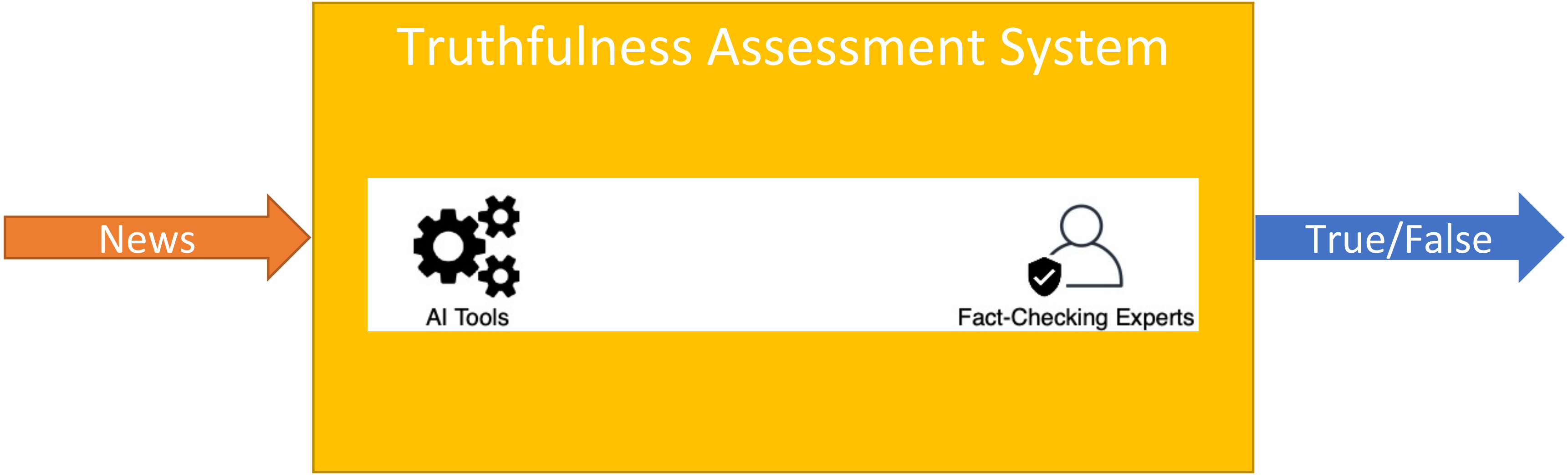
AI vs. Crowd vs. Experts



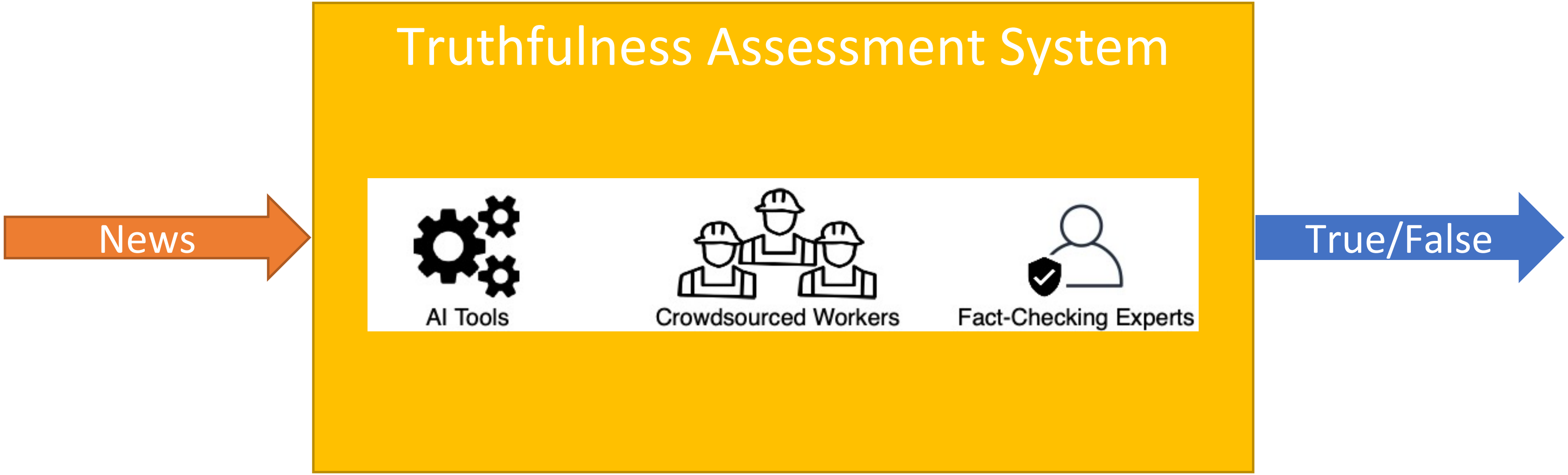
AI + Experts


News
Truthfulness Assessment System
AI Tools
Fact-Checking Experts
True/False

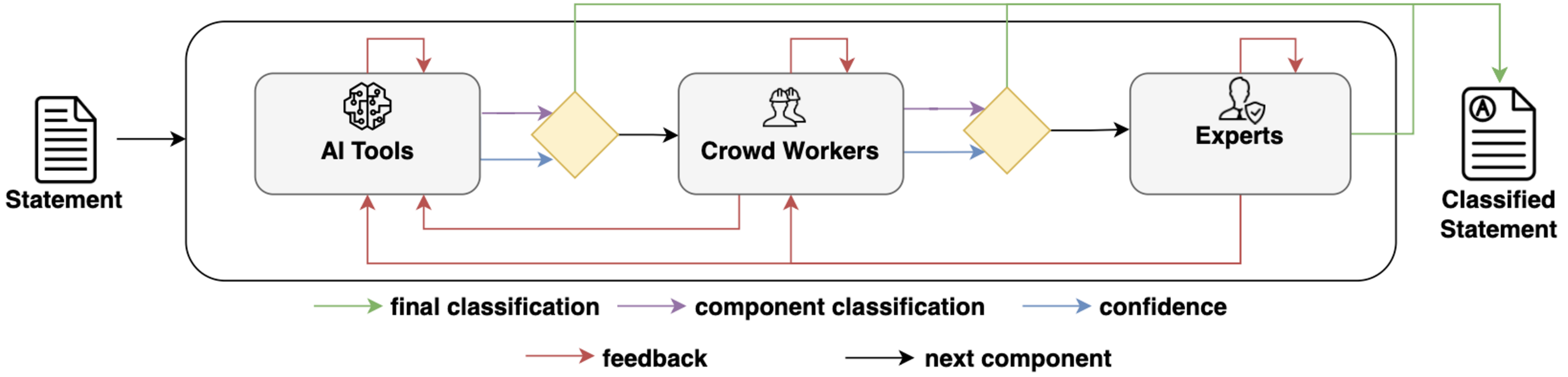
AI + Crowd + Experts



AI + Crowd + Experts



AI + Crowd + Experts



After the panel

- I did not plan to spend too much time on this but after the panel...
- "It is impossible to reach all the users that have been reached by misinformation if platforms do not cooperate" [Giovanni]
- Well, it is impossible anyway given that we do not have enough fact-chekers!

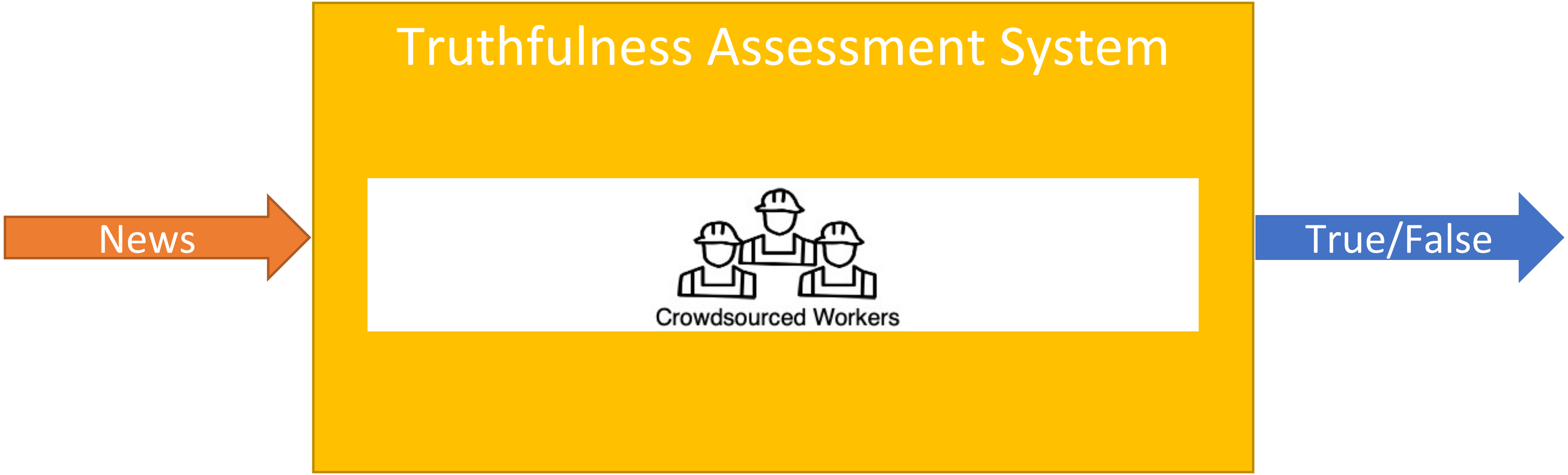
AI + Experts

- Stage 1: AI
  - Output: Truthfulness value + Confidence
  - (+ more: explanations…)
- Stage 2 (if low Confidence): Experts
  - Output: Truthfulness value + Confidence
  - (+ more: explanations, motivations, …)

AI + Crowd + Experts

- Stage 1: AI
  - Output: Truthfulness value + Confidence
  - (+ more: explanations…)
- Stage 2 (if low Confidence): Crowd
  - Output: Truthfulness value + Confidence
  - (+ more: explanations, motivations, …)
- Stage 3 (if low Confidence): Experts
  - Output: Truthfulness value + Confidence
  - (+ more: explanations, motivations, …)

AI + Crowd + Experts



Main RQ

- Can crowdsourcing work?
- Can the crowd be put in the right conditions to assess truthfulness?

Overall approach

- We take some statements, with ground truth
- We ask crowd workers to assess the truthfulness
  - In a sort of "controlled" situation
- We compute agreement with ground truth

- (several variations / versions)

Experimental design (one version)

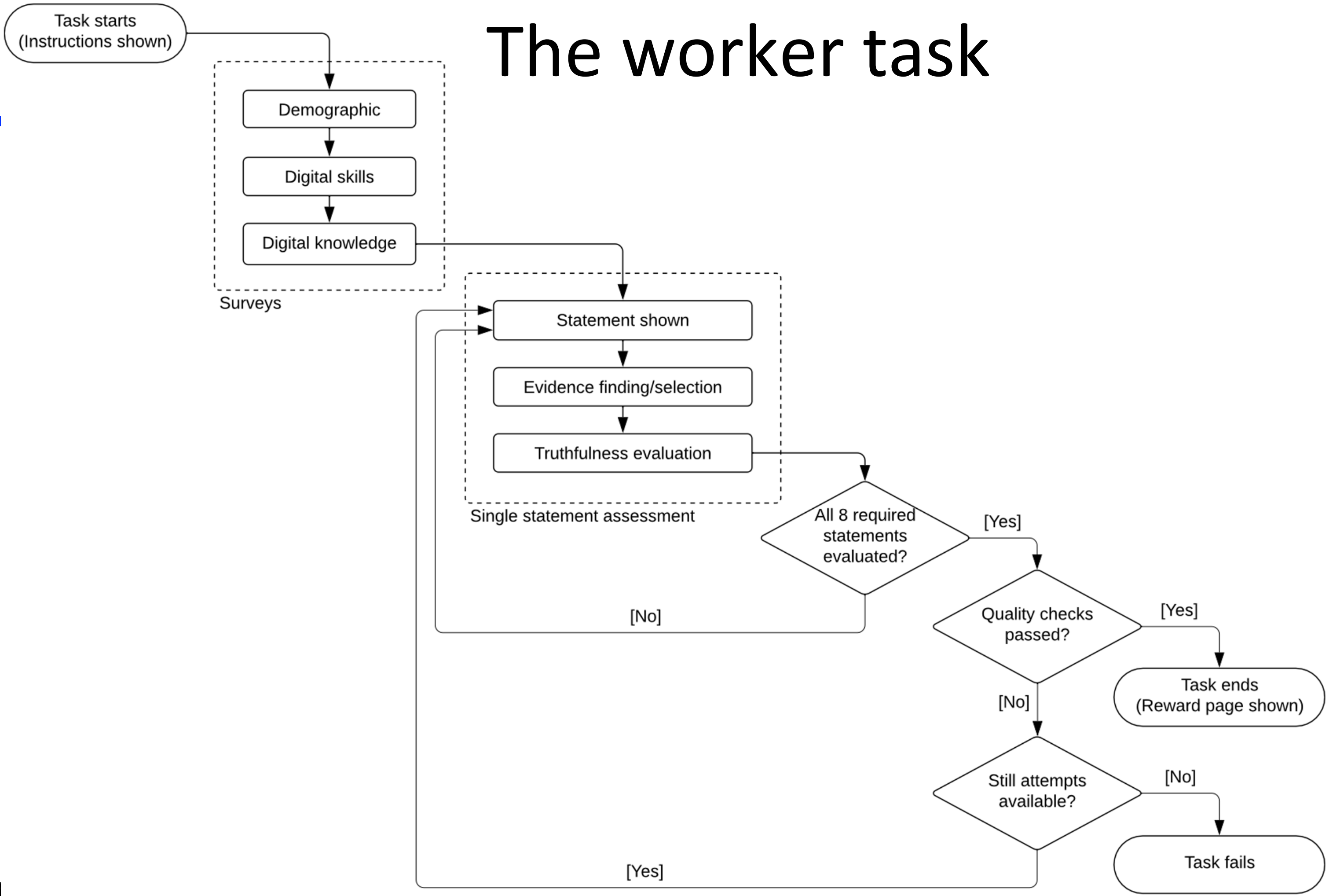
- Crowdsourcing platform:
  - Amazon's Mechanical Turk
  - Prolific
- Each worker judges the truthfulness of 8 (6 + 2) statements
  - (randomization to avoid bias)
- Redundancy
  - Each statement judged by 10 workers
- Quality checks:
  - 2 statements are equal for all workers: one clearly true and one clearly false
  - Time, Actions
  - …

Experimental design (one version)

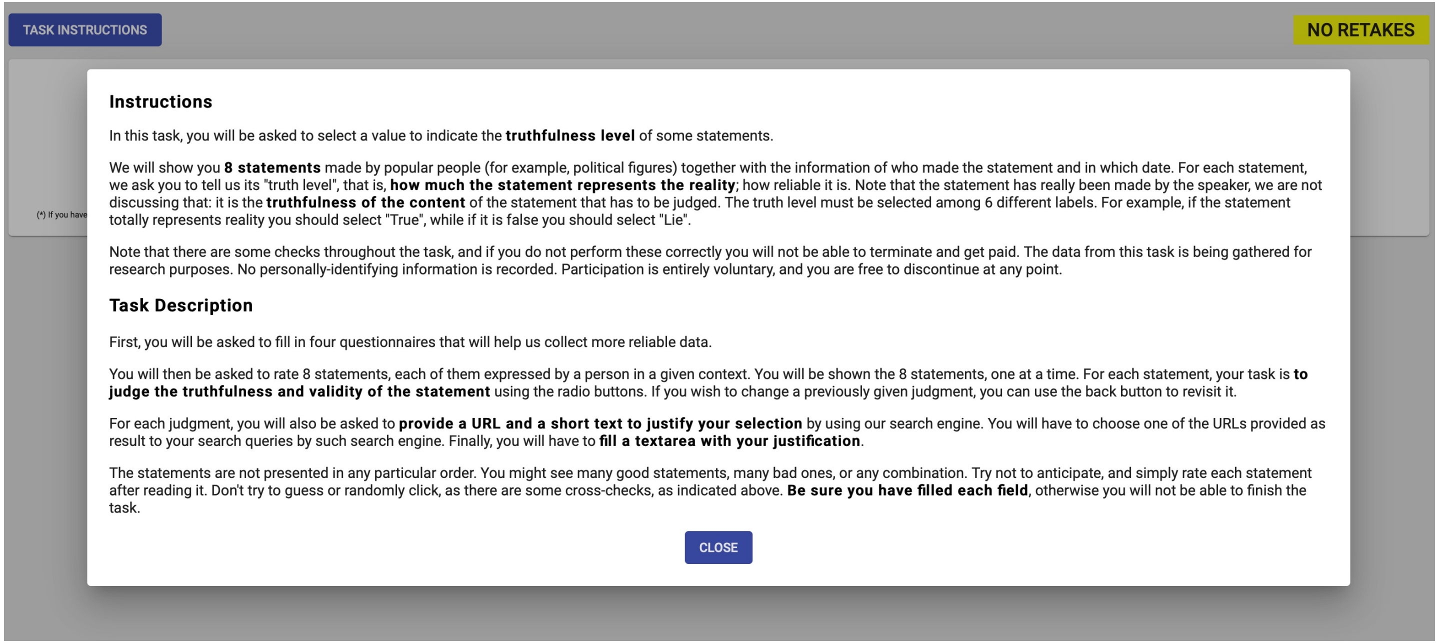
- Datasets
  - Politifact (https://www.politifact.com)
  - (But not only)
- In total
  - 120 statements, 20 for each Politifact category (+ the two for quality check)
  - 120 / 6 * 10 = 200 workers, each expressing 8 assessments, 1,600 assessments in total
- We also ask for evidence
  - Custom search engine to avoid workers finding Politifact pages with evidence

POLITIFACT
The Poynter Institute

The worker task



Screenshots – What the workers are doing



TASK INSTRUCTIONS

NO RETAKES

1 Questionna... 2 Questionna... 3 Questionna... 4 Questionna... 5 Docume... 6 Docume... 7 Docume... 8 Docume... 9 Docume... 10 Docume... 11 Docume... 12 Docume... 13 Outco...

Please answer each question

1) What is your age range?

- 0-18
- 19-25
- 26-35
- 36-50
- 50-80
- 80

2) What is the highest level of school you have completed or the highest degree you have received?

- High school incomplete or less
- High school graduate or GED (includes technical/vocational training that doesn't towards college credit)
- Some college (some community college, associate's degree)
- Four year college degree/bachelor's degree
- Some postgraduate or professional schooling, no postgraduate degree
- Postgraduate or professional degree, including master's, doctorate, medical or law degree

3) Last year what was your total family income from all sources, before taxes?

- Less than 10,000
- 10,000 to less than 20,000
- 20,000 to less than 30,000
- 30,000 to less than 40,000
- 40,000 to less than 50,000
- 50,000 to less than 75,000
- 75,000 to less than 100,000
- 100,000 to less than 150,000
- 150,000 or more

4) In general, would you describe your political views as

- Very conservative
- Conservative

- Some postgraduate or professional schooling, no postgraduate degree
- Postgraduate or professional degree, including master's, doctorate, medical or law degree

3) Last year what was your total family income from all sources, before taxes?

- Less than 10,000
- 10,000 to less than 20,000
- 20,000 to less than 30,000
- 30,000 to less than 40,000
- 40,000 to less than 50,000
- 50,000 to less than 75,000
- 75,000 to less than 100,000
- 100,000 to less than 150,000
- 150,000 or more

4) In general, would you describe your political views as

- Very conservative
- Conservative
- Moderate
- Liberal
- Very liberal

5) In politics today, do you consider yourself a

- Republican
- Democrat
- Independent
- Something else

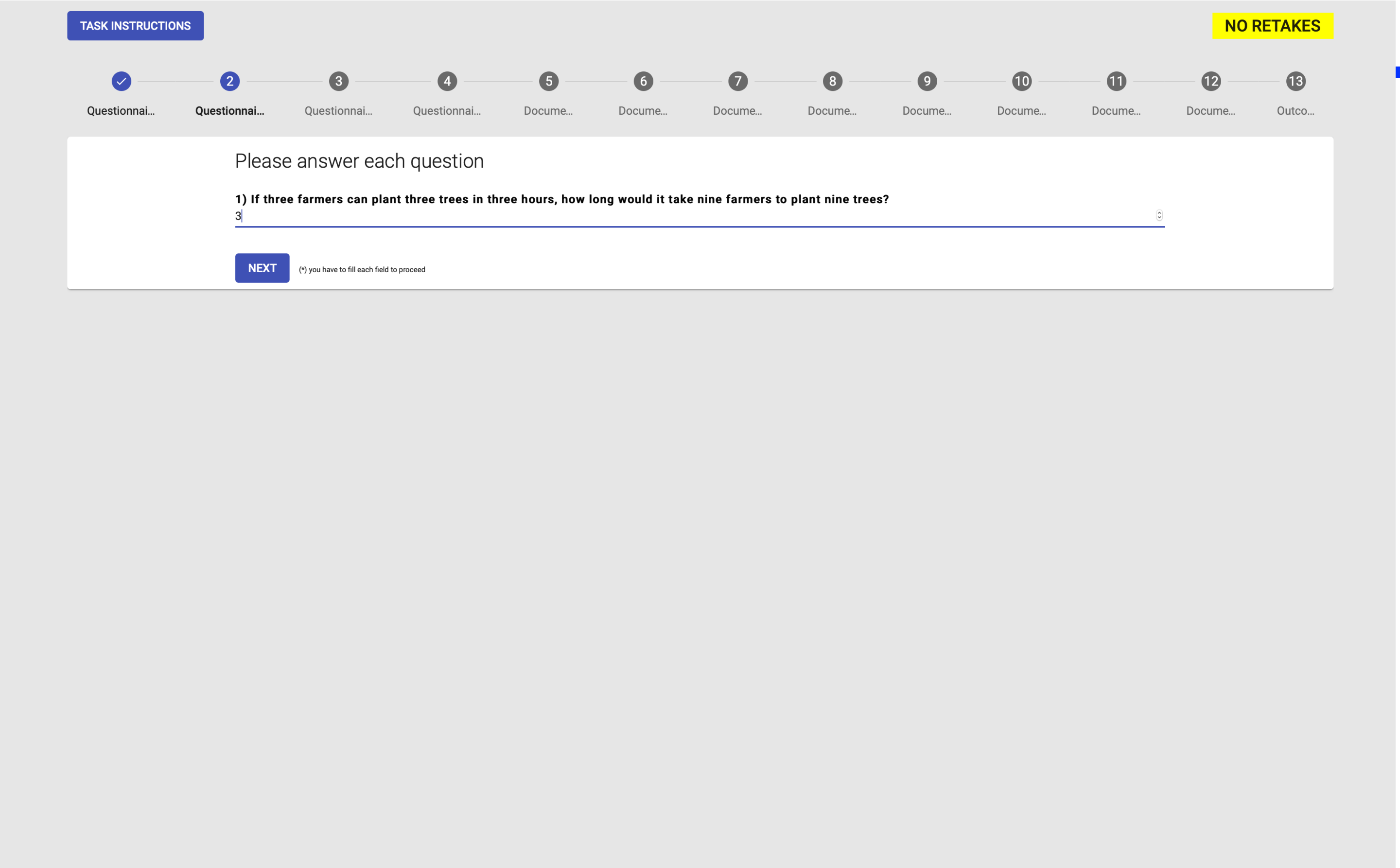
6) Should the U.S. build a wall along the southern border?

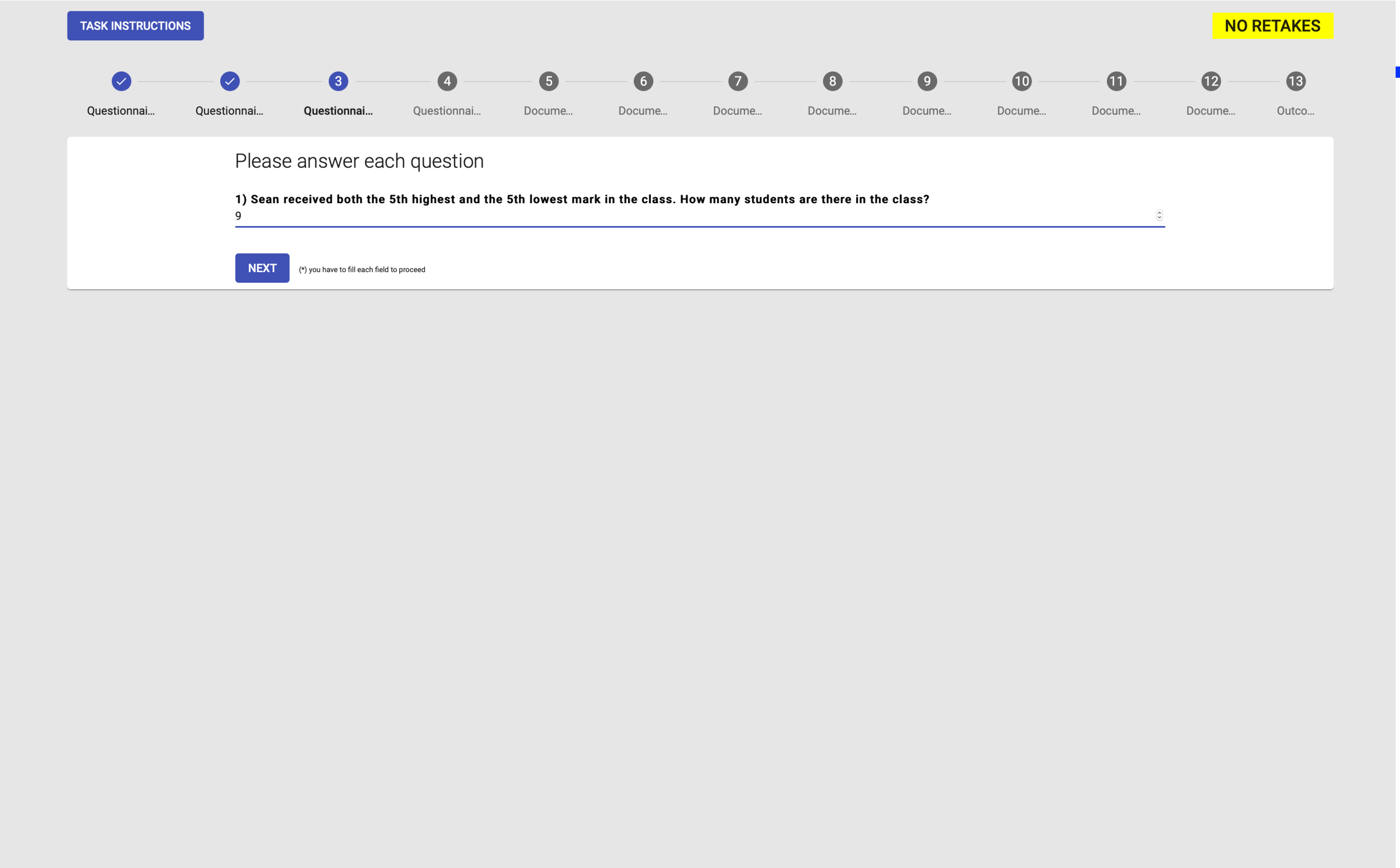
- Agree
- Disagree
- No opinion either way

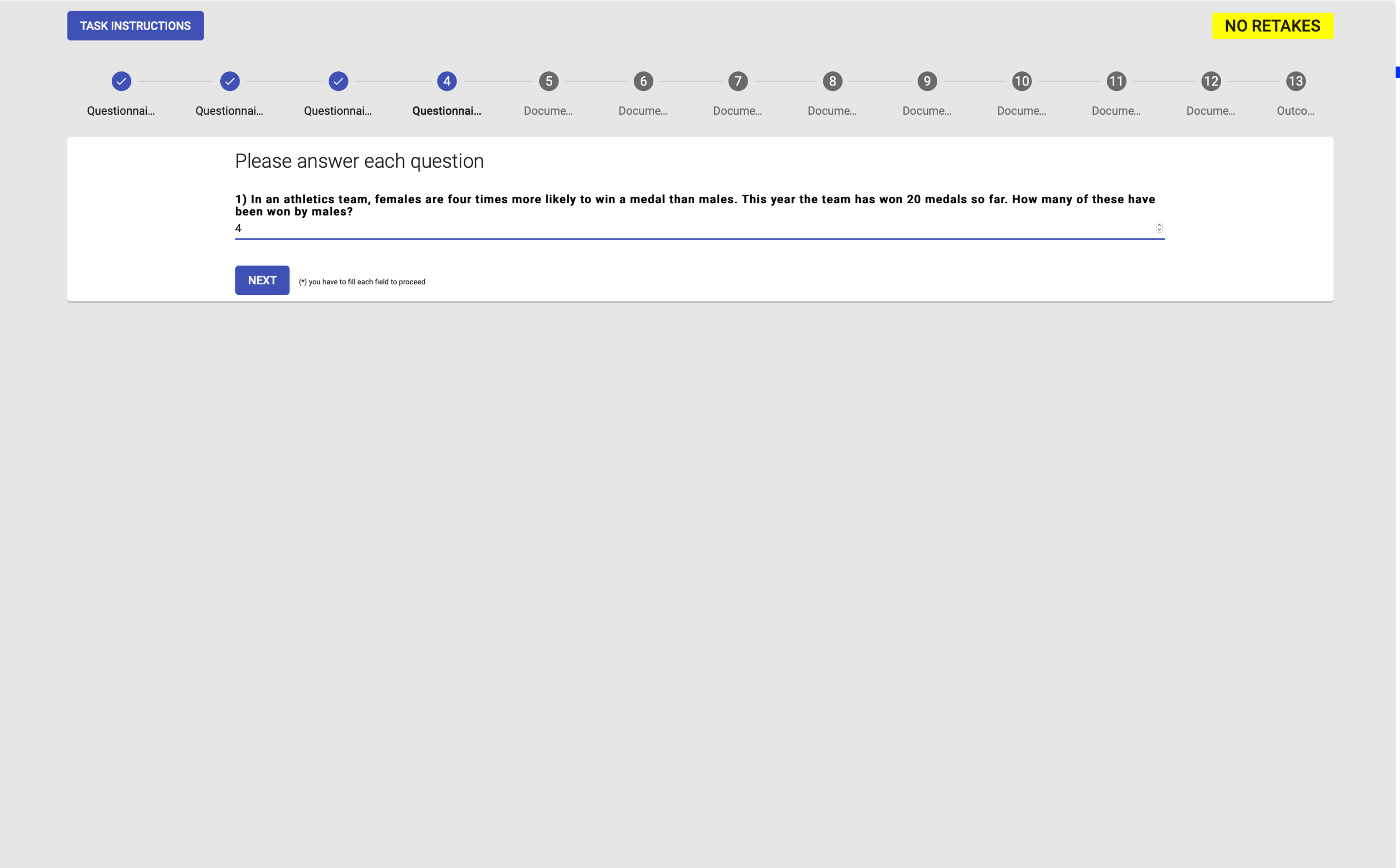
7) Should the government increase environmental regulations to prevent climate change?

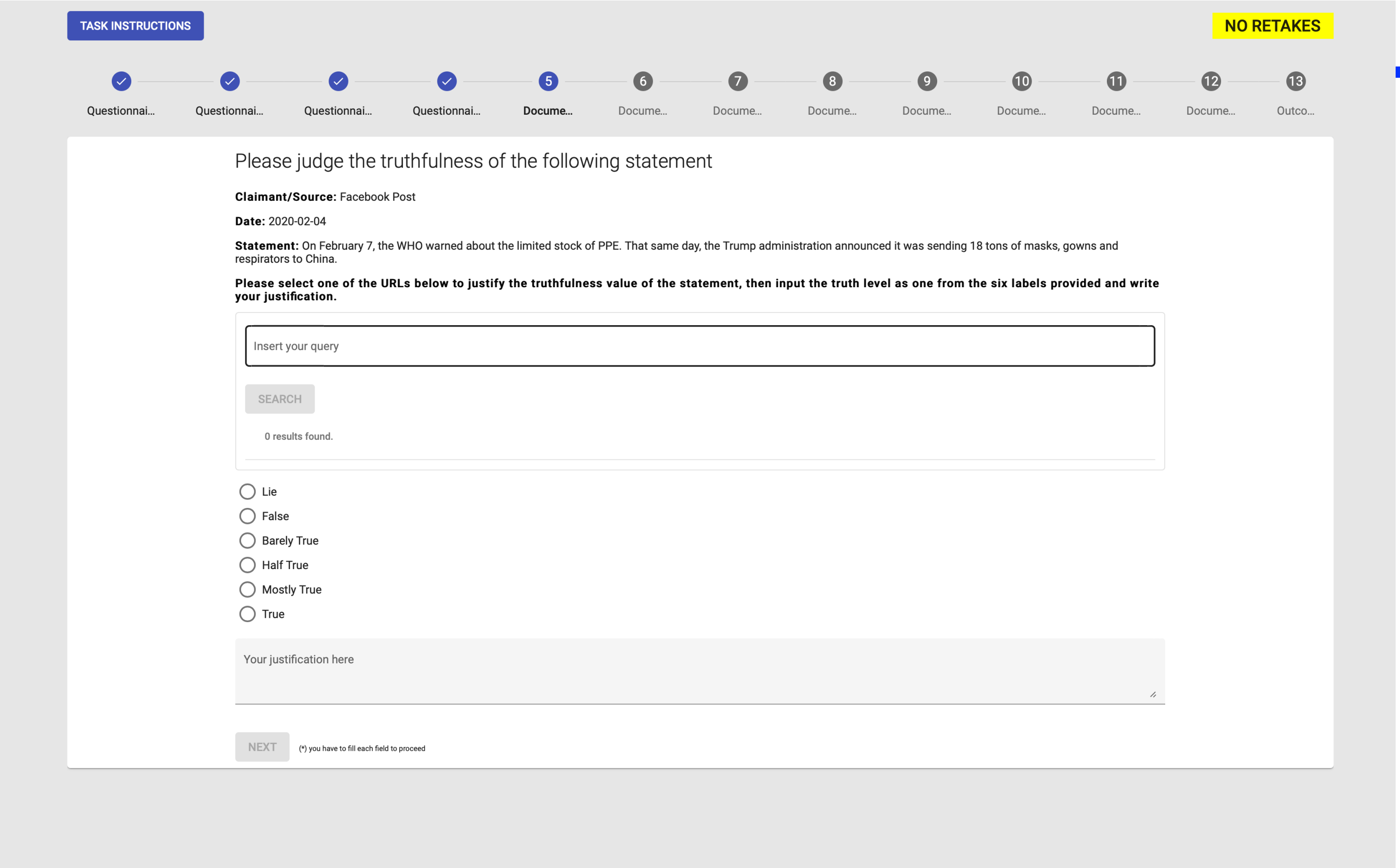
- Agree
- Disagree
- No opinion either way

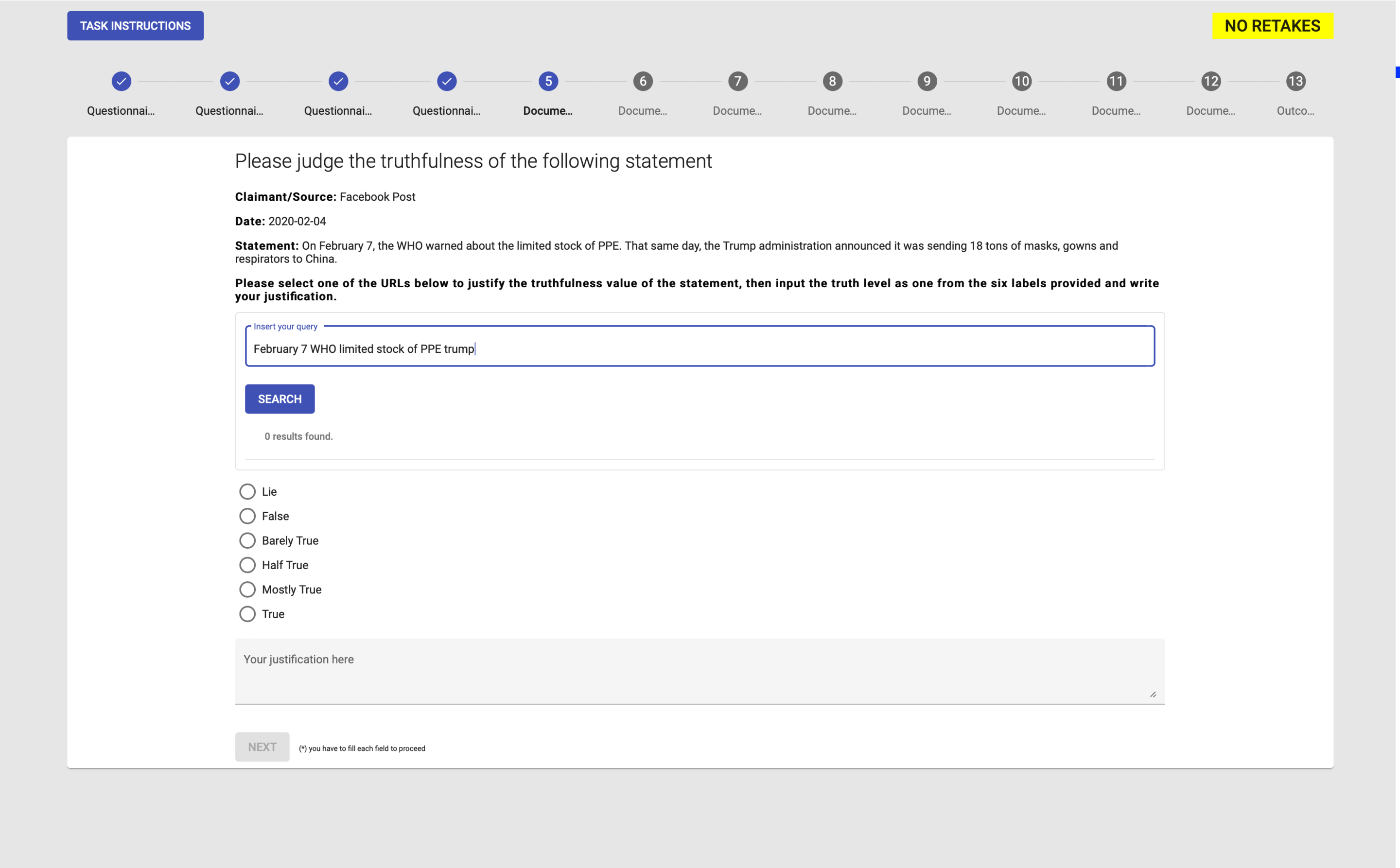
NEXT (*) you have to fill each field to proceed


TASK INSTRUCTIONS
NO RETAKES
2
3
4
5
6
7
8
9
10
11
12
13
Questionnai...
Questionnai...
Questionnai...
Questionnai...
Docume...
Docume...
Docume...
Docume...
Docume...
Docume...
Docume...
Docume...
Outco...
Please answer each question
1) If three farmers can plant three trees in three hours, how long would it take nine farmers to plant nine trees?
3
NEXT
(*) you have to fill each field to proceed


TASK INSTRUCTIONS
NO RETAKES
Questionnai...
Questionnai...
3 Questionnai...
4 Questionnai...
5 Docume...
6 Docume...
7 Docume...
8 Docume...
9 Docume...
10 Docume...
11 Docume...
12 Docume...
13 Outco...
Please answer each question
1) Sean received both the 5th highest and the 5th lowest mark in the class. How many students are there in the class?
9
NEXT
(*) you have to fill each field to proceed


TASK INSTRUCTIONS
NO RETAKES
Questionnai...
Questionnai...
Questionnai...
4
Questionnai...
5
Docume...
6
Docume...
7
Docume...
8
Docume...
9
Docume...
10
Docume...
11
Docume...
12
Docume...
13
Outco...
Please answer each question
1) In an athletics team, females are four times more likely to win a medal than males. This year the team has won 20 medals so far. How many of these have been won by males?
4
NEXT
(*) you have to fill each field to proceed


TASK INSTRUCTIONS
NO RETAKES
Questionnai...
Questionnai...
Questionnai...
Questionnai...
5 Docume...
6 Docume...
7 Docume...
8 Docume...
9 Docume...
10 Docume...
11 Docume...
12 Docume...
13 Outco...
Please judge the truthfulness of the following statement
Claimant/Source: Facebook Post
Date: 2020-02-04
Statement: On February 7, the WHO warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.
Please select one of the URLs below to justify the truthfulness value of the statement, then input the truth level as one from the six labels provided and write your justification.
Insert your query
SEARCH
0 results found.
Lie
False
Barely True
Half True
Mostly True
True
Your justification here
NEXT
(*) you have to fill each field to proceed


TASK INSTRUCTIONS
NO RETAKES
Questionnai... Questionnai... Questionnai... Questionnai...
5 Docume...
6 Docume...
7 Docume...
8 Docume...
9 Docume...
10 Docume...
11 Docume...
12 Docume...
13 Outco...
Please judge the truthfulness of the following statement
Claimant/Source: Facebook Post
Date: 2020-02-04
Statement: On February 7, the WHO warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.
Please select one of the URLs below to justify the truthfulness value of the statement, then input the truth level as one from the six labels provided and write your justification.
Insert your query
February 7 WHO limited stock of PPE trump
SEARCH
0 results found.
Lie
False
Barely True
Half True
Mostly True
True
Your justification here
NEXT
(*) you have to fill each field to proceed

TASK INSTRUCTIONS

NO RETAKES

Questionna... Questionna... Questionna... Questionna... 5 Docume... 6 Docume... 7 Docume... 8 Docume... 9 Docume... 10 Docume... 11 Docume... 12 Docume... 13 Outco...

Please judge the truthfulness of the following statement

Claimant/Source: Facebook Post

Date: 2020-02-04

Statement: On February 7, the WHO warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.

Please select one of the URLs below to justify the truthfulness value of the statement, then input the truth level as one from the six labels provided and write your justification.

Insert your query

February 7 WHO limited stock of PPE trump

SEARCH

46 results found.

Did the Trump Administration Send 18 Tons of PPE to China ...
https://www.snopes.com/fact-check/us-tons-ppe-china/
Did the Trump Administration Send 18 Tons of PPE to China in Early 2020? ... of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of ...

Select

Trump administration sent protective medical gear to China ...
https://www.msn.com/en-us/health/medical/trump-administration-sent-protective-medical-gear-to-china-while-he-was-minimizing-the-virus-threat-to-us/ar-BB11SSuo
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher ...

Select

Trump admin sent medical gear to China while he minimized ...
https://dnyuz.com/2020/03/30/trump-admin-sent-medical-gear-to-china-while-he-minimized-the-virus-threat-to-us/
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment.. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 tons (more than 35,000 pounds) of "masks ...

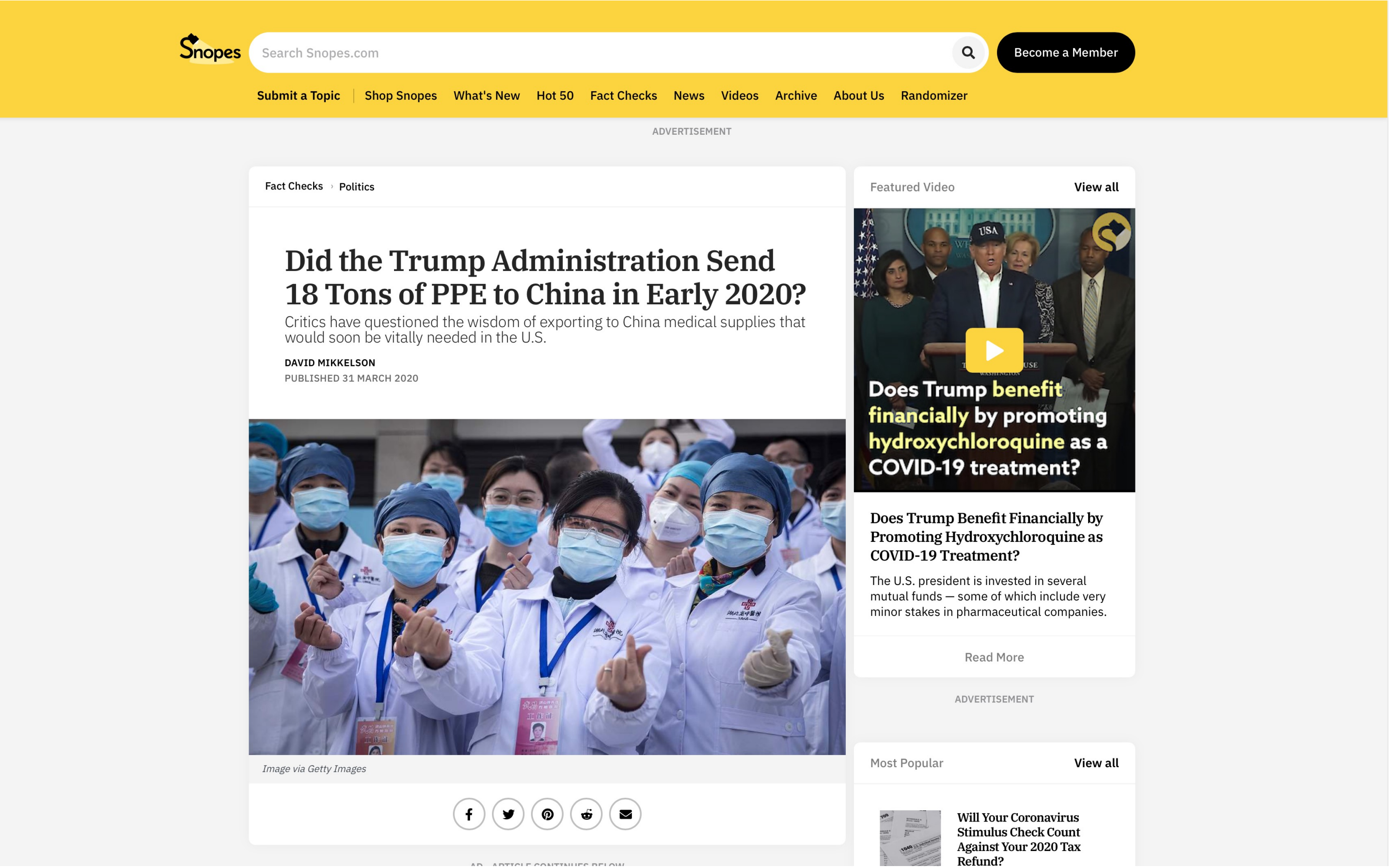
Select

America Has PPE Shortage Because Trump Sent Ours to China ...
https://www.themarysue.com/shortage-of-ppe-because-trump-sent-our-stockpile-to-china/
America Has a Shortage of PPE Because Trump Sent Our Stockpile to China Among all of Trump's poor decisions, this might be the deadliest.

Select

Reports that Trump sent PPE equipment to China Feb 4 ...
https://2peasrefugees.boards.net/thread/104967/reports-trump-sent-equipment-china
There are reports that Trump sent 17.8 tons of PPE equipment to China on February 4. (tweet from Pompeo @secretary Pompeo February 7. There are also reports that Ottawa

Select



Snopes

Search Snopes.com

Become a Member

Submit a Topic | Shop Snopes | What's New | Hot 50 | Fact Checks | News | Videos | Archive | About Us | Randomizer

ADVERTISEMENT

Fact Checks › Politics

Did the Trump Administration Send 18 Tons of PPE to China in Early 2020?

Critics have questioned the wisdom of exporting to China medical supplies that would soon be vitally needed in the U.S.

DAVID MIKKELSON

PUBLISHED 31 MARCH 2020

Image via Getty Images

Featured Video | View all

Does Trump benefit financially by promoting hydroxychloroquine as a COVID-19 treatment?

Does Trump Benefit Financially by Promoting Hydroxychloroquine as COVID-19 Treatment?

The U.S. president is invested in several mutual funds — some of which include very minor stakes in pharmaceutical companies.

Read More

ADVERTISEMENT

Most Popular | View all

Will Your Coronavirus Stimulus Check Count Against Your 2020 Tax Refund?

Claim

The U.S. facilitated the sending of nearly 17.8 tons of donated medical supplies to China to combat the spread of the novel coronavirus in early 2020.

Rating

True

About this rating

Do you rely on Snopes reporting? Become a member today.

Origin

As governments fight the COVID-19 pandemic, Snopes is ***fighting*** *an "infodemic" of rumors and misinformation, and you can help.* ***Read*** *our coronavirus fact checks.* ***Submit*** *any questionable rumors and "advice" you encounter.* ***Become a Founding Member*** *to help us hire more fact-checkers. And, please, follow the* ***CDC*** *or* ***WHO*** *for guidance on protecting your community from the disease.*

During the COVID-19 coronavirus disease pandemic in March 2020, social media users began **sharing** a tweet ostensibly posted by U.S. Rep. Maxine Waters (D-CA) that

Did Bill Gates Admit Vaccinations Are Designed So Governments Can Depopulate the World?

Does Ivanka Trump Hold a Trademark on Coffins in China?

Did Michigan's Governor Ban the Sale of American Flags?

Is Bill Gates Being Sued by India Over Vaccination Deaths?

View all

AD - ARTICLE CONTINUES BELOW

Claim

The U.S. facilitated the sending of nearly 17.8 tons of donated medical supplies to China to combat the spread of the novel coronavirus in early 2020.

Rating

True

About this rating

Do you rely on Snopes reporting? Become a member today.

AD - ARTICLE CONTINUES BELOW

Origin

*As governments fight the COVID-19 pandemic, Snopes is **fighting** an "infodemic" of rumors and misinformation, and you can help. **Read** our coronavirus fact checks. **Submit** any questionable rumors and "advice" you encounter. **Become a Founding Member** to help us hire more fact-checkers. And, please, follow the **CDC** or **WHO** for guidance on protecting your community from the disease.*

During the COVID-19 coronavirus disease pandemic in March 2020, social media users began **sharing** a tweet ostensibly posted by U.S. Rep. Maxine Waters (D-CA) that

Did Bill Gates Admit Vaccinations Are Designed So Governments Can Depopulate the World?

Does Ivanka Trump Hold a Trademark on Coffins in China?

Did Michigan's Governor Ban the Sale of American Flags?

Is Bill Gates Being Sued by India Over Vaccination Deaths?

View all

TASK INSTRUCTIONS

NO RETAKES

Questionna... Questionna... Questionna... Questionna... 5 Docume... 6 Docume... 7 Docume... 8 Docume... 9 Docume... 10 Docume... 11 Docume... 12 Docume... 13 Outco...

Please judge the truthfulness of the following statement

Claimant/Source: Facebook Post

Date: 2020-02-04

Statement: On February 7, the WHO warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.

Please select one of the URLs below to justify the truthfulness value of the statement, then input the truth level as one from the six labels provided and write your justification.

Insert your query

February 7 WHO limited stock of PPE trump

SEARCH

46 results found.

Did the Trump Administration Send 18 Tons of PPE to China ...
https://www.snopes.com/fact-check/us-tons-ppe-china/
Did the Trump Administration Send 18 Tons of PPE to China in Early 2020? ... of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of ...
Select

Trump administration sent protective medical gear to China ...
https://www.msn.com/en-us/health/medical/trump-administration-sent-protective-medical-gear-to-china-while-he-was-minimizing-the-virus-threat-to-us/ar-BB11SSuo
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher ...
Select

Trump admin sent medical gear to China while he minimized ...
https://dnyuz.com/2020/03/30/trump-admin-sent-medical-gear-to-china-while-he-minimized-the-virus-threat-to-us/
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment.. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 tons (more than 35,000 pounds) of "masks ...
Select

America Has PPE Shortage Because Trump Sent Ours to China ...
https://www.themarysue.com/shortage-of-ppe-because-trump-sent-our-stockpile-to-china/
America Has a Shortage of PPE Because Trump Sent Our Stockpile to China Among all of Trump's poor decisions, this might be the deadliest.
Select

Reports that Trump sent PPE equipment to China Feb 4 ...
https://2peasrefugees.boards.net/thread/104967/reports-trump-sent-equipment-china
There are reports that Trump sent 17.8 tons of PPE equipment to China on February 4. (tweet from Pompeo @secretary Pompeo February 7. There are also reports that Ottawa
Select

[Reports that Trump sent PPE equipment to China Feb 4 ...]
https://2peasrefugees.boards.net/thread/104967/reports-trump-sent-equipment-china
There are reports that Trump sent 17.8 tons of PPE equipment to China on February 4. (tweet from Pompeo @secretary Pompeo February 7. There are also reports that Ottawa also sent almost that amount during the same time period. In addition, THEY brought the virus back to the US.

Select

[Sick of the Slant - A Fair Look at the News of the Day ...]
https://www.facebook.com/pg/SickOfTheSlant/posts/
(Albert) "On February 7, the (World Health Organization) warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.

Select

[CNN - Breaking News, Latest News and Videos]
https://lite.cnn.com/en/article/h_c9ec764aa333434b8543e28250957b40
Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 tons (more than 35,000 pounds) of "masks, gowns, gauze ...

Select

[Why did Trump send 18 tons of medical equipment to China...]
https://247sports.com/college/texas-tech/board/102960/Contents/why-did-trump-send-18-tons-of-medical-equipment-to-china-145529702/?page=1
Why did Trump send 18 tons of medical equipment to China...

Select

[Trump Sends 35,000 Pounds of PPE to China | GopherHole ...]
https://www.forums.gopherhole.com/boards/threads/trump-sends-35-000-pounds-of-ppe-to-china.95940/
Trump Sends 35,000 Pounds of PPE to China. Thread starter stocker08; Start date 25 minutes ago; stocker08 Well-known member. Joined Feb 2, 2009 Messages 18,001 Reaction score 261 Points 83. 25 minutes ago #1 Why is this being glossed over? Trump has been a complete fool throughout this entire debacle. He's minimized this every step of the way ...

Select

[US Shipping China Covid 19 Medical Supplies?]
https://www.debatepolitics.com/covid-19-coronavirus-disease/392029-us-shipping-china-covid-19-medical-supplies.html
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 ...

Select

Items per page: 10 1 – 10 of 46

- ○ Lie
- ○ False
- ○ Barely True
- ○ Half True
- ○ Mostly True
- ● True

Your justification here

NEXT (*) you have to fill each field to proceed

[Reports that Trump sent PPE equipment to China Feb 4 ...]
https://2peasrefugees.boards.net/thread/104967/reports-trump-sent-equipment-china
There are reports that Trump sent 17.8 tons of PPE equipment to China on February 4. (tweet from Pompeo @secretary Pompeo February 7. There are also reports that Ottawa also sent almost that amount during the same time period. In addition, THEY brought the virus back to the US.
Select

[Sick of the Slant - A Fair Look at the News of the Day ...]
https://www.facebook.com/pg/SickOfTheSlant/posts/
(Albert) "On February 7, the (World Health Organization) warned about the limited stock of PPE. That same day, the Trump administration announced it was sending 18 tons of masks, gowns and respirators to China.
Select

[CNN - Breaking News, Latest News and Videos]
https://lite.cnn.com/en/article/h_c9ec764aa333434b8543e28250957b40
Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 tons (more than 35,000 pounds) of "masks, gowns, gauze ...
Select

[Why did Trump send 18 tons of medical equipment to China...]
https://247sports.com/college/texas-tech/board/102960/Contents/why-did-trump-send-18-tons-of-medical-equipment-to-china-145529702/?page=1
Why did Trump send 18 tons of medical equipment to China...
Select

[Trump Sends 35,000 Pounds of PPE to China | GopherHole ...]
https://www.forums.gopherhole.com/boards/threads/trump-sends-35-000-pounds-of-ppe-to-china.95940/
Trump Sends 35,000 Pounds of PPE to China. Thread starter stocker08; Start date 25 minutes ago; stocker08 Well-known member. Joined Feb 2, 2009 Messages 18,001 Reaction score 261 Points 83. 25 minutes ago #1 Why is this being glossed over? Trump has been a complete fool throughout this entire debacle. He's minimized this every step of the way ...
Select

[US Shipping China Covid 19 Medical Supplies?]
https://www.debatepolitics.com/covid-19-coronavirus-disease/392029-us-shipping-china-covid-19-medical-supplies.html
Concerns about a dwindling supply of PPE are not new. Back on February 7, the World Health Organization sounded alarm bells about "the limited stock of PPE," noting demand was 100 times higher than normal for this equipment. Yet the same day as the WHO warning, the Trump administration announced that it was transporting to China nearly 17.8 ...
Select

Items per page: 10 1 – 10 of 46

- ○ Lie
- ○ False
- ○ Barely True
- ○ Half True
- ○ Mostly True
- ● True

Your justification here
The U.S. facilitated the sending of nearly 17.8 tons of donated medical supplies to China to combat the spread of the novel coronavirus in early 2020.

NEXT (*) you have to fill each field to proceed

TASK INSTRUCTIONS

NO RETAKES

Questionna... | Questionna... | Questionna... | Questionna... | Docume... | 6 Docume... | 7 Docume... | 8 Docume... | 9 Docume... | 10 Docume... | 11 Docume... | 12 Docume... | 13 Outco...

Please judge the truthfulness of the following statement

Claimant/Source: Abraham Lincoln

Date: 2020-04-01

Statement: If I were not a Formula One pilot, I would be a football player.

Please select one of the URLs below to justify the truthfulness value of the statement, then input the truth level as one from the six labels provided and write your justification.

Insert your query

Abraham Lincoln

SEARCH

49 results found.

[Abraham Lincoln - Wikipedia]
https://en.wikipedia.org/wiki/Abraham_Lincoln
Abraham Lincoln was born on February 12, 1809, the second child of Thomas Lincoln and Nancy Hanks Lincoln, in a one-room log cabin on Sinking Spring Farm near Hodgenville, Kentucky.: 20–22 He was a descendant of Samuel Lincoln, an Englishman who migrated from Hingham, Norfolk, to its namesake, Hingham, Massachusetts, in 1638.
Select

[Abraham Lincoln - Facts, Birthday & Assassination - HISTORY]
https://www.history.com/topics/us-presidents/abraham-lincoln
Abraham Lincoln, a self-taught lawyer, legislator and vocal opponent of slavery, was elected 16th president of the United States in November 1860, shortly before the outbreak of the Civil War. He ...
Select

[Abraham Lincoln | Biography, Facts, History, & Childhood ...]
https://www.britannica.com/biography/Abraham-Lincoln
Abraham Lincoln was a member of the Whig Party and later a Republican.He believed that the government's job was to do what a community of people could not do for themselves. One of his greatest preoccupations as a political thinker was the issue of self-governance and the promise and problems that could arise from it.
Select

[Abraham Lincoln | The White House]
https://www.whitehouse.gov/about-the-white-house/presidents/abraham-lincoln/
Abraham Lincoln became the United States' 16th President in 1861, issuing the Emancipation Proclamation that declared forever free those slaves within the Confederacy in 1863. Lincoln warned the ...
Select

[Abraham Lincoln Facts for Kids]
https://kids.kiddle.co/Abraham_Lincoln
Abraham Lincoln (February 12 1809 – April 15 1865) was the 16th President of the United States.He served as president from 1861 to 1865, during the American Civil War.Just five days after most of the Confederate forces had surrendered and the war was ending, John Wilkes Booth assassinated Lincoln.. Lincoln was the first president of the United States to be assassinated.
Select

[Abraham Lincoln Facts for Kids](https://kids.kiddle.co/Abraham_Lincoln)
https://kids.kiddle.co/Abraham_Lincoln
Abraham Lincoln (February 12 1809 – April 15 1865) was the 16th President of the United States.He served as president from 1861 to 1865, during the American Civil War.Just five days after most of the Confederate forces had surrendered and the war was ending, John Wilkes Booth assassinated Lincoln.. Lincoln was the first president of the United States to be assassinated.
Select

TOP 25 QUOTES BY ABRAHAM LINCOLN (of 1141) | A-Z Quotes
https://www.azquotes.com/author/8880-Abraham_Lincoln
"Abraham Lincoln: Speeches & Writings Part 2: 1859-1865: Library of America #46", p.627, Library of America 1 Copy quote. I do the very best I can, I mean to keep going. If the end brings me out all right, then what is said against me won't matter. If I'm wrong, ten angels swearing I was right won't make a difference.
Select

Biography of Abraham Lincoln, 16th U.S. President
https://www.thoughtco.com/abraham-lincoln-16th-president-united-states-104273
Abraham Lincoln (February 12, 1809–April 15, 1865) was the 16th president of the United States, serving from 1861 to 1865. During his time in office, the nation fought the Civil War, which claimed hundreds of thousands of lives. One of Lincoln's greatest accomplishments was the abolition of slavery in 1864.
Select

50 interesting facts about Abraham Lincoln's life ...
https://constitutioncenter.org/blog/50-shades-of-abraham-lincoln-2/
Abraham Lincoln, the 16th president, has a birthday today. If you are a Lincoln fan, here are some cool facts, including Lincoln's career as an inventor, his love of animals, and his one losing appearance before the Supreme Court.
Select

Abraham Lincoln's Assassination - HISTORY
https://www.history.com/topics/american-civil-war/abraham-lincoln-assassination
Abraham Lincoln. Abraham Lincoln, a self-taught lawyer, legislator and vocal opponent of slavery, was elected 16th president of the United States in November 1860, shortly before the outbreak of ...
Select

Abraham Lincoln - Simple English Wikipedia, the free ...
https://simple.wikipedia.org/wiki/Abraham_Lincoln
Abraham Lincoln (February 12, 1809 – April 15, 1865) was an American politician. He was the 16th President of the United States.He was president from 1861 to 1865, during the American Civil War.Just five days after most of the Confederate forces had surrendered and the war was ending, John Wilkes Booth assassinated Lincoln for being a tyrant.Lincoln was the first president of the United ...
Select

Items per page: 10 1 – 10 of 49

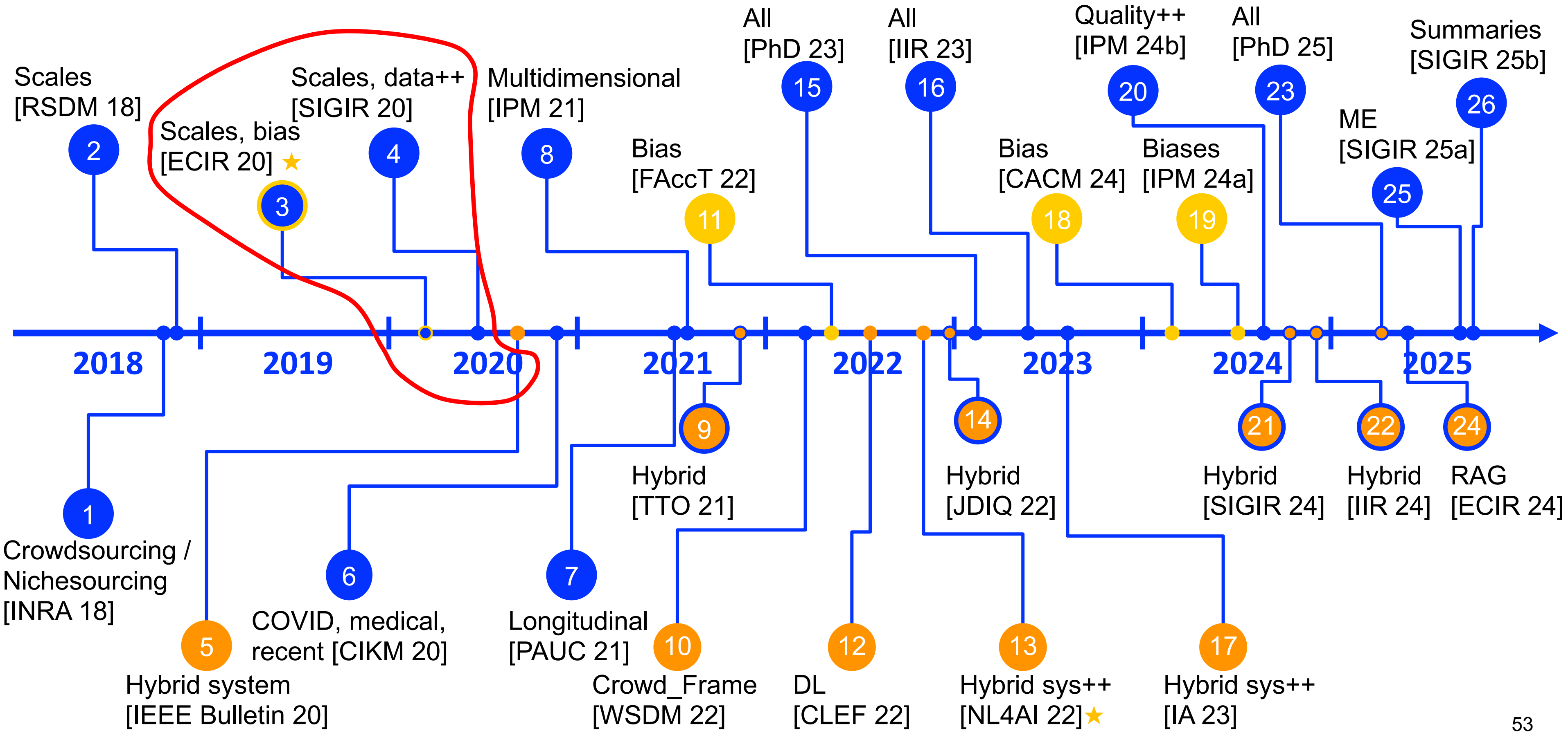
- Lie
- False
- Barely True
- Half True
- Mostly True
- True

Your justification here
Abraham Lincoln was born on February 12, 1809,

BACK NEXT (*) you have to fill each field to proceed

Now, does it work?

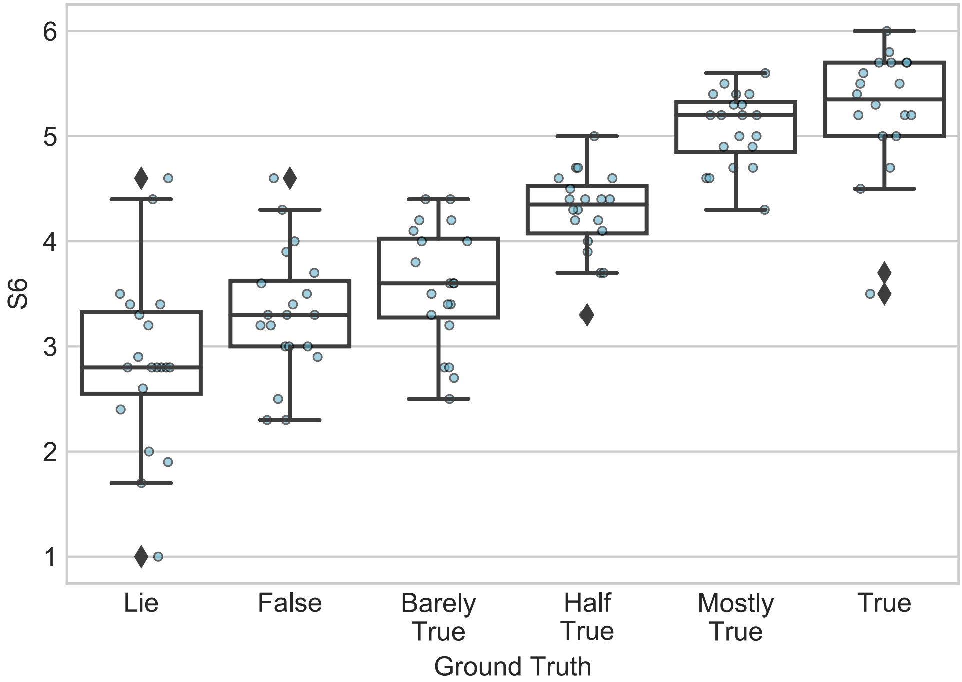
Timeline



2018
2019
2020
2021
2022
2023
2024
2025
1
Crowdsourcing / Nichesourcing [INRA 18]
2
Scales [RSDM 18]
3
Scales, bias [ECIR 20] ★
4
Scales, data++ [SIGIR 20]
5
Hybrid system [IEEE Bulletin 20]
6
COVID, medical, recent [CIKM 20]
7
Longitudinal [PAUC 21]
8
Multidimensional [IPM 21]
9
Hybrid [TTO 21]
10
Crowd_Frame [WSDM 22]
11
Bias [FAccT 22]
12
DL [CLEF 22]
13
Hybrid sys++ [NL4AI 22]★
14
Hybrid [JDIQ 22]
15
All [PhD 23]
16
All [IIR 23]
17
Hybrid sys++ [IA 23]
18
Bias [CACM 24]
19
Biases [IPM 24a]
20
Quality++ [IPM 24b]
21
Hybrid [SIGIR 24]
22
Hybrid [IIR 24]
23
All [PhD 25]
24
RAG [ECIR 24]
25
ME [SIGIR 25a]
26
Summaries [SIGIR 25b]

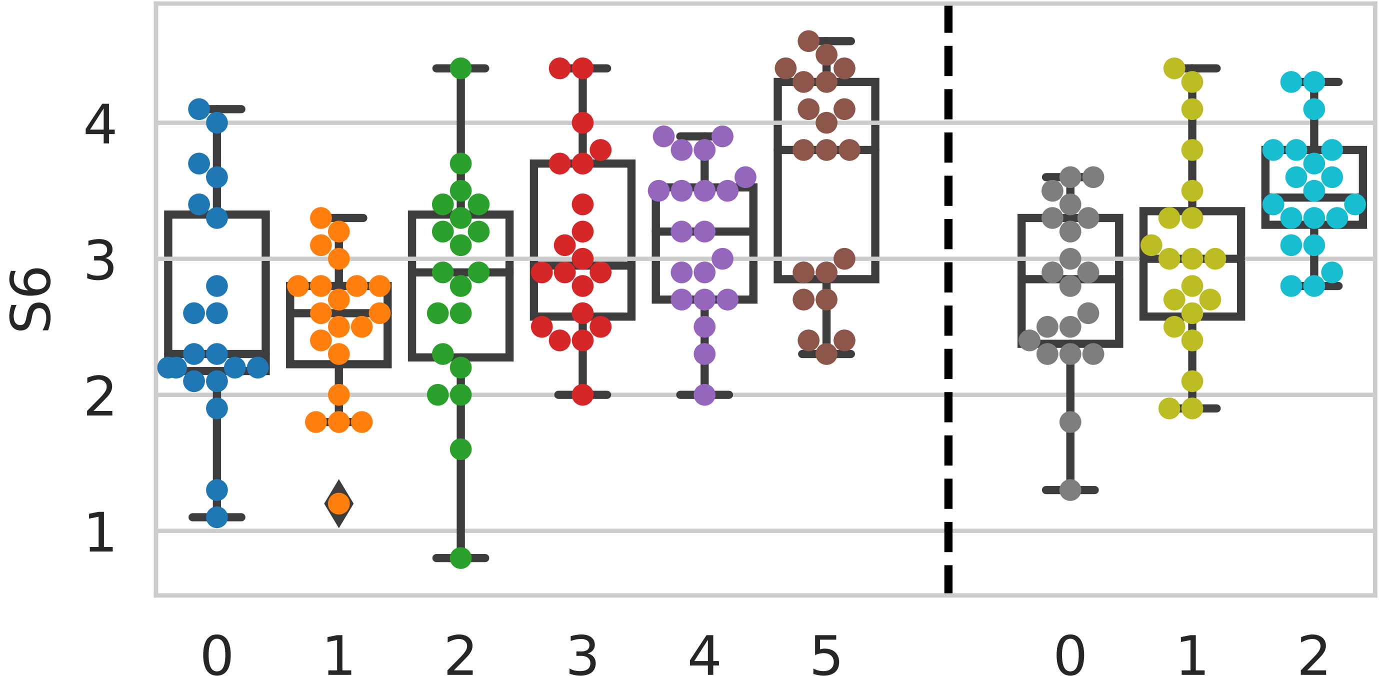
Agreement with Ground truth 3. [ECIR 2020]

Crowd



(more on the scale later)

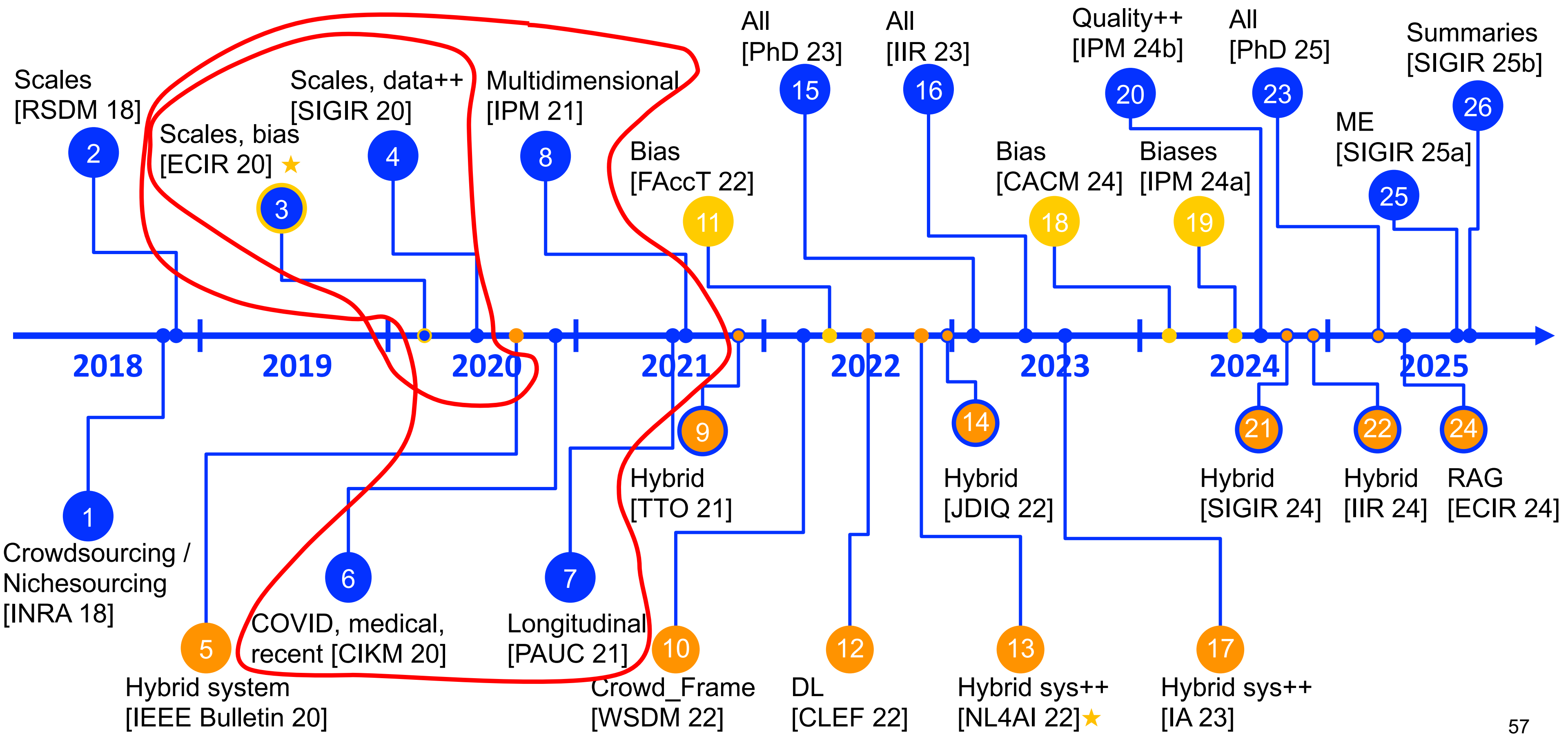
Another result 4. [SIGIR 2020]



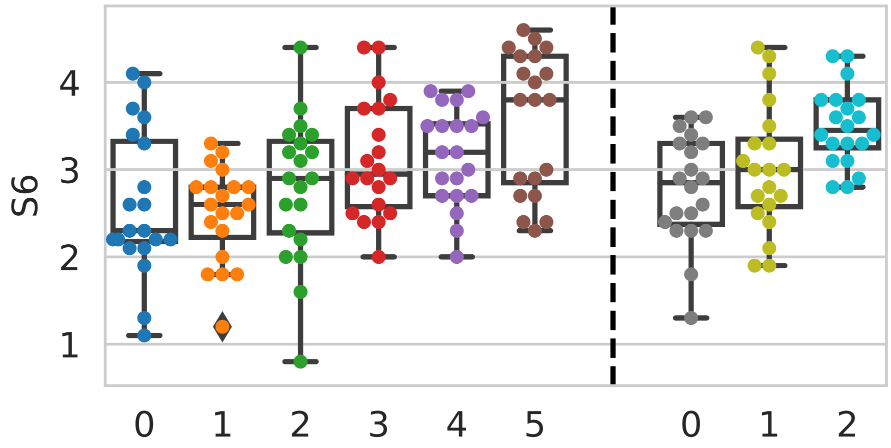
What we told ourselves

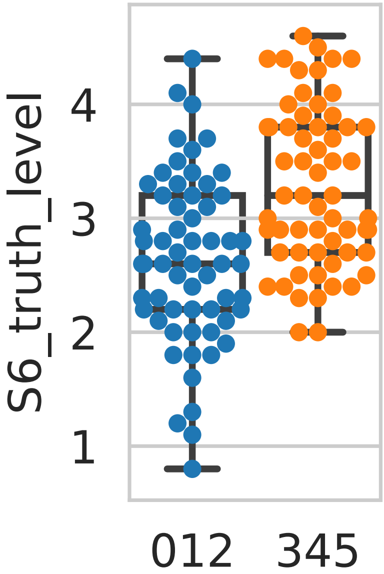
- "Increasing medians"… bla bla bla … "there is a signal" … bla bla bla …
- Yes we published some papers
- And some more in the following years
  - Multidimensional scale
  - COVID-related statements (medical, recent, sensitive)
  - Longitudinal
  - …

Timeline: we're here



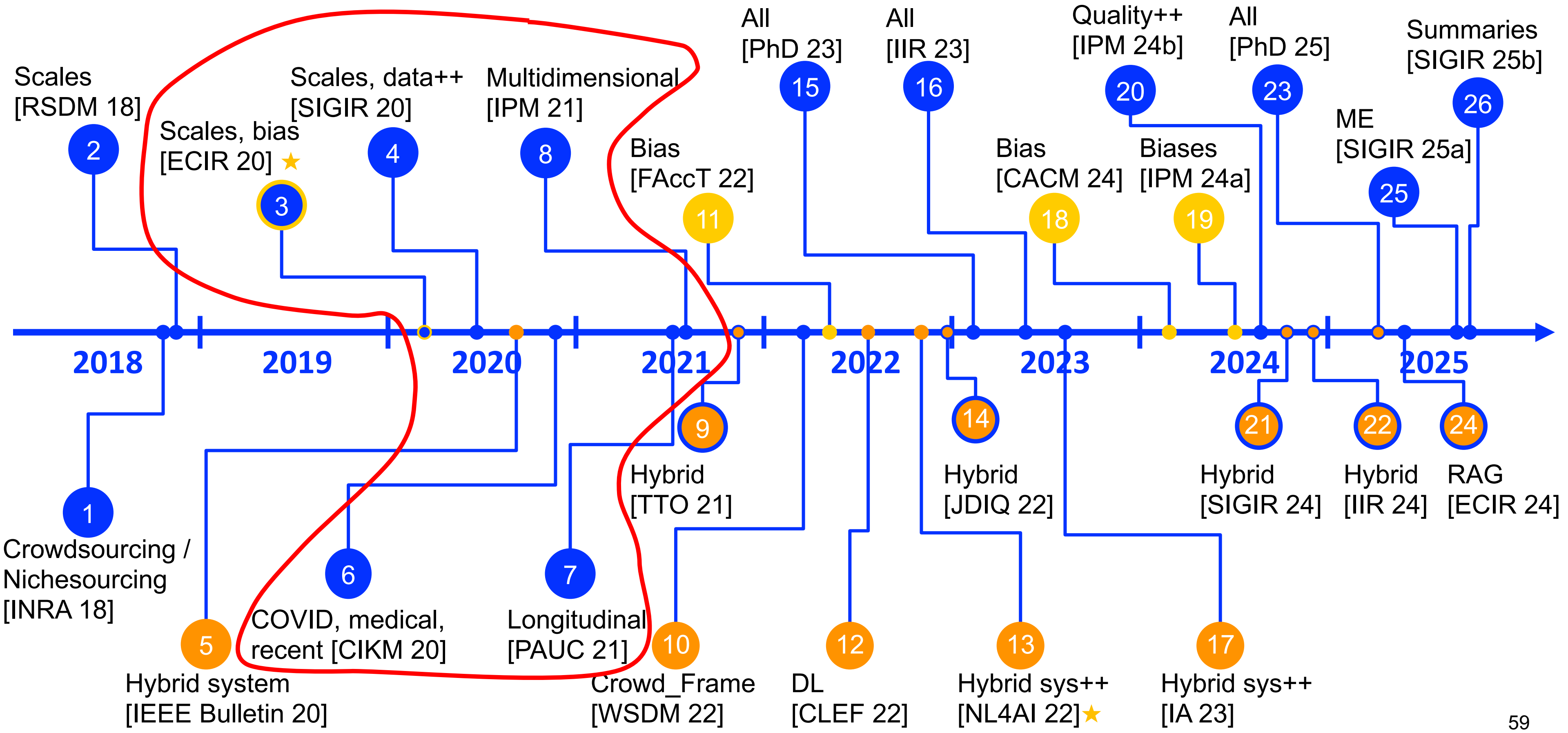
The sad truth





Binary accuracy: 0.6

We. Did. Not. Stop.



Motivation: Does crowdsourcing actually work?

Study	Binary Accuracy
3. [ECIR 2020]	0.841 (*)
4. [SIGIR 2020]	0.627
8. [IPM 2021]	0.571
11. [FAccT 2022]	0.580

(*) But the Politifact evidence was easily retrievable…

Motivation: Does crowdsourcing actually work?

Study	Binary Accuracy
3. [ECIR 2020]	**0.841 (*)**
4. [SIGIR 2020]	**0.627**
8. [IPM 2021]	**0.571**
11. [FAccT 2022]	**0.580**
Allen et al. (Science Advances 2021)	0.826

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances, 7*(36), eabf4393. https://doi.org/10.1126/sciadv.abf4393

(*) But the Politifact evidence was easily retrievable…

Motivation: Does crowdsourcing actually work?

Study	Binary Accuracy	Method
3. [ECIR 2020]	**0.841 (*)**	Crowd
4. [SIGIR 2020]	**0.627**	
8. [IPM 2021]	**0.571**	
11. [FAccT 2022]	**0.580**	
Allen et al. (Science Advances 2021)	0.826	
Hu et al. (AI Open 2022)	[0.531 - 0.904]	AI

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances, 7*(36), eabf4393. https://doi.org/10.1126/sciadv.abf4393

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI Open, 3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

(*) But the Politifact evidence was easily retrievable…

Real motivation!

Study	Binary Accuracy	Method
3. [ECIR 2020]	**0.841 (*)**	Crowd
4. [SIGIR 2020]	**0.627**	
8. [IPM 2021]	**0.571**	
11. [FAccT 2022]	**0.580**	
Allen et al. (Science Advances 2021)	0.826	
Hu et al. (AI Open 2022)	[0.531 - 0.904]	AI

We're at 0.6, the others (even AI) are at 0.8, WTF?

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances, 7*(36), eabf4393. https://doi.org/10.1126/sciadv.abf4393

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI Open, 3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

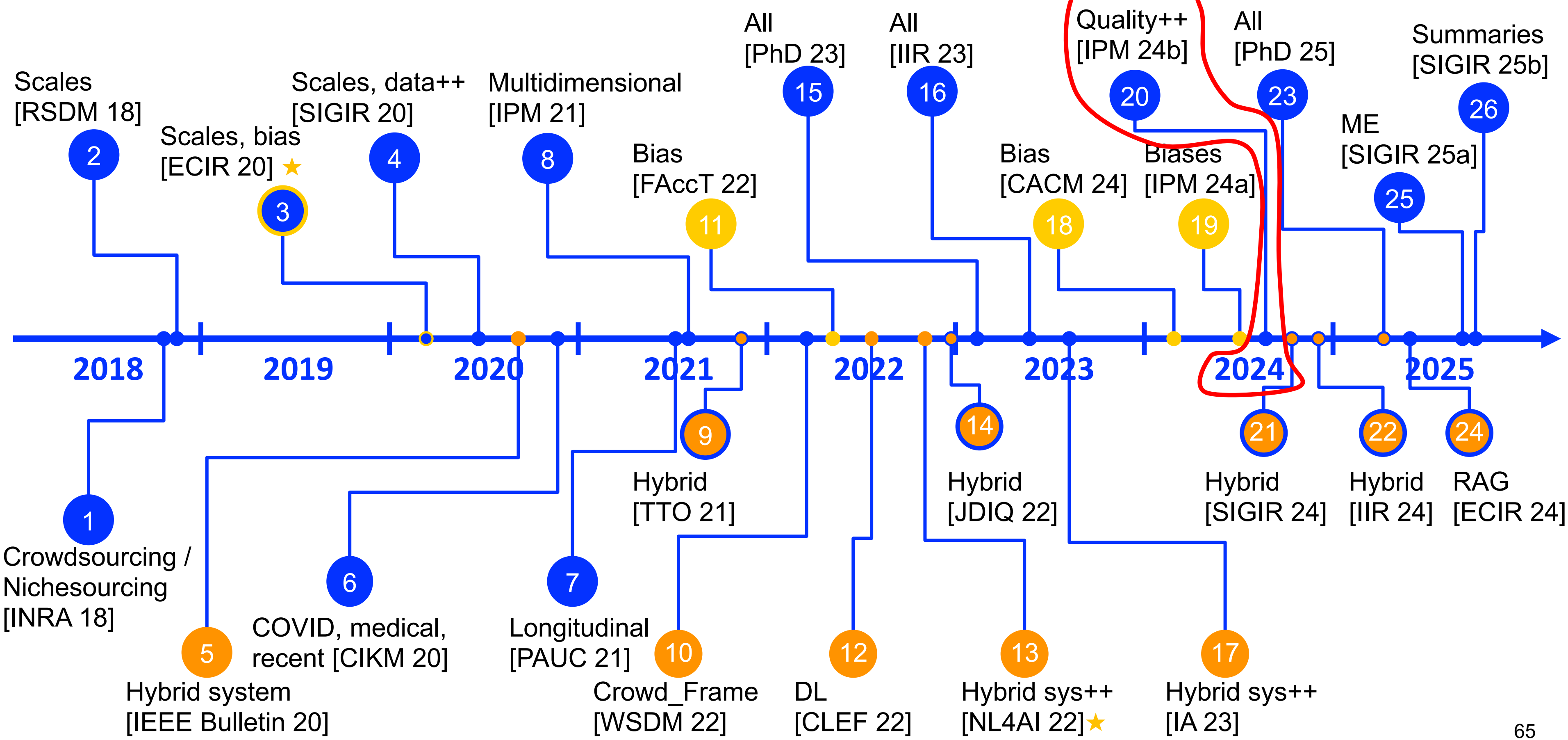
(*) But the Politifact evidence was easily retrievable…

Improvements

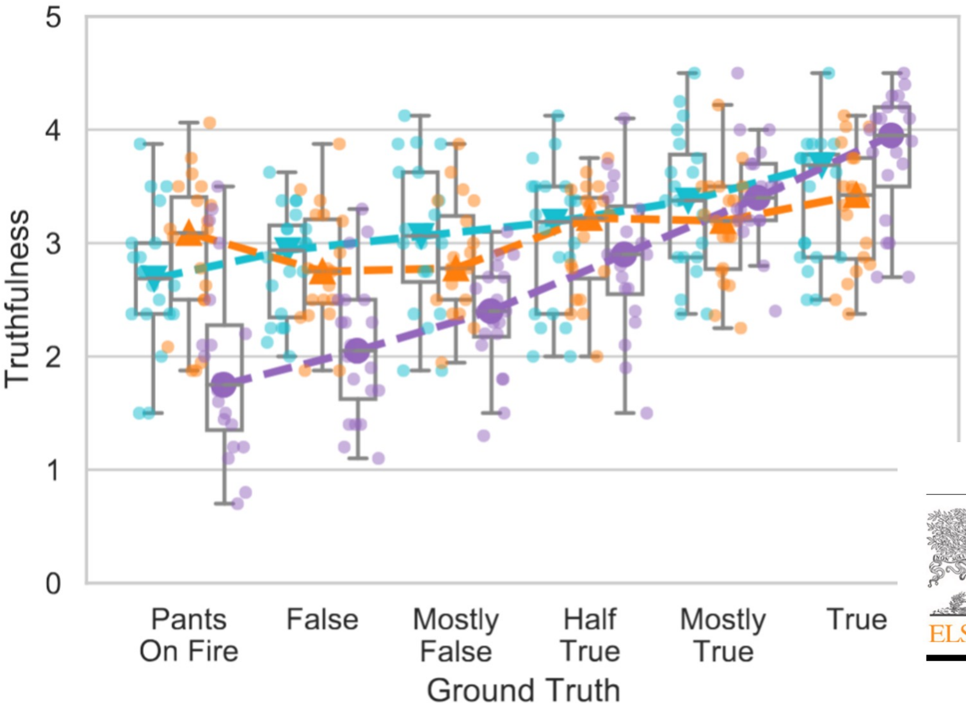
- Better crowd
  - Amazon's Mechanical Turk → Prolific
- Better instructions
  - Rewording, shortening, more direct
- Better UI
  - Some small improvements

- (even after 5+ years of re-doing the same experiments!)

Timeline: we are here...



(Better) Results!



Information Processing and Management 61 (2024) 103792

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

ELSEVIER

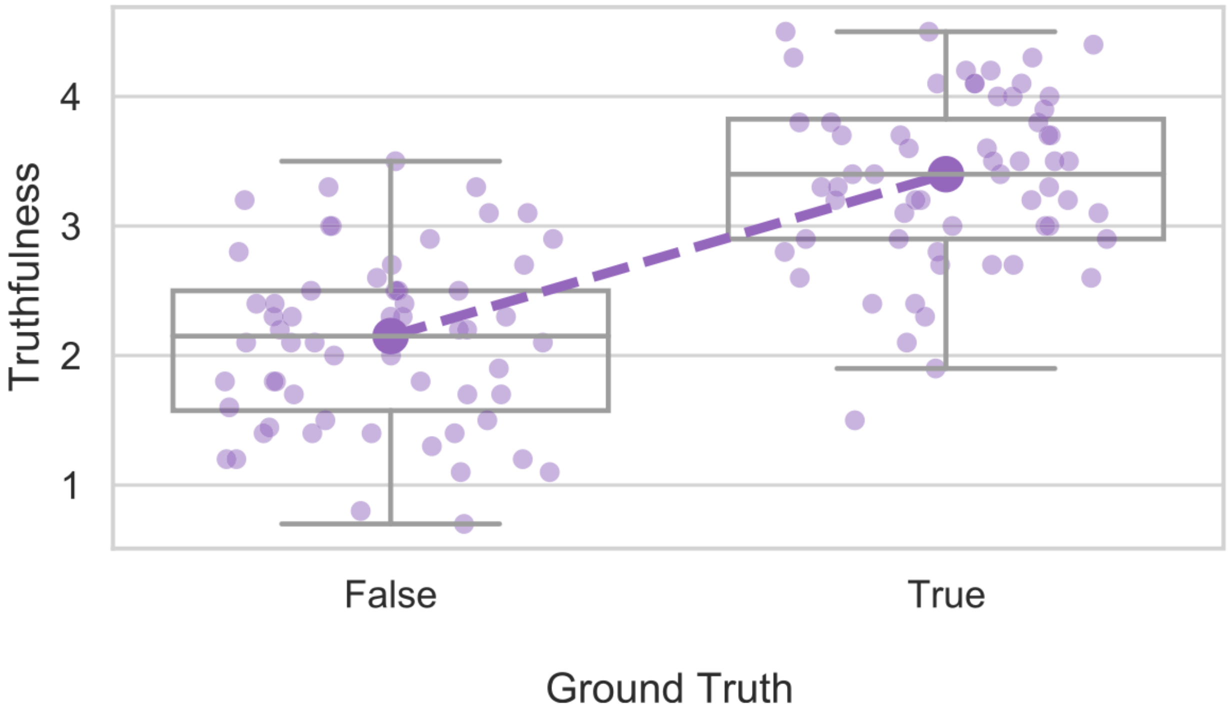
Check for updates

Crowdsourced Fact-checking: Does It Actually Work?

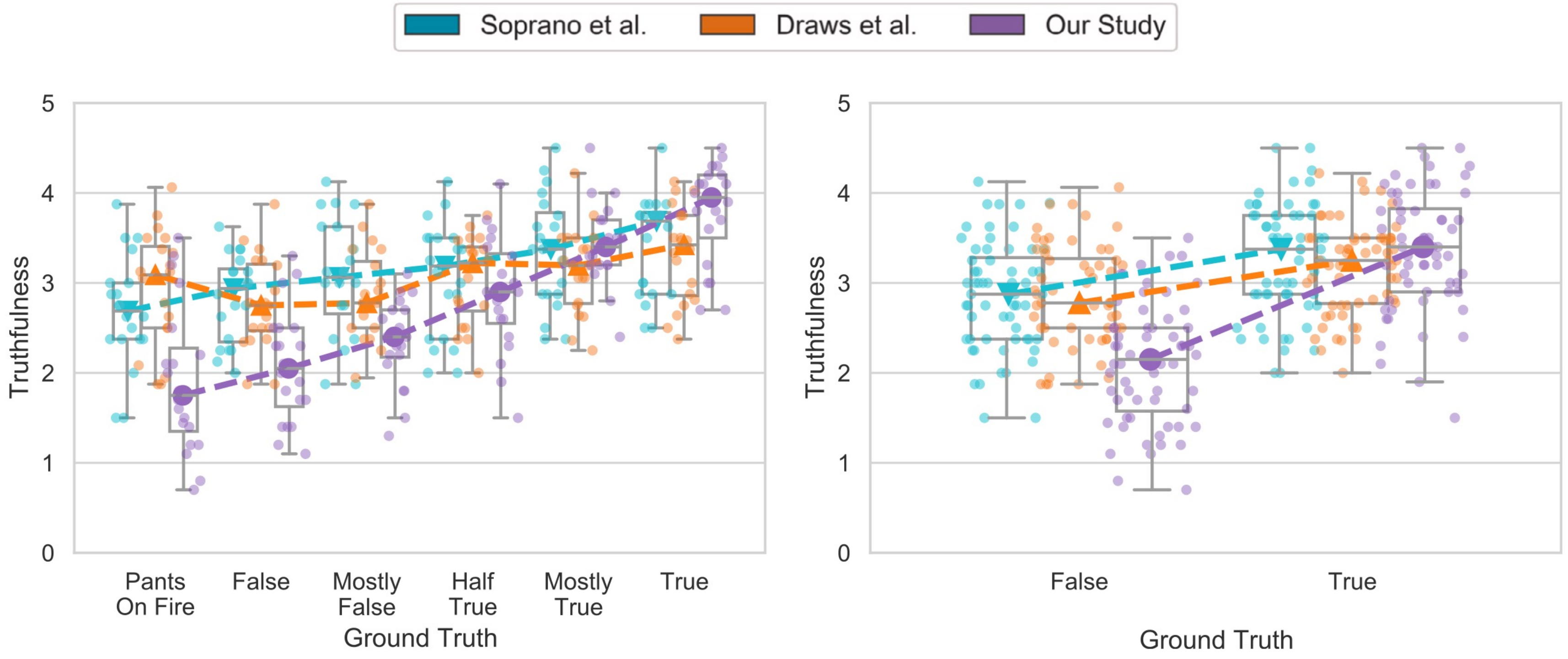
David La Barbera [a,*], Eddy Maddalena [a], Michael Soprano [a], Kevin Roitero [a], Gianluca Demartini [b], Davide Ceolin [c], Damiano Spina [d], Stefano Mizzaro [a]

[a] *University Of Udine, Via Delle Scienze 206, Udine, Italy*
[b] *The University of Queensland, St Lucia Queensland 4072, Brisbane, Australia*
[c] *Centrum Wiskunde & Informatica (CWI), Science Park 123, Amsterdam, The Netherlands*
[d] *RMIT University, 124 La Trobe St, Melbourne, Australia*

Aaaaaah, 0.8

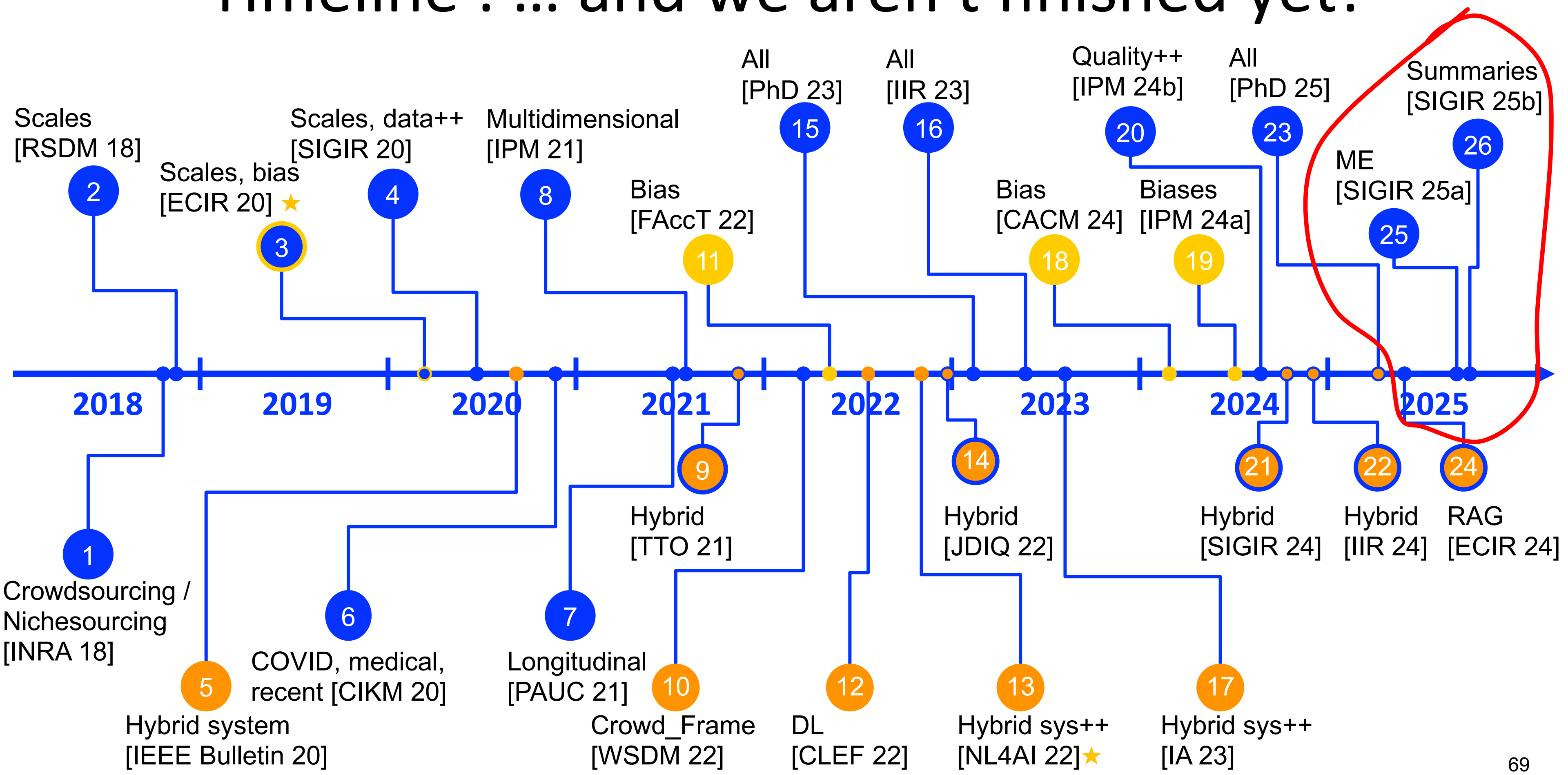


Aaaaaah, 0.8



Soprano et al.
Draws et al.
Our Study
5
4
3
2
1
0
Truthfulness
Pants On Fire
False
Mostly False
Half True
Mostly True
True
Ground Truth
5
4
3
2
1
0
Truthfulness
False
True
Ground Truth

Timeline : … and we aren't finished yet!



Outline

- Intro
- Crowdsourcing for fact-checking
- Truthfulness Scales and Magnitude estimation

How to express truthfulness?

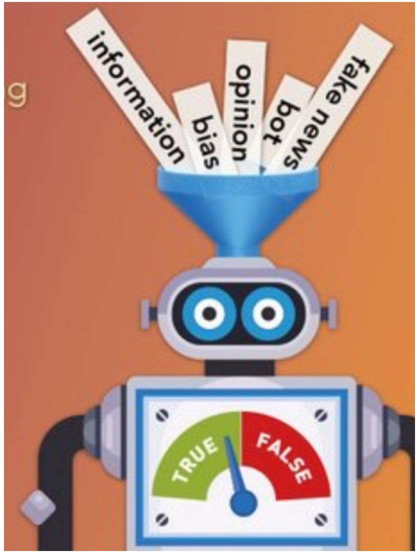
- In technical terms, which **scale** to use?

Binary scale

- Traditional solution: binary scale
- **Binary truthfulness**

True

False



Non-binary scales

- 3-levels scale
  - True
  - Neither false not true
  - False

Non-binary scales

- 4-levels scale
  - Completely false
  - More false than true
  - More true than false
  - Completely true

It's not just categories

- 4-level ordinal category scale
  - Completely false
  - More false than true
  - More true than false
  - Completely true

It's not just ranking

- 4-level ordinal category scale
  - Completely false
  - More false than true
  - More true than false
  - Completely true

Also IRL

- Politifact: https://www.politifact.com
- 6 levels scale

POLITIFACT

POLITIFACT
The Poynter Institute

Subtle differences

2 (n) levels	3 (n) ordered levels
Binary (n-ary) classification	Ordinal classification
All errors are equal	Misclassifying in the adjacent category is a smaller error than misclassifying in the next one - False - Neither true nor false - True
Standard metrics (F1, Accuracy)	No standard metrics

No standard metrics?!

- Accuracy
  - Assumes all errors are equal!
- MSE, RMSE, MAE, …?
  - Assume equally spaced categories!
- CEM!?
  - (but that's for another talk)

An Effectiveness Metric for Ordinal Classification: Formal Properties and Experimental Results

Enrique Amigó
UNED
Madrid, Spain
enrique@lsi.uned.es

Julio Gonzalo
UNED
Madrid, Spain
julio@lsi.uned.es

Stefano Mizzaro
University of Udine
Udine, Italy
mizzaro@uniud.it

Jorge Carrillo-de-Albornoz
UNED
Madrid, Spain
jcalbornoz@lsi.uned.es

0, 1, 2, 3, 4, 5?

0	**1**	**2**	**3**	**4**	**5**
0	**1**	**3**	**5**	**7**	**8**
0	**1**	**10**	**100**	**1000**	**10000**

- (more later on this)

We even tried 100 levels

- 101 actually

Can The Crowd Identify Misinformation Objectively? The Effects of Judgment Scale and Assessor’s Background

Kevin Roitero
roitero.kevin@spes.uniud.it
University of Udine
Udine, Italy

Michael Soprano
michael.soprano@outlook.com
University of Udine
Udine, Italy

Shaoyang Fan
fsysean@gmail.com
The University of Queensland
Brisbane, Australia

Damiano Spina
damiano.spina@rmit.edu.au
RMIT University
Melbourne, Australia

Stefano Mizzaro
mizzaro@uniud.it
University of Udine
Udine, Italy

Gianluca Demartini
demartini@acm.org
The University of Queensland
Brisbane, Australia

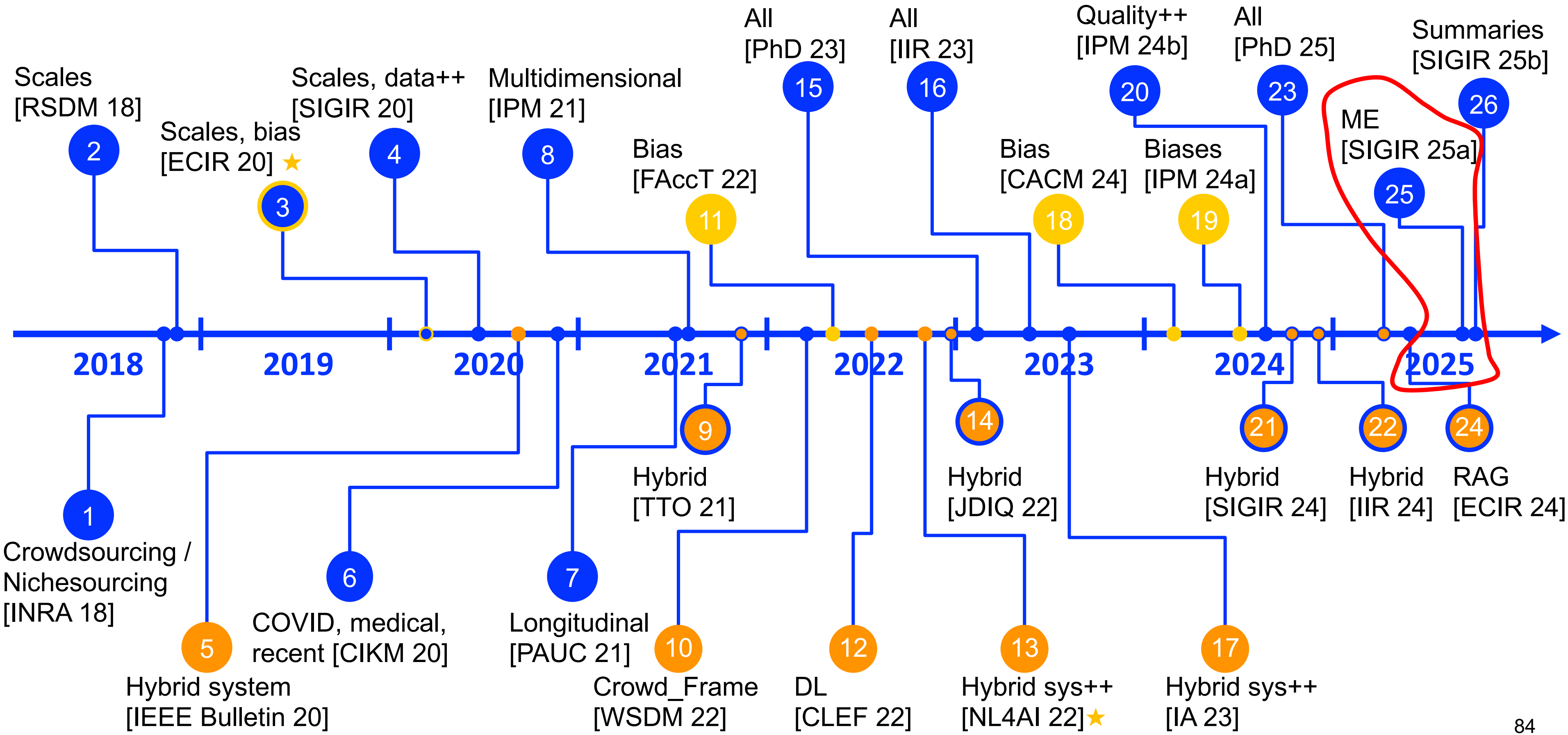
Even more?!

- So,
  - 2
  - 3
  - 4
  - 5
  - 6
  - 101
- Even more!?

Yes. More than 101.

Timeline

Forthcoming SIGIR 2025 paper

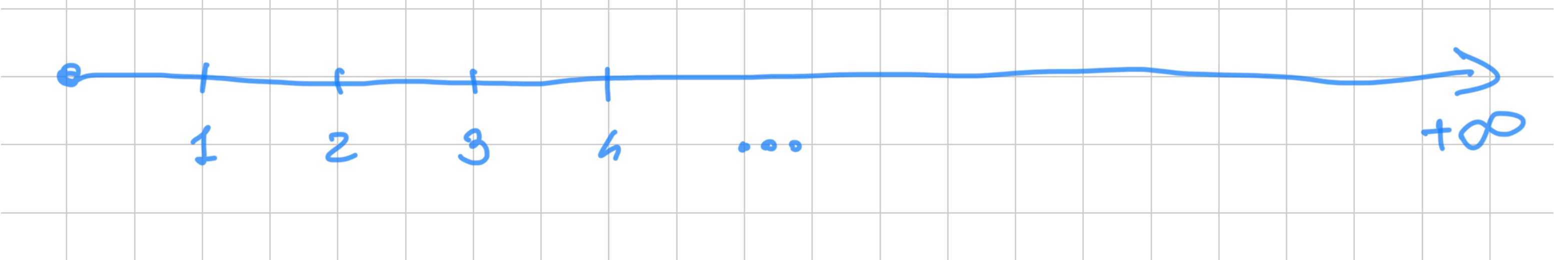


Magnitude Estimation (ME)

- A psychophysical scaling technique for measuring perception
- Stimuli at different levels of intensity are presented to an observer
- The intensity of each stimulus is rated by the assignment of a number, depending on the perceived intensity
  - "Given a stimulus, assign it a number (whole o fraction)"
- Developed by Stanley Stevens at Harvard in 1950s

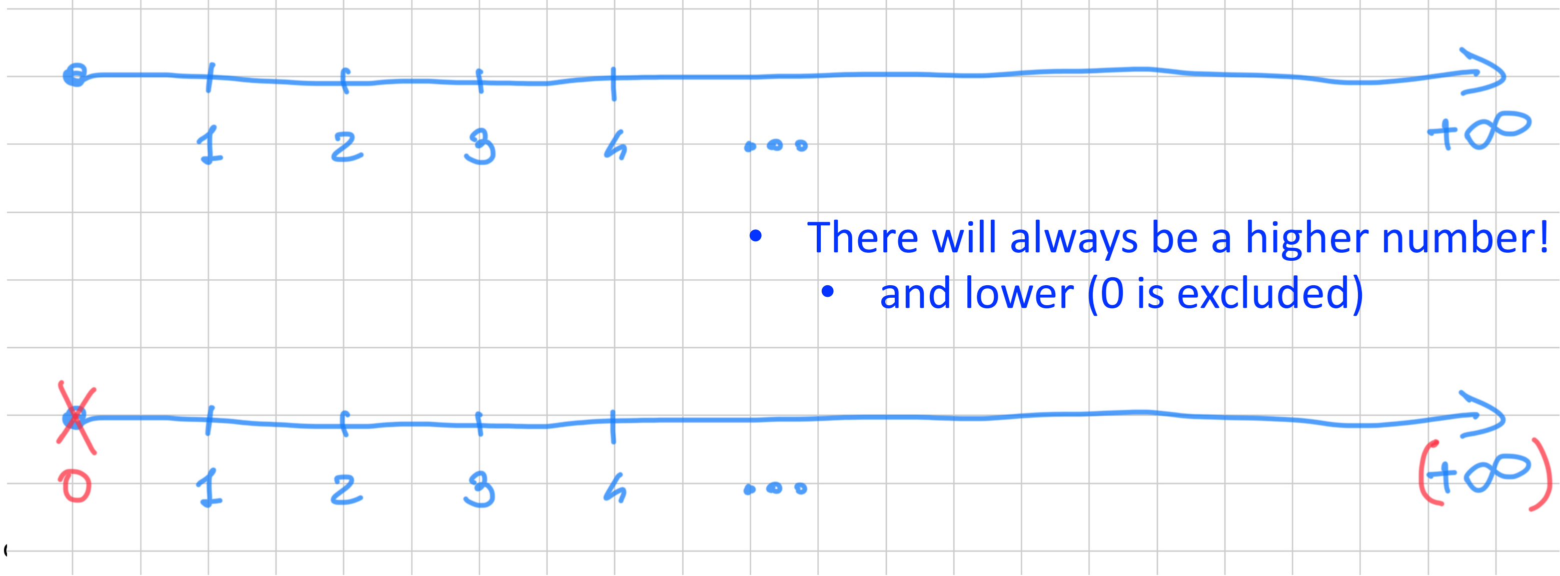
Magnitude Estimation (ME)…

- Unlimited (≠ "category scale with many categories")
  - either]-∞, +∞[or
  -]0, +∞[(we used this one)



Magnitude Estimation (ME)…

- Unlimited (≠ “category scale with many categories”)
  - either]-∞, +∞[or
  -]0, +∞[(we used this one)



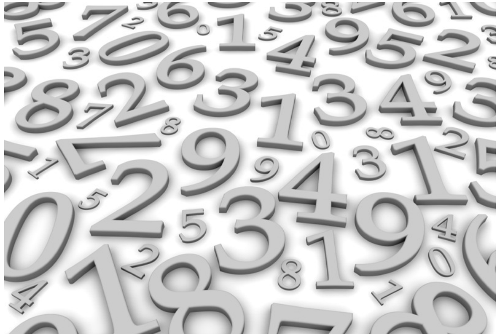
- There will always be a higher number!
  - and lower (0 is excluded)

ME used for

- Initially for physical stimuli, such as:
  - brightness of a light
  - frequency of a sound
  - etc.
- These have an underlying measurable quantity
- Also applied to stimuli that are not physically quantifiable, such as:
  - levels of pain or emotional stress (medicine)
  - severity of crimes, appropriateness of punishments (law)
  - importance of Swedish monarchs (sociology)
  - wine tasting
  - usability of an interface (HCI)
  - judging grammaticality of sentences (linguistics)
  - relevance assessments

ME

- Leads to ratio scale: **ratios** of the assigned numbers is what's important
  - If one item is assigned a 50, and another 10, then it can be inferred that the former is 5 times more … than the latter
  - Supports all mathematical operations, and parametric statistical analysis
- ME truthfulness assessment should be more precise than binary or ordinal assessments
  - The granularity of the scale is chosen by the assessor, and not constrained by pre-determined levels
  - Assessors cannot run out of categories



Issues with ME

- My “inner scale” is different from yours
- Cultural background might affect which numbers are used
  - E.g., school marks over 0–10 (in Italy) vs. A–F vs. 0–100 vs. …
  - Round number tendency (prefer 20 to 21)
- Using ME for truthfulness assessment?
  - "A claim more true (false) than an already absolutely true (false) one?!"
- It. Can’t. Work.
  - But we tried! (and we are the first!)

RQs

- Can the crowd express reliable truthfulness assessments using ME?
- Are there any advantages to do so?

Comparison with S6

- Experimental design identical to a previous study
  - 20. [IPM 2024b]
  - Apart from the scale: we had used Politifact 6-level scale (S6 from now on)
- We knew it worked 😬
  - (actually, many studies)
- We can compare ME with S6 🧐

Information Processing and Management 61 (2024) 103792

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

ELSEVIER

Check for updates

Crowdsourced Fact-checking: Does It Actually Work?

David La Barbera [a,*], Eddy Maddalena [a], Michael Soprano [a], Kevin Roitero [a], Gianluca Demartini [b], Davide Ceolin [c], Damiano Spina [d], Stefano Mizzaro [a]

[a] University Of Udine, Via Delle Scienze 206, Udine, Italy
[b] The University of Queensland, St Lucia Queensland 4072, Brisbane, Australia
[c] Centrum Wiskunde & Informatica (CWI), Science Park 123, Amsterdam, The Netherlands
[d] RMIT University, 124 La Trobe St, Melbourne, Australia

Normalization

- S6 values were converted into [0, 5]
  - (yes, more on that later)
- Each ME value is normalized into [0, 5]
  - Using the max & min values by each worker

$$\alpha_{\text{norm}} = \frac{\alpha - \min(A_w)}{\max(A_w) - \min(A_w)} \cdot 5$$

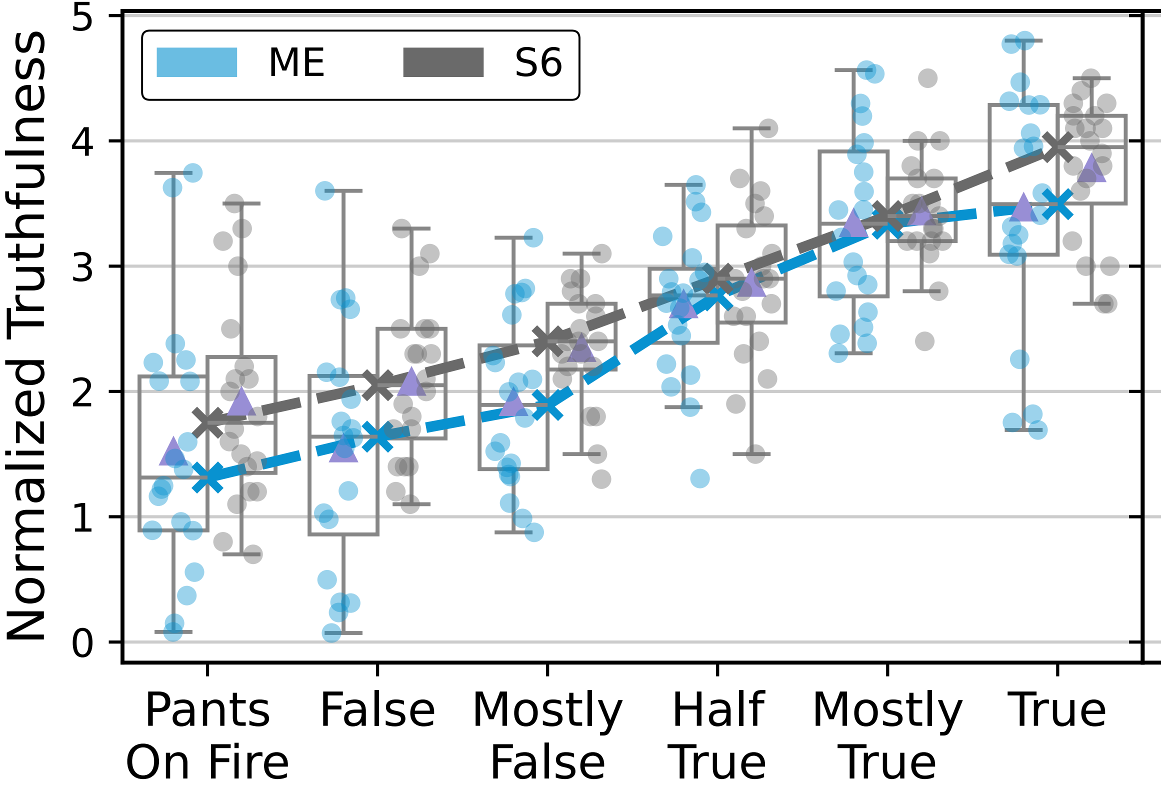
Aggregation

- 10 individual scores → 1 aggregated score
- We tried several alternatives
  - Weighted mean resulted more effective

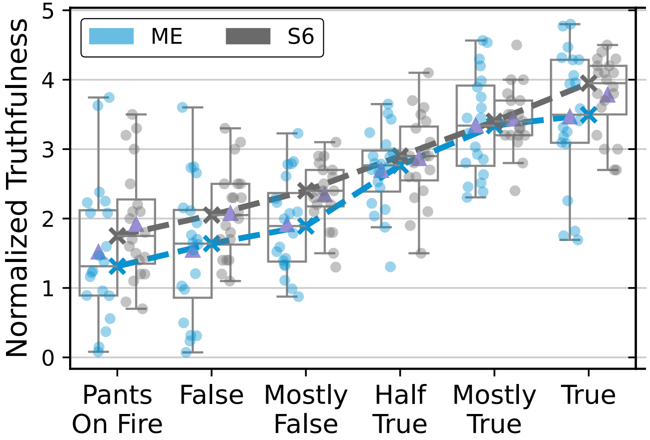
Table 2: Comparison of aggregation functions across effectiveness measures.

	$Accuracy_2$	$Accuracy_3$	$Accuracy_6$	External	$Pairwise_2$	$Pairwise_3$	$Pairwise_6$	MAE	MSE
wmean	**0.80**	**0.60**	**0.32**	0.63	**0.89**	**0.83**	**0.79**	**0.97**	**1.59**
median	0.73	0.59	0.28	**0.67**	0.87	0.82	0.78	1.02	1.73
mean	0.75	0.57	0.29	0.53	0.86	0.82	0.77	1.07	1.78
gmean	0.57	0.41	0.21	-0.15	0.26	0.25	0.24	2.10	7.06

ME vs. S6 & ground truth

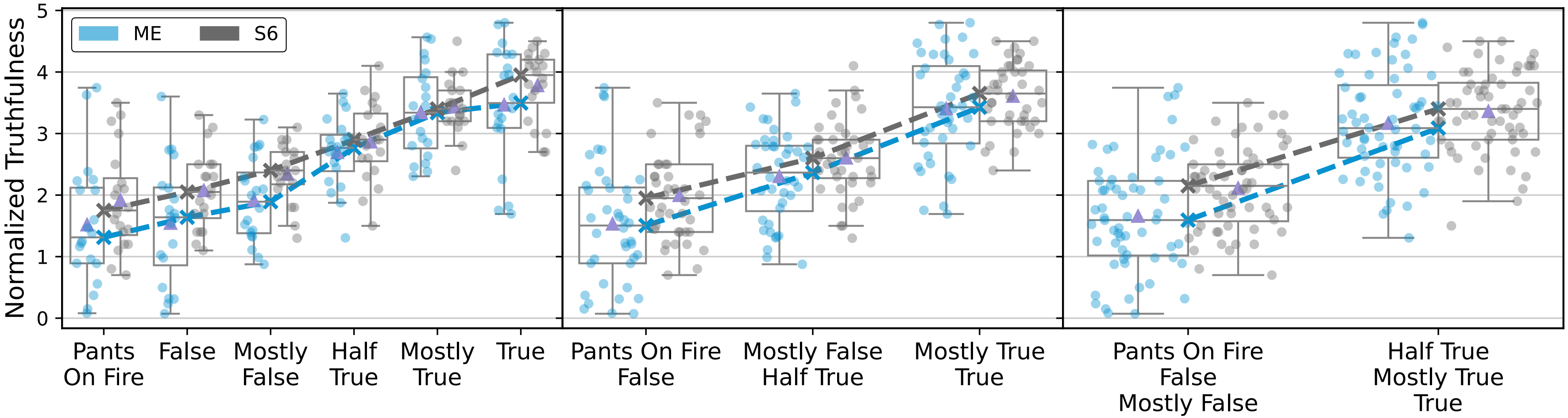


ME vs. S6 & ground truth



- ME: increasing medians/means
- ME vs. S6: similar trends
  - ME lower for low truthfulness values, but lower also for high truthfulness value

ME vs. S6 & ground truth



ME vs. S6: numbers

	Individual	
Measure	**S_6**	**ME**
Accuracy$_2$	0.65	0.63
Accuracy$_3$	0.51	0.46†
Accuracy$_6$	0.29	0.24‡
MAE	1.35	1.66‡
MSE	3.45	4.83‡
Pairwise$_2$	0.61	0.63
Pairwise$_3$	0.58	0.61
Pairwise$_6$	0.56	0.58
External	0.39	0.34
Internal	0.29	0.15‡

Remember Ordinal Classification? No metric works, but we use all of them 😬

ME vs. S6: numbers

Measure	Individual		Aggregated	
	S_6	ME	S_6	ME
Accuracy_2	0.65	0.63	0.83	0.80
Accuracy_3	0.51	$0.46^{\dagger}$	0.60	0.60
Accuracy_6	0.29	$0.24^{\ddagger}$	0.37	0.32
MAE	1.35	$1.66^{\ddagger}$	0.97	0.97
MSE	3.45	$4.83^{\ddagger}$	1.48	1.59
Pairwise_2	0.61	0.63	0.89	0.89
Pairwise_3	0.58	0.61	0.85	$0.83^{\ddagger}$
Pairwise_6	0.56	0.58	0.81	$0.79^{\ddagger}$
External	0.39	0.34	0.61	0.63^{*}
Internal	0.29	$0.15^{\ddagger}$	0.22	0.10^{*}

Remember Ordinal Classification? No metric works, but we use all of them 😬

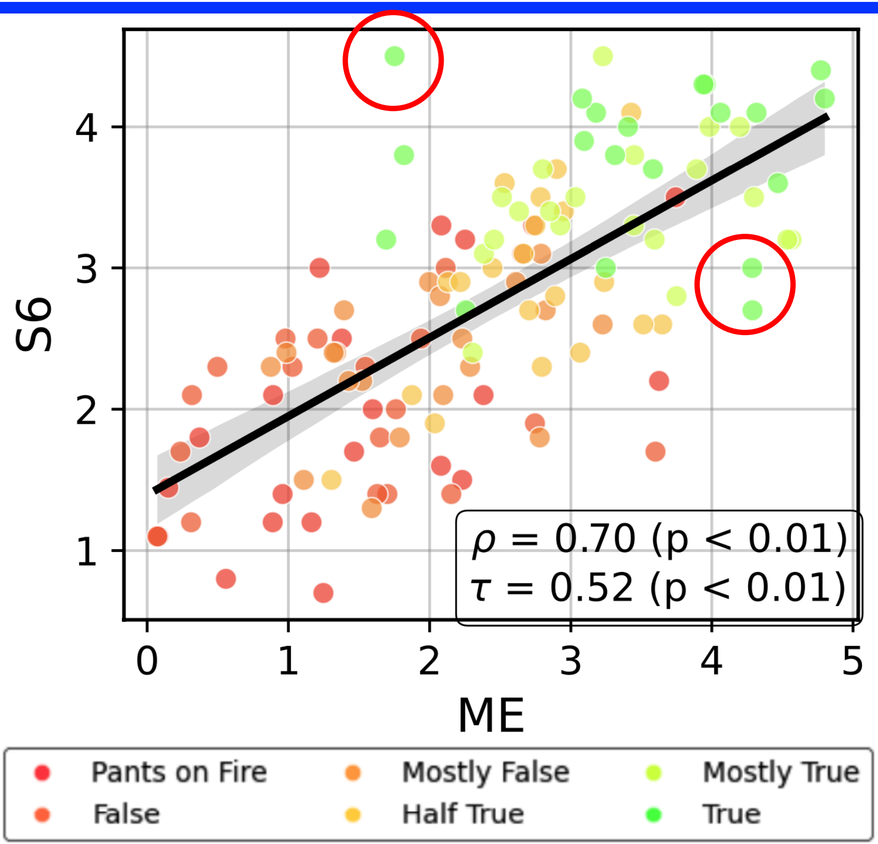
First result

- ME allows to gather reliable truthfulness assessments
- No evidence that ME is worse than S6

Beyond ME vs. S6

- Can ME tell us something that S6 can not?
- Possible, ME is more fine-grained
- Is the additional information useful?
- Yes, in two ways

1. Complementarity



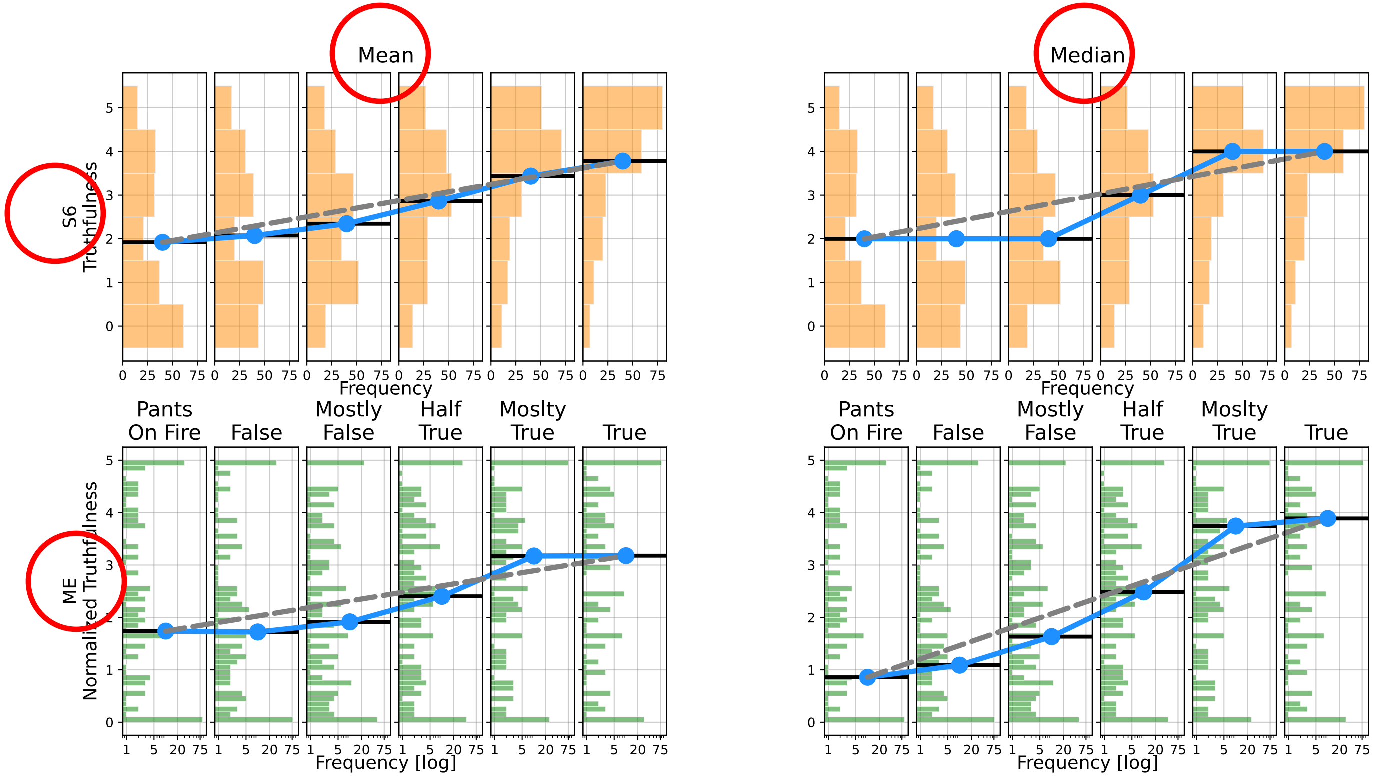
- If ME and S6 were equivalent, dots would be on the diagonal
  - They are not
  - (ME covers more [0,5])
- → Combination of ME and S6?

2. Perception of truthfulness values

- 0, 1, 2, 3, 4, 5?
- Actually, no.

The Magnitude of Truth

Distributions of individual normalized values

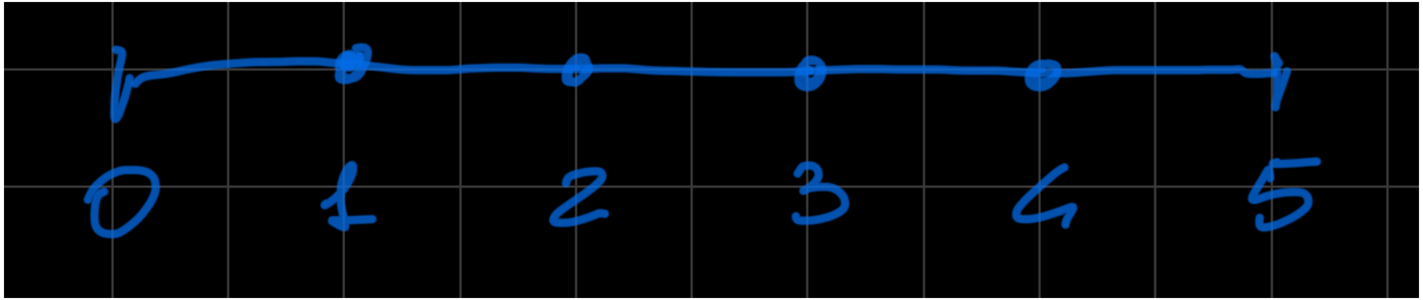


Remember 0, 1, 2, 3, 4, 5?

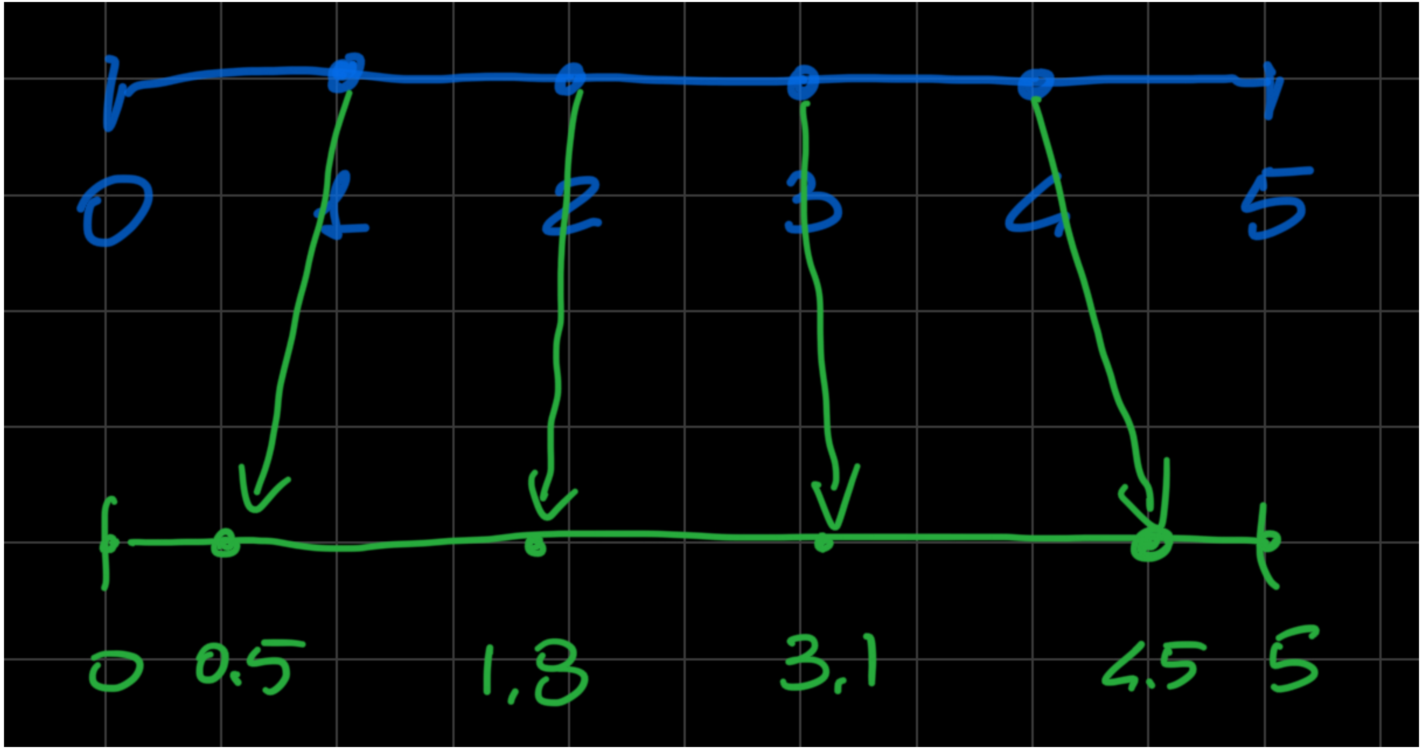
- Pants On Fire ~ False
- Superlinear increase
  - False → Mostly False → Half True → Mostly True
- Mostly True ~ True
- Maybe 0, 0.5. 1.8, 3.1, 4.5, 5?!
  - (even with an "official" scale)
- And maybe the middle of the scale is around Half True
  - Steepest increase

- Less than 1
- More than 1
- Less than 1

From this…



… to something like this



So, Magnitude Estimation

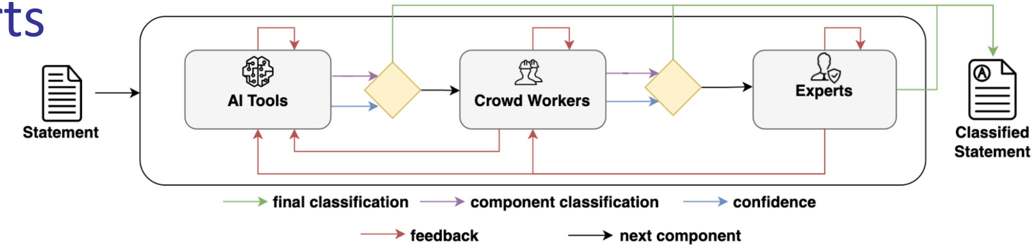
- In terms of accuracy, we couldn't find any compelling reason to prefer S6 to ME
  - (ok, ok, neither the other way around)
- But, ME did provide more information
  - It appears that workers judge differently
  - Useful indication on the scale used by Politifact

Conclusions & Lessons learned

- Lesson 1. Crowdsourcing is a viable alternative (?)
  - Is 0.8 enough?! 🤷‍♂️
- Lesson 2. Improvements are always possible
  - Hard work is needed
  - "Live. Die. Repeat." → "Try. Fail. Repeat."
- Lesson 3. Magnitude Estimation can be used
  - Even though it looked crazy

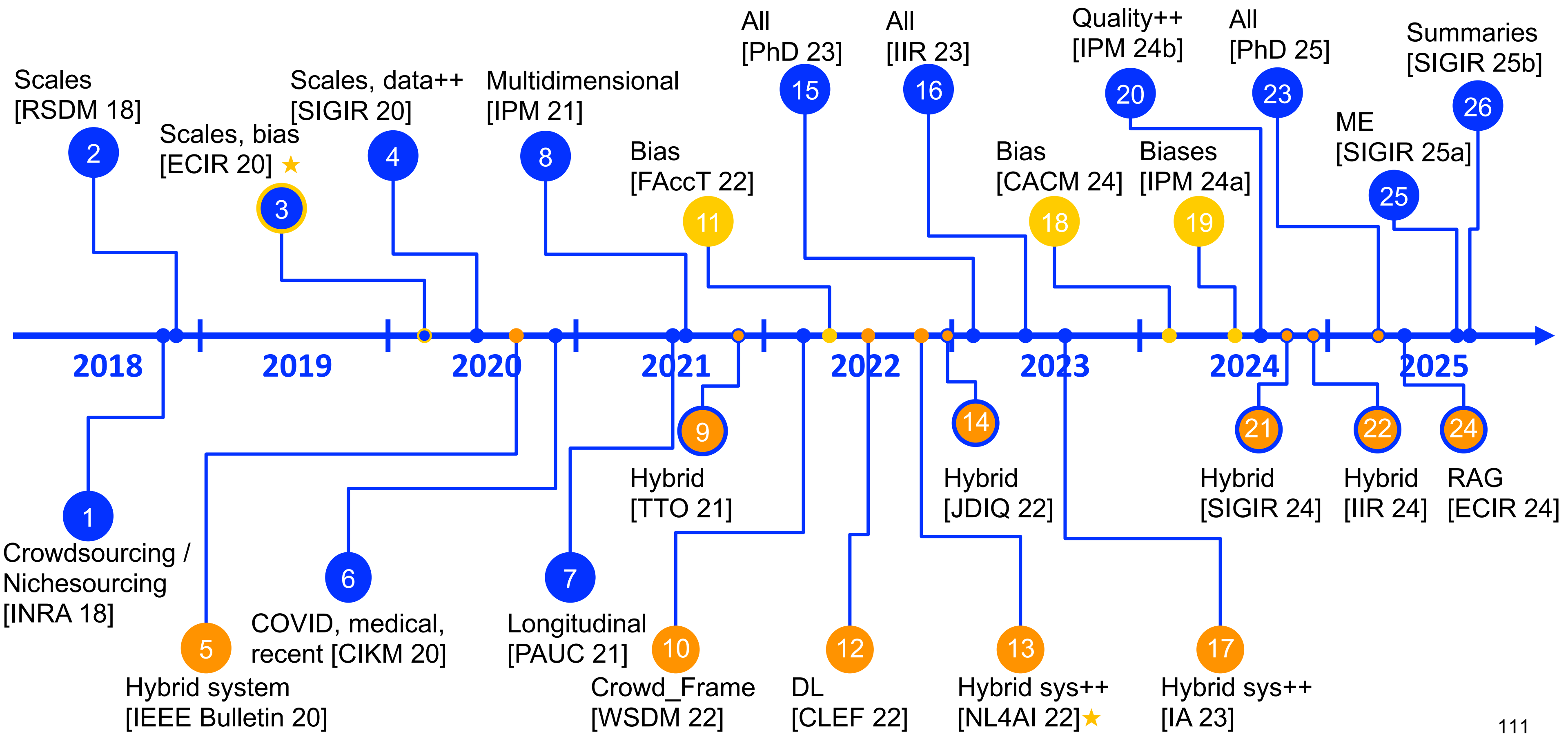
Future

- Changes/Improvement to the task
  - E.g., workers interacting?
- Combination AI / Crowd / Experts
  - AI, Gen AI, LLMs, …
  - So far inconclusive evidence
  - E.g., Confidence seems not useful



- Deep fakes (video, multimedia)
  - (First preliminary results: we are doomed! 😱)
- Summary of evidence: → [SIGIR 2025b]
- Metrics for ordinal classification

Timeline, a last time 😅



Biblio – Our papers, grouped

- Several similar experiments
  - 1. Crowdsourcing vs. Nichesourcing [INRA 2018]
  - 2. Scales [RDSM 2018]
  - 3. Initial, a few data, worker background, scales, bias [ECIR 2020] ★
  - 4. More data, scales [SIGIR 2020]
  - 6. COVID-19: medical, recent news [CIKM 2020]
  - 7. Longitudinal: [PAUC 2021]
  - 8. Multidimensional: Measure more aspects [IPM 2021]
  - 20. Quality > (Prolific, UI, …) [IPM 2024b]
  - 25. ME [SIGIR 2025a]
  - 26. Summaries [SIGIR 2025b]
- Implementation – Hybrid system
  - 5. First proposal [Bulletin 2020]
  - 10. Crowd_Frame [WSDM 2022]
  - 12. Deep Learning [CLEF 2022]
  - 13. System description [NL4AI2022] ★
  - 17. Hybrid system ++ [IA 2023]
- Hybrid approaches Crowd + LLMs
  - 9. First experiments [TTO 2021]
  - 14. Hybrid [JDIQ 2022]
  - 21. Hybrid++ [SIGIR 2024]
  - 22. Hybrid [IIR 2024]
  - 24. RAG [ECIR 2025]
- Bias
  - 11. Bias [FAccT 2022]
  - 18. Bias management [CACM 2023]
  - 19. Biases in Fact-checking [IPM 2024a]
- All
  - 15. PhD Thesis Soprano [PhD 2023]
  - 16. Summary [IIR 2023]
  - 23. PhD Thesis La Barbera [PhD 2025]

(details in next slides)

1. [INRA 2018]

- E. Maddalena, D. Ceolin, S. Mizzaro. "Multidimensional News Quality: A Comparison of Crowdsourcing and Nichesourcing." In *CIKM Workshops* -- Proceedings of the 6th International Workshop on News Recommendation and Analytics (INRA 2018), 2018.
- http://ceur-ws.org/Vol-2482/paper17.pdf

2. [RDSM 2018]

- K. Roitero, G. Demartini, S. Mizzaro, D. Spina. How Many Truth Levels? Six? One Hundred? Even More? Validating Truthfulness of Statements via Crowdsourcing. In *CIKM Workshops* -- 2nd International Workshop on Rumours and Deception in Social Media (RDSM). 2018.
- https://www.damianospina.com/publication/roitero-2018-how/roitero-2018-how.pdf

3. [ECIR 2020]

- D. La Barbera, K. Roitero, G. Demartini, S. Mizzaro, D. Spina. Crowdsourcing Truthfulness: The Impact of Judgment Scale and Assessor Bias. Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020. 207-214
- Video:

https://www.youtube.com/watch?v=9wFFMcplvjk

ECIR 2020 @ecir2020 · 1g

#ECIR2020 Best **Short** Paper **Award** goes to "Crowdsourcing Truthfulness; The Impact of Judgment Scale and Assessor Bias" authored by
David la Barbera, Kevin Roitero, Damiano Spina, Stefano Mizzaro, Gianluca Demartini. Congratulations!!

1 3 20

4. [SIGIR 2020]

- K. Roitero, M. Soprano, S. Fan, D. Spina, S. Mizzaro, G. Demartini. Can The Crowd Identify Misinformation Objectively? The Effects of Judgment Scale and Assessor's Background. Proceedings of the 43st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020). Xi'an, China (Online). July 25-30, 2020, https://doi.org/10.1145/3397271.3401112
- Video:

https://www.youtube.com/watch?v=D10EtrThvbc

5. [Bulletin 2020]

- G. Demartini, S. Mizzaro, D. Spina. Human-in-the-loop artificial intelligence for fighting online misinformation: challenges and opportunities. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 43.3 (2020): 65-74.
- https://www.damianospina.com/publication/demartini-2020-human/demartini-2020-human.pdf

6. [CIKM 2020]

- K. Roitero, M. Soprano, B. Portelli, D. Spina, V. Della Mea, G. Serra, S. Mizzaro, G. Demartini. The COVID-19 Infodemic: Can the Crowd Judge Recent Misinformation Objectively?, Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM 2020). Galway, Ireland (Online). October 19-23, 2020, https://doi.org/10.1145/3340531.3412048
- Video: https://dl.acm.org/doi/10.1145/3340531.3412048

7. [PAUC 2021]

- K. Roitero, M. Soprano, B. Portelli, M. De Luise, D. Spina, V. Della Mea, G. Serra, S. Mizzaro, G. Demartini. Can the crowd judge truthfulness? A longitudinal study on recent misinformation about COVID-19. *Personal and Ubiquitous Computing* (2021): 1-31.
- https://www.damianospina.com/publication/roitero-2021-can/roitero-2021-can.pdf

8. [IPM 2021]

- M. Soprano, K. Roitero, D. La Barbera, D. Ceolin, D. Spina, S. Mizzaro, G. Demartini. The many dimensions of truthfulness: Crowdsourcing misinformation assessments on a multidimensional scale. *Information Processing & Management*, *58*(6), 102710, 2021.
- https://arxiv.org/pdf/2108.01222.pdf

9. [TTO 2021]

- Y. Qu, K. Roitero, S. Mizzaro, D. Spina, G. Demartini. Human-in-the-Loop Systems for Truthfulness: A Study of Human and Machine Confidence. *TTO 2021* (2021): 40.
- https://truthandtrustonline.com/wp-content/uploads/2021/11/TTO_2021_proceedings.pdf#page=52

10. [WSDM 2022]

- M. Soprano, K. Roitero, F. Bombassei De Bona, S. Mizzaro. Crowd_Frame: A Simple and Complete Framework to Deploy Complex Crowdsourcing Tasks Off-the-shelf. Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22), 1605-1608, 2022. https://doi.org/10.1145/3488560.3502182

11. [FAccT 2022]

- T. Draws, D. La Barbera, M. Soprano, K. Roitero, D. Ceolin, A. Checco, S. Mizzaro. 2022. The Effects of Crowd Worker Biases in Fact-Checking Tasks. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’22), June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 17 pages.
- https://doi.org/10.1145/3531146.3534629

12. [CLEF 2022]

- D. La Barbera, K. Roitero, J. Mackenzie, D. Spina, G. Demartini, S. Mizzaro. BUM at CheckThat! 2022: a composite deep learning approach to fake news detection using evidence retrieval, Working Notes of CLEF, 2022

13. [NL4AI 2022]

- D. La Barbera, K. Roitero, S. Mizzaro. A Hybrid Human-In-The-Loop Framework for Fact Checking, Proceedings of the Sixth Workshop on Natural Language for Artificial Intelligence (NL4AI 2022) co-located with 21th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2022), November 2022
- http://sag.art.uniroma2.it/NL4AI/wp-content/uploads/2022/11/paper4.pdf

AILC

NL4AI

Associazione Italiana per l'Intelligenza Artificiale

CERTIFICATE

FOR BEST PAPER

NL4AI 2022

This certificate is proudly awarded by

David La Barbera, Kevin Roitero and Stefano Mizzaro

"A Hybrid Human-In-The-Loop Framework for Fact Checking"

Chairs

Debora Nozza, Lucia Passaro, Marco Polignano

14. [JDIQ 2022]

- Y. Qu, D. La Barbera, K. Roitero, S. Mizzaro, D. Spina, G. Demartini. Combining Human and Machine Confidence in Truthfulness Assessment. ACM Journal of Data and Information Quality (JDIQ), 2022.

15. [PhD 2023]

- M. Soprano. In Crowd Veritas: Leveraging Human Intelligence To Fight Misinformation. PhD Thesis, University of Udine, 2023
  - (I was just the supervisor)

16. [IIR 2023]

- D. La Barbera, M. Soprano, K. Roitero, E. Maddalena, S. Mizzaro. Fact-Checking at Scale with Crowdsourcing: Experiments and Lessons Learned. Proceedings of the 13th Italian Information Retrieval Workshop, CEUR-WS.org, volume=3448, pages 85-90, 2023

17. [IA 2023]

- D. La Barbera, K. Roitero, S. Mizzaro. Combining human intelligence and machine learning for fact-checking: Towards a hybrid human-in-the-loop framework. Intelligenza Artificiale, 2023, IOS Press.

18. [CACM 2023]

- G. Demartini, K. Roitero, S. Mizzaro. Data Bias Management, *Communications of the ACM*, Volume 67(1), pagg. 28-32, 2023
- https://cacm.acm.org/opinion/data-bias-management/

19. [IPM 2024a]

- M. Soprano, K. Roitero, D. La Barbera, D. Ceolin, D. Spina, G. Demartini, S. Mizzaro. Cognitive Biases in Fact-Checking and Their Countermeasures: A Review. Information Processing & Management, Volume 61(3), May 2024
- https://www.sciencedirect.com/science/article/pii/S0306457324000323

20. [IPM 2024b]

- D. La Barbera, E. Maddalena, M. Soprano, K. Roitero, G. Demartini, D. Ceolin, D. Spina, S. Mizzaro. Crowdsourced Fact-Checking: Does it Actually Work?, *Information Processing & Management*, Volume 61(5), September 2024
- https://www.sciencedirect.com/science/article/pii/S0306457324001523

21. [SIGIR 2024]

- X. Zeng, D. La Barbera, K. Roitero, A. Zubiaga, S. Mizzaro. Combining Large Language Models and Crowdsourcing for Hybrid Human-AI Misinformation Detection, Proceedings of the 47th International ACM SIGIR, pagg. 2332-2336, 2024
- https://dl.acm.org/doi/pdf/10.1145/3626772.3657965

22. [IIR 2024]

- K. Roitero, M. Soprano, D. La Barbera, E. Maddalena, S. Mizzaro. Enhancing Fact-Checking: From Crowdsourced Validation to Integration with Large Language Models. Proceedings of the 14th Italian Information Retrieval Workshop, IIR 2024, 2024

23. [PhD 2025]

- D. La Barbera. Human-AI Collaboration in Fact-Checking. PhD Thesis, University of Udine, 2025
  - (I was just the supervisor)

24. [ECIR 2025]

- F. Bombassei De Bona, D. La Barbera, S. Mizzaro, K. Roitero. A Comparative Analysis of Retrieval-Augmented Generation and Crowdsourcing for Fact-Checking. In: Hauff, C., et al. Advances in Information Retrieval. ECIR 2025. Lecture Notes in Computer Science, vol 15574. Springer, Cham, 2025. https://doi.org/10.1007/978-3-031-88714-7_44

25. [SIGIR 2025a]

- M. Soprano, D. E. Tapu, D. La Barbera, K. Roitero, S. Mizzaro. The Magnitude of Truth: On Using Magnitude Estimation for Truthfulness Assessment. Accepted as full paper at SIGIR 2025 🎉

26. [SIGIR 2025b]

- K. Roitero, D. Wright, M. Soprano, I. Augenstein, S. Mizzaro. Efficiency and Effectiveness of LLM-Based Summarization of Evidence in Crowdsourced Fact-Checking. Accepted as full paper at SIGIR 2025 🎉