

# LLMs as Democratizing Platform to Combat Disinformation



---

The Pennsylvania State University, USA

**Dongwon Lee, Ph.D.**

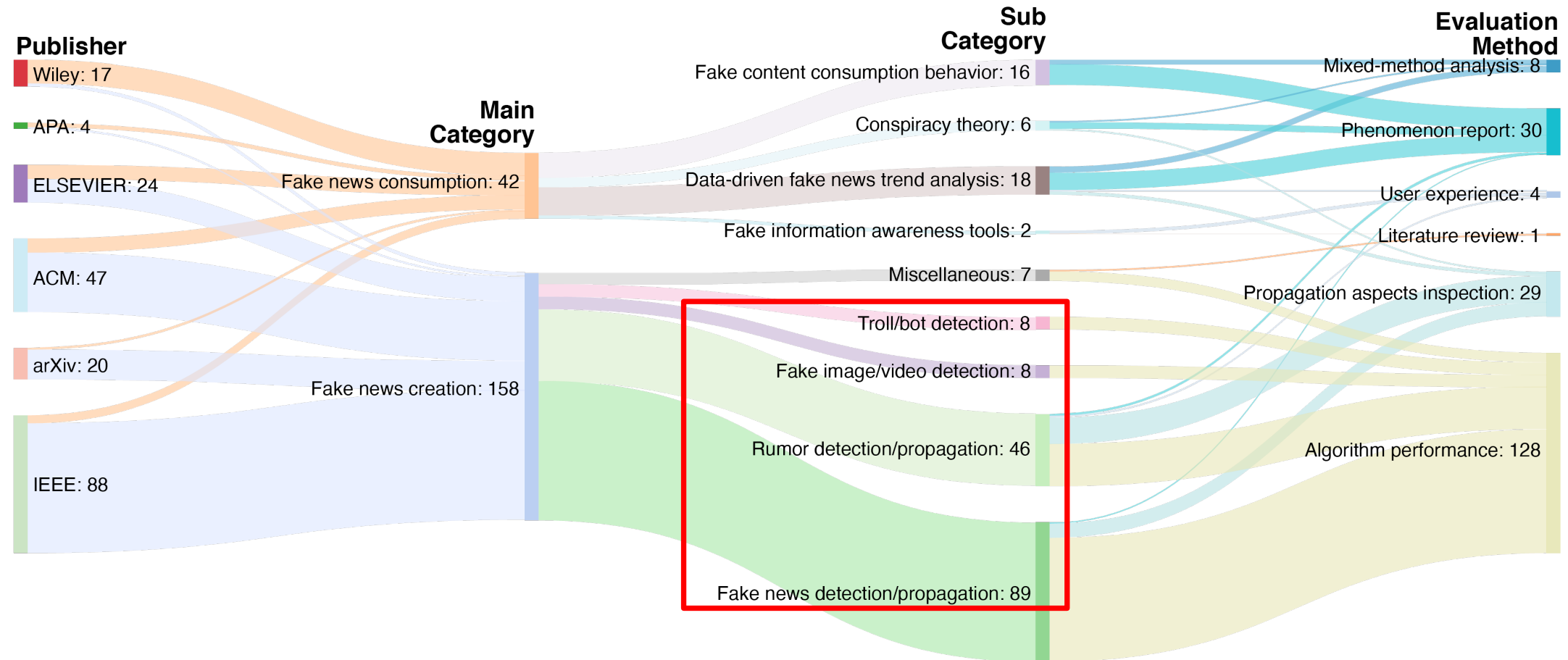
[dongwon@psu.edu](mailto:dongwon@psu.edu)

**PennState**



*[Evolving Technological, Legal and Social Solutions to Counter Disinformation in Social Media - A Workshop](#) @ April 11, 2024*

# Extensive disinformation research in academia



A visualization of 200 fake news-related papers published in IEEE, ACM, ELSEVIER, arXiv, Wiley, APA from 2010 to 2020

# Deployed in-house solutions in industry

- Social platforms have in-house disinfo detectors deployed

## Meta has a new AI tool to fight misinformation—and it's using Wikipedia to train itself

Published Wed, Jul 13 2022•11:19 AM EDT •  
 Updated Wed, Jul 20 2022•12:11 PM EDT

Company

## An update on reporting potential misinformation on Twitter

By [Twitter Philippines](#)

Monday, 17 January 2022    

 Meta AI

APPLICATIONS | INTEGRITY | Research | Publications | People | Ever

## Here's how we're using AI to help detect misinformation

November 19, 2020



 Watch

**Register**

Sign In

## Instagram fact-check: Can a new flagging tool stop fake news?

13 September 2019

By **Shayan Sardarizadeh**, BBC Monitoring

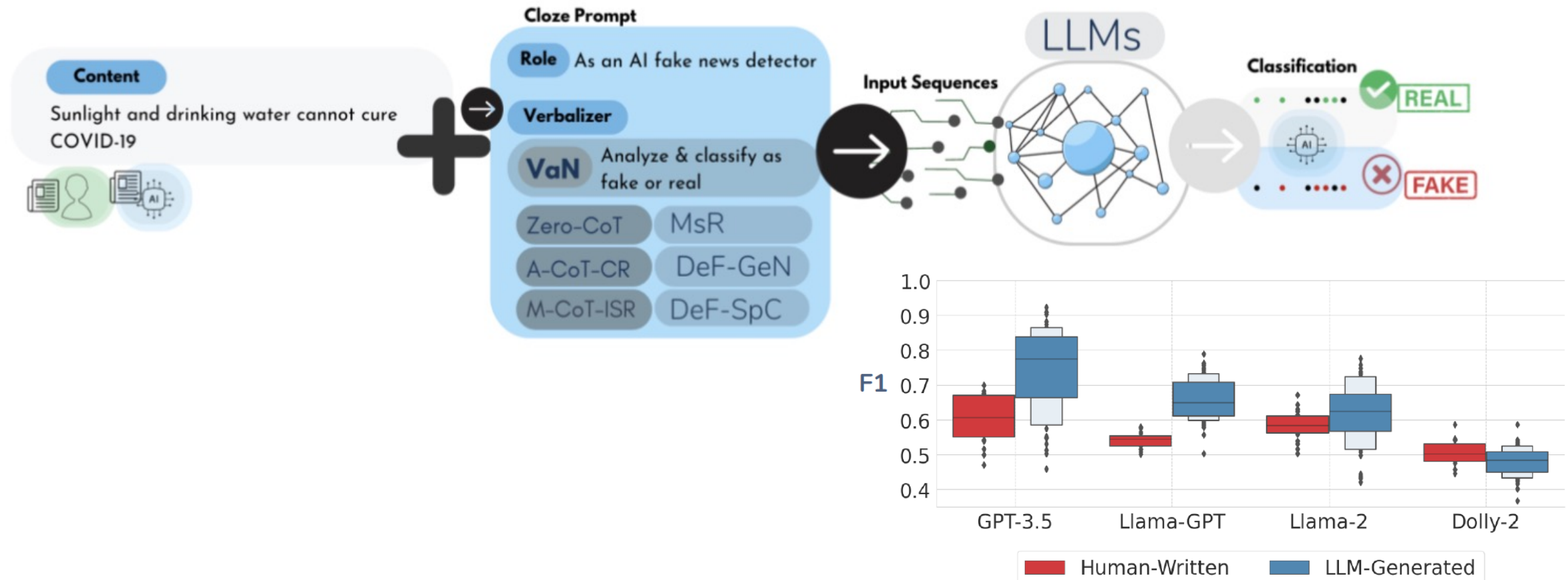
 Share

# Challenges for technological interventions

- Solutions emerging from academia have
  - constrained functionalities
  - limited adoption—ie, most stay in labs only as prototypes
- Industrial platforms employ proprietary in-house solutions
  - lack standardized approaches, transparency, and trust
- “Long-tail” users from low-resource regions have more challenges
  - existing solutions tend to be biased toward “big-head” users
- Gap between innovative technological advancements in research and their implementation in real-world contexts presents significant challenge

# LLMs as platform to combat disinformation

- Can LLMs detect disinformation that other LLMs or humans actors generate?



# LLMs as **democratizing** platform

- LLM is becoming universal AI layer, brokering between humans and tasks
- Despite current issues in LLMs
  - biases, memorizations, hallucinations, ...
- LLMs can be an effective standardized platform to combat disinformation
  - across vendors, social platforms, formats, ...
  - across languages, nations, ...