Characterizing healthcare workers vaccine narratives on X (Twitter).

Leonardo Heyerdahl – Postdoctoral researcher,
Anthropology and Ecology of Disease Emergence Unit, Global Health Department, Institut Pasteur

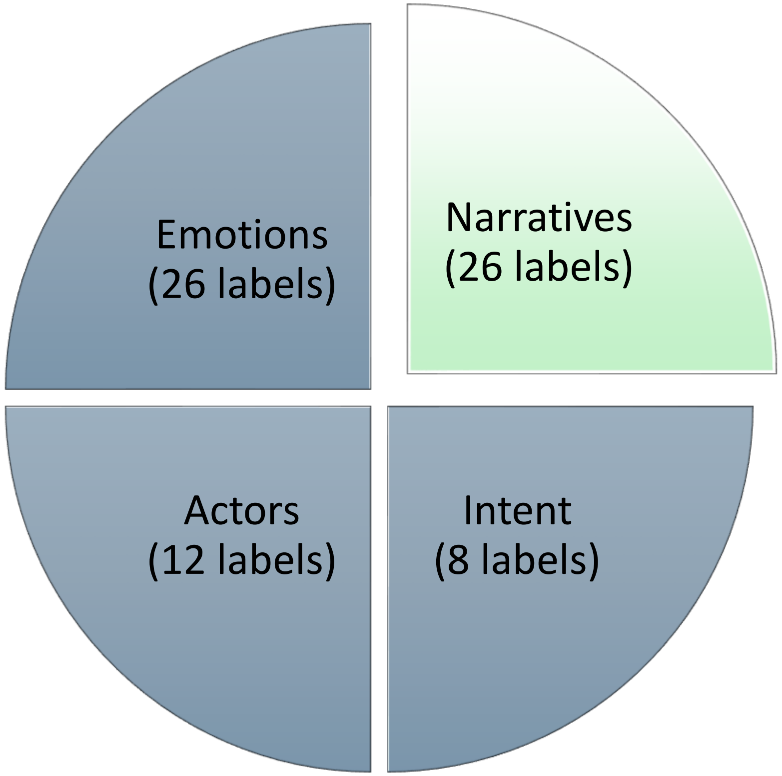
Les professionnels de santé, acteurs et sujets de la vaccination

1ÈRE JOURNÉE D'ÉTUDES DU RÉSEAU SHS-VACCINATION

PARIS 24 JANVIER 2025

Objective

- To characterize online vaccine discourses among healthcare workers at scale using custom taxonomies.



Description of the AEE Twitter database

Daily collection of tweets (COVID, vaccination, drugs, masks, etc.)

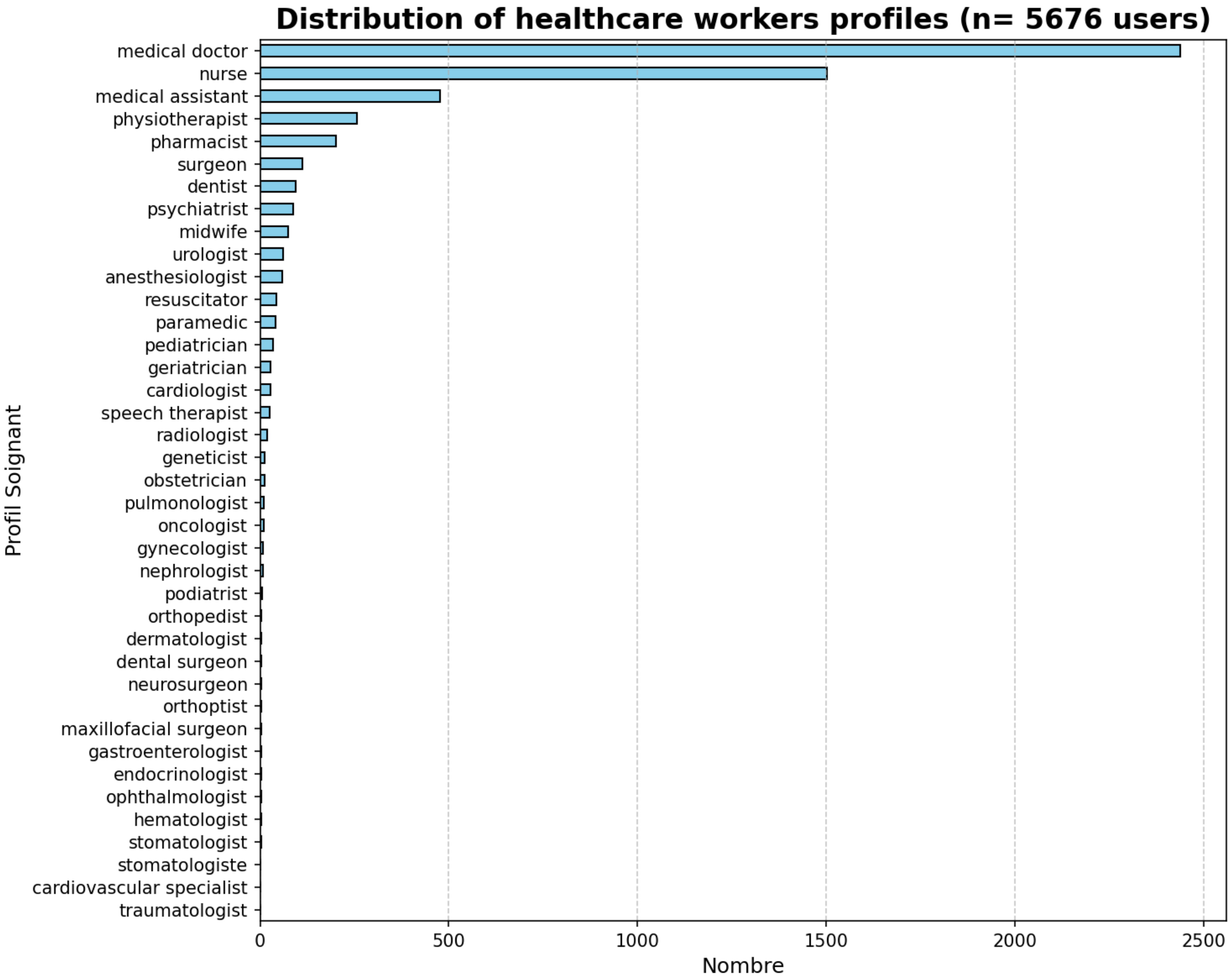
- From February 2020 – present
- Vaccination: >12 million organic tweets

Previous uses:

- Hybrid social listening
- Periodic thematic analysis
- Triangulation: interviews, questionnaires (RCCOVID, RECOVER, Transvaxx, VCF Unspoken)

Sample discussed today:

- Healthcare profiles: filtered by Twitter bio / medical professions nomenclature
- Period: 2020–2023
- ~ 94,000 tweets



Vaccine narratives taxonomy – ACME

Topics	labels
Safety	safe
	unsure_safe
	not_safe
Usefulness	useful
	unsure_useful
	not_useful
Accessibility	not_accessible
	unsure_accessible
	accessible
	waste
Pathogenic Risk	risky_pathogen
	unsure_risky_pathogen
	not_risky_pathogen
Polarization	criticizing_pro_vax
	criticizing_anti_vax
	unsanitary_other
	de-escalating
Actors' benevolence & competence	trust_actors
	unsure_trust_actors
	not_trust_actors
Probity	vaccine_profiteering
	conspiracy_ideation
Legitimacy	reactance
	coercion_legitimate
information	generic
	candid_question

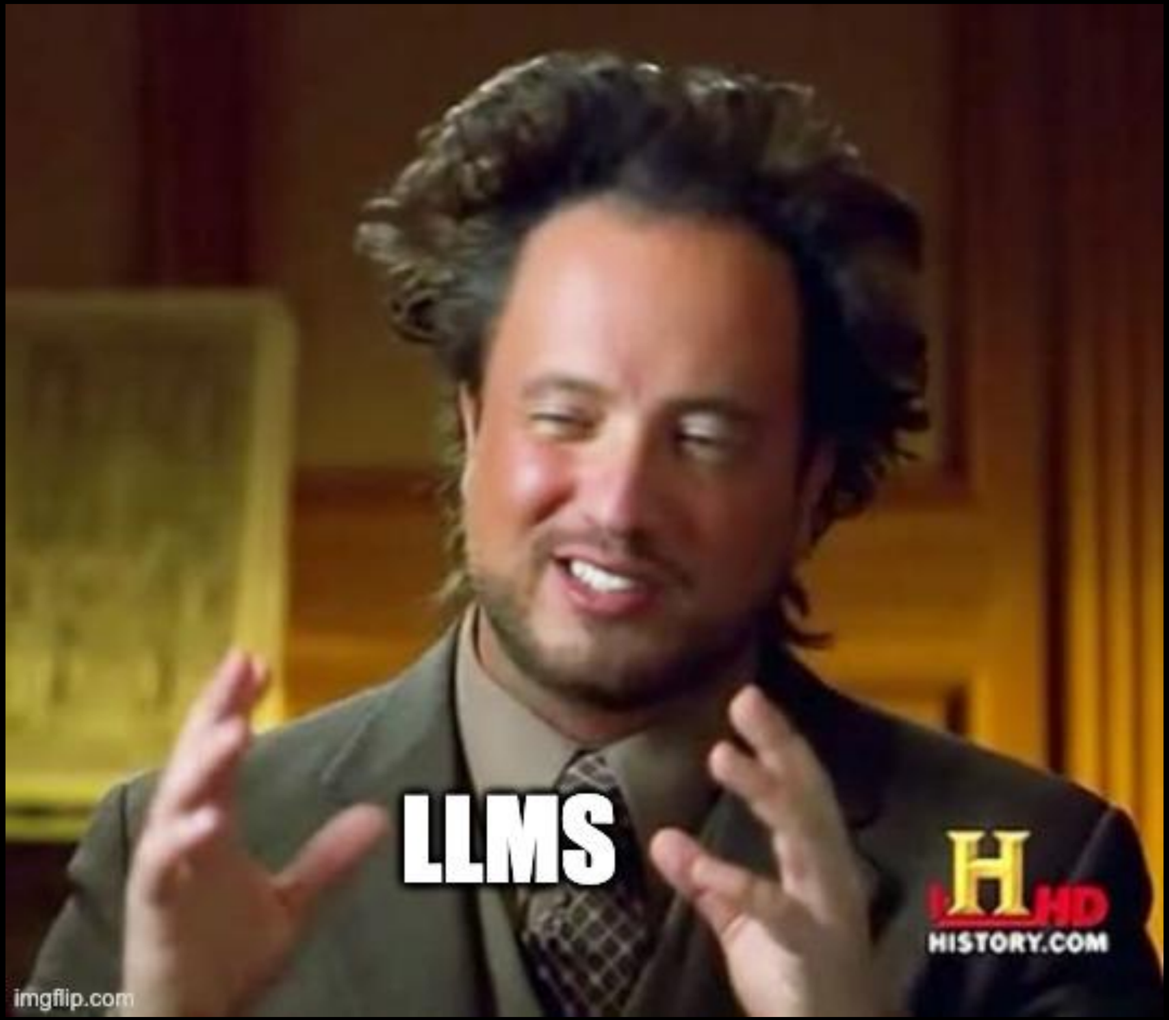
Hand labeled (qualitatively coded) 600 messages.

~ 93,400 to go.

How to proceed?

Evolution of supervised classification options

Era	Techniques	Accuracy	Training/Inference Speed	Volume of Labeled Data Required
Classic Text Mining (2008-2013)	Bag of Words, TF-IDF (-> modelling with RF, XGBOOST using the vectors)	Moderate	Fast	Several thousand labeled texts per label
Word Embeddings (2013-2018)	Word2Vec, GloVe	Improved	Moderate	>500 Hundred to thousands of labeled texts per label
Deep Learning & Transfer Learning (>2018)	LSTM, Transformer models (e.g., BERT, CamemBERT)	Very high	Moderate	>500 Hundred to a few thousand labelled texts
>2020	?	?	?	?


LLMS
HISTORY.COM
imgflip.com

Evolution of supervised classification options

Era	Techniques	Accuracy	Training/Inference Speed	Volume of Labeled Data Required
Classic Text Mining (2008-2013)	Bag of Words, TF-IDF (-> modelling with RF, XGBOOST using the vectors)	Moderate	Fast	Several thousand labeled texts per class
Word Embeddings (2013-2018)	Word2Vec, GloVe	Improved	Moderate	>500 Hundred to thousands of labeled texts per class
Deep Learning & Transfer Learning (>2018)	LSTM, Transformer models (e.g., BERT)	Very high	Moderate	>500 Hundred to a few thousand labelled texts
Large Language Models (>2020)	**Few-shot, zero-shot learning (e.g., GPT-4, Gemini)**	**Moderate ‹-› Very High**	**Relatively slow**	**Minimal (dozens to hundreds of texts), or even zero-shot**

Conditions to integrating Large Language Models (LLMs)

Autonomy

- Augmenting –not replacing– analytical capacity -> deductive coding
- Ability to evaluate different models, prompting strategies
- Keeping data within research group

Closed models

ChatGPT

Gemini Google

Claude

Pros

- State of the art, highly efficient models
- Local compute not requiered

Cons

- Closed models
- Data & confidentiality ?

Open source models

Llama 3

MISTRAL AI_

Command R+

[...]

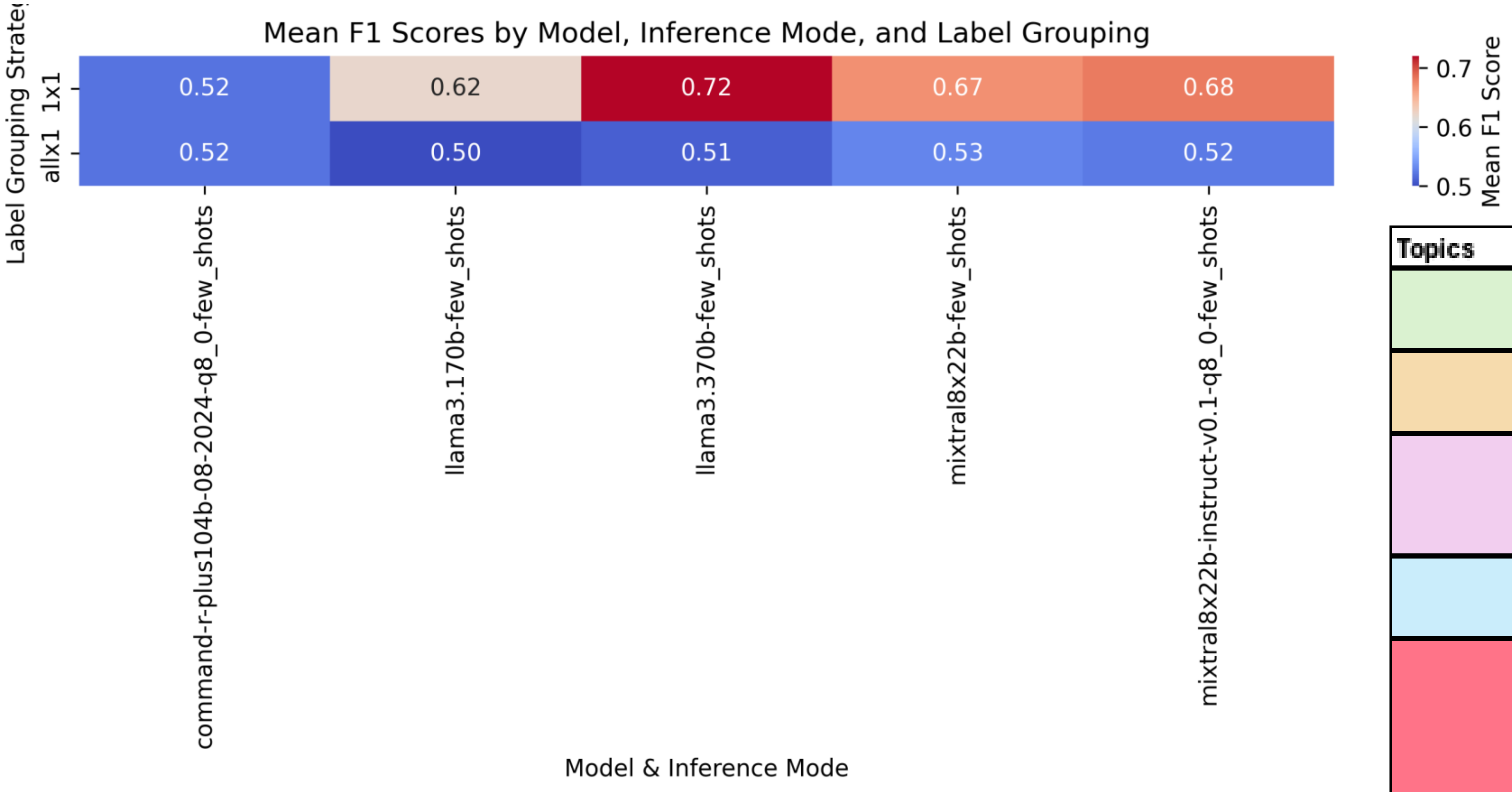
Pros

- Open source
- Run locally
- Data is not shared

Cons

- Less accurate
- Requieres local compute

LLMs Evaluation

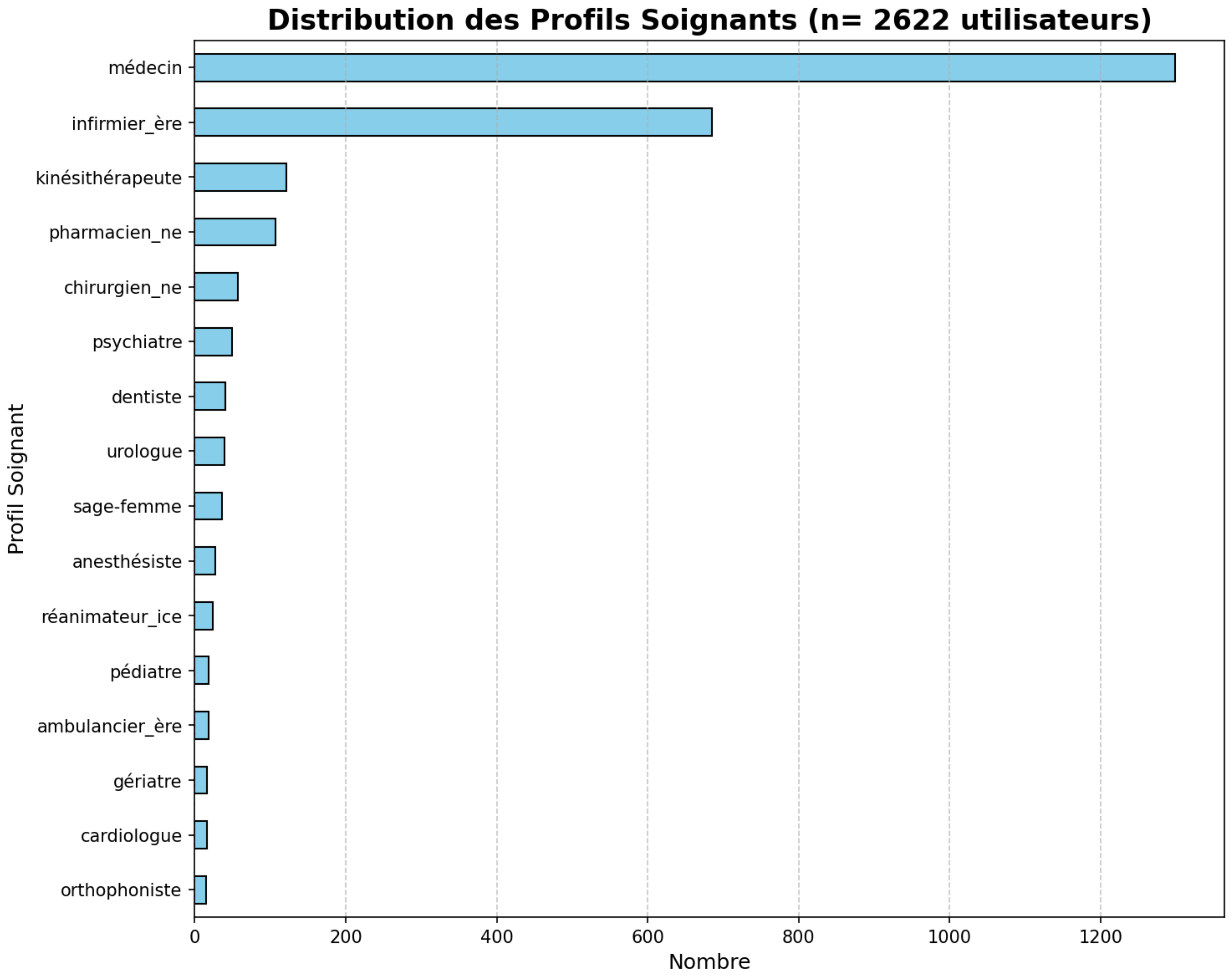


Mixture of models without *uncertain* labels F1 = **0.79**

Topics	labels
Safety	safe
	not_safe
Usefulness	useful
	not_useful
Accessibility	not_accessible
	accessible
	waste
Pathogenic Risk	risky_pathogen
	not_risky_pathogen
Polarization	criticizing_pro_vax
	criticizing_anti_vax
	unsanitary_other
	de-escalating
Actors' benevolence & competence	trust_actors
	not_trust_actors
Probity	vaccine_profiteering
	conspiracy_ideation
Legitimacy	reactance
	coercion_legitimate
information	generic
	candid_question

- Gains through prompt engineering:
    - Contextualizing the task
    - Disambiguating (Collaborative Qual Coding: human / machine) – Ongoing

Healthcare worker profiles — within the LLM labelled dataset (22,863 tweets)



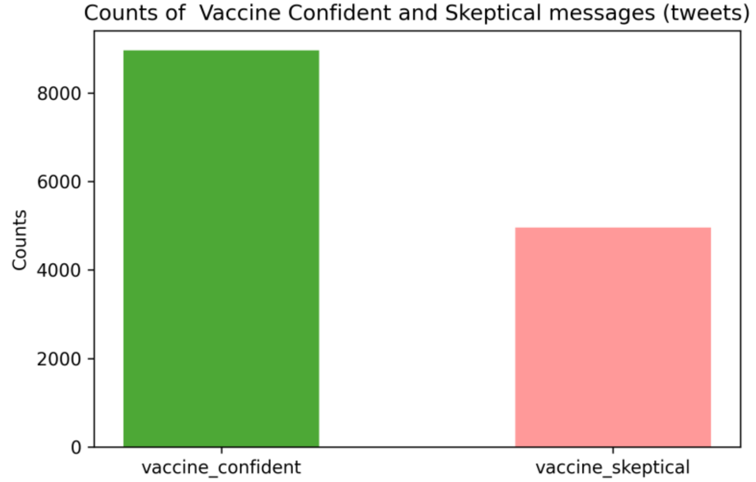
MD over represented:

- Status / desirability?
- Specialists not detailing?

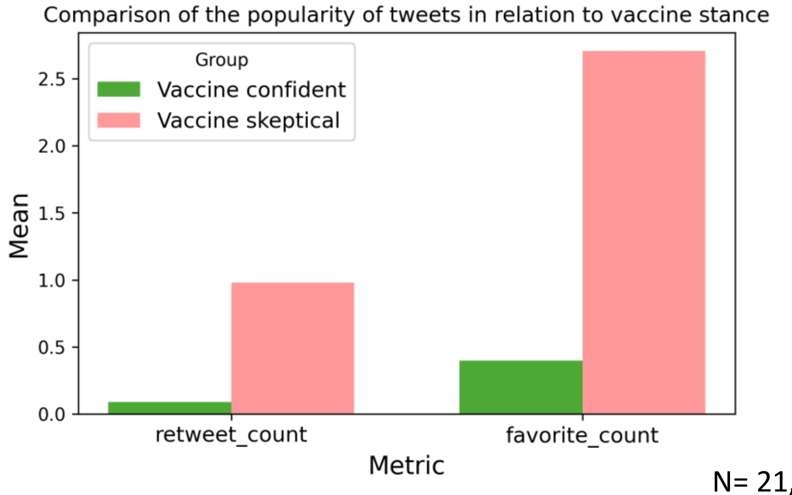
HCW Vaccine skepticism spreads faster

Overarching *vaccine_skeptical*, *vaccine_confident* variables created based on key labels (not_safe, safe..)

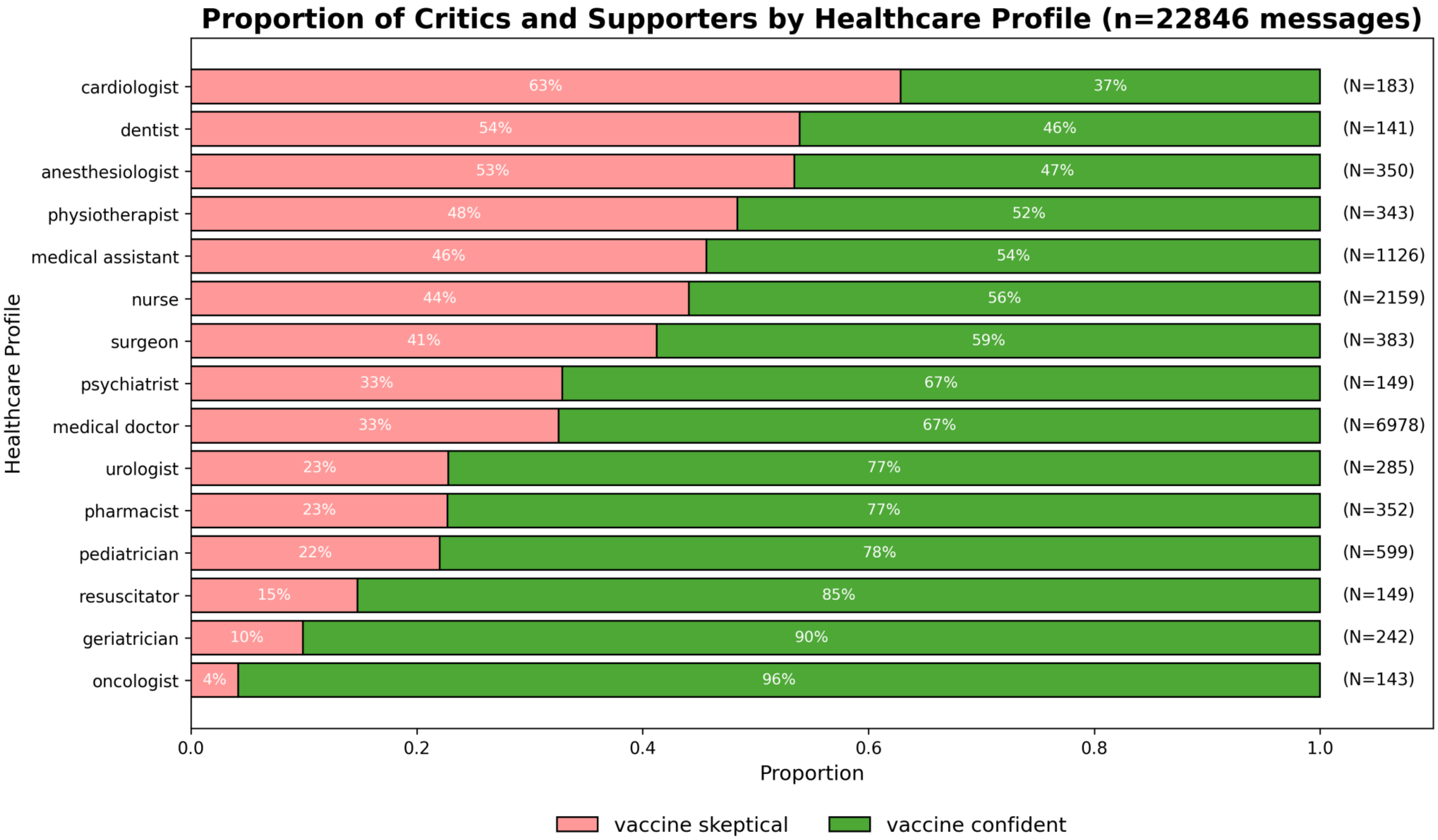
Overall, messages are predominantly vaccine confident (nearly x2)



But vaccine skeptical messages are more shared (RT) and receive more favorites.

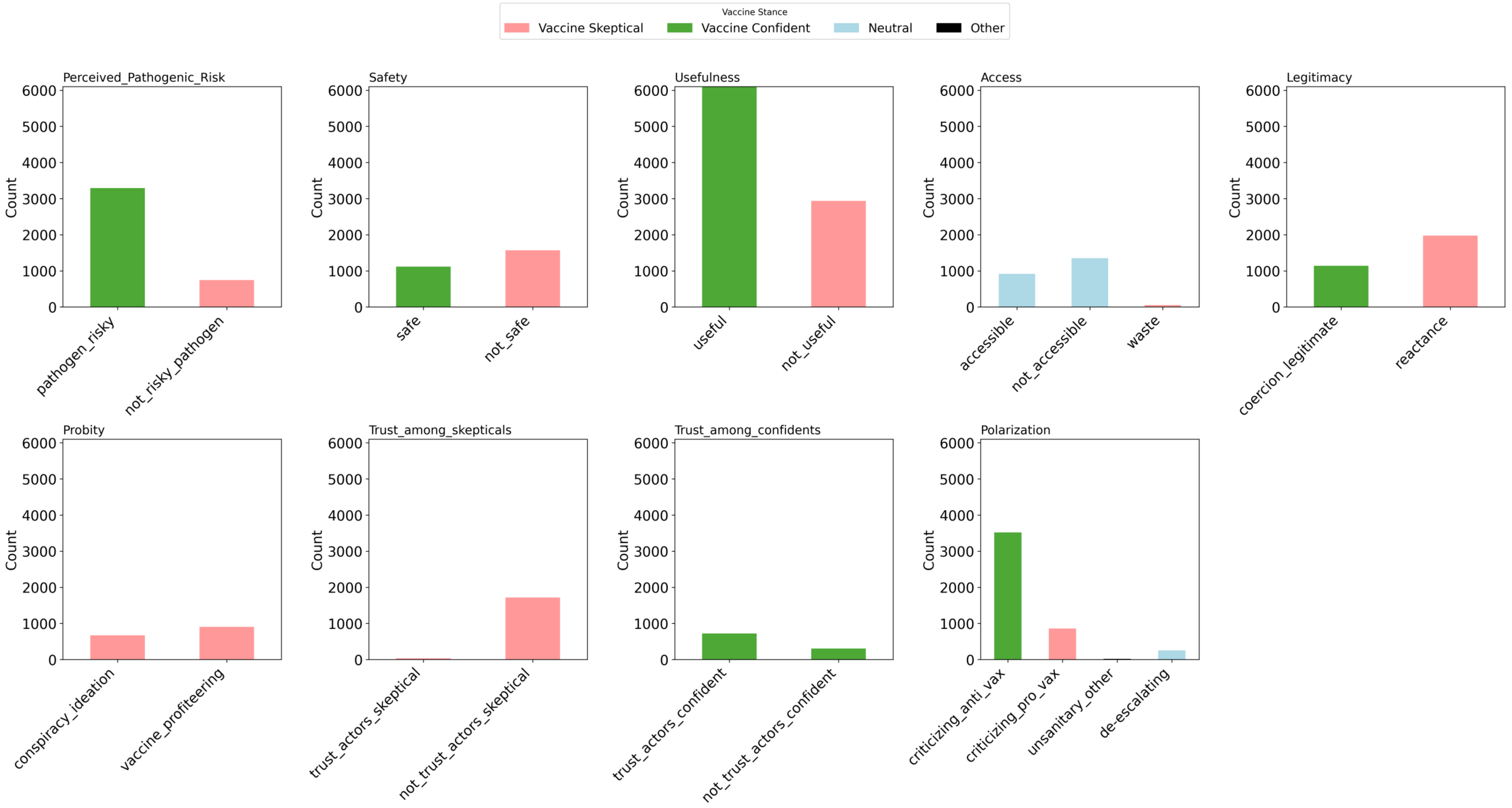


N= 21,923 HCW messages



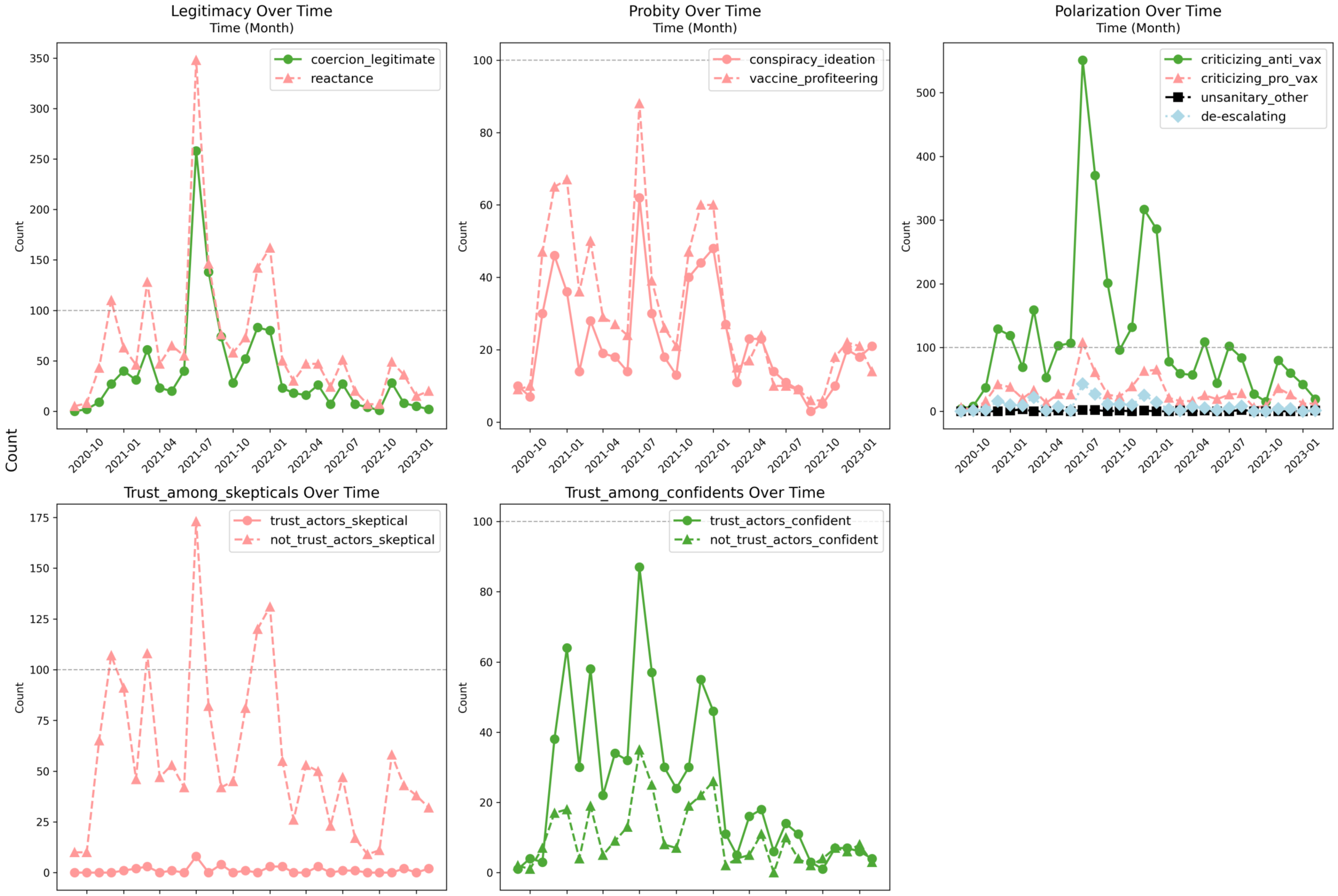
Proportion of Critics and Supporters by Healthcare Profile (n=22846 messages)
Healthcare Profile
cardiologist
dentist
anesthesiologist
physiotherapist
medical assistant
nurse
surgeon
psychiatrist
medical doctor
urologist
pharmacist
pediatrician
resuscitator
geriatrician
oncologist
63%
37%
54%
46%
53%
47%
48%
52%
46%
54%
44%
56%
41%
59%
33%
67%
33%
67%
23%
77%
23%
77%
22%
78%
15%
85%
10%
90%
4%
96%
(N=183)
(N=141)
(N=350)
(N=343)
(N=1126)
(N=2159)
(N=383)
(N=149)
(N=6978)
(N=285)
(N=352)
(N=599)
(N=149)
(N=242)
(N=143)
0.0
0.2
0.4
0.6
0.8
1.0
Proportion
vaccine skeptical
vaccine confident

Label distributions (profile, volume, time)



N= 22,846 HCW messages

N= 22,846 HCW messages



N= 22,846 HCW messages

Briefly

- Limitations
  - Preliminary
  - Big data is *thin* data: need to ground / triangulate
- Next steps
  - Enhance human / machine intercoder reliability (disambiguation, prompt engineering, new models).
  - Integration
    - ACME
    - DEV-SHP (ANR, > Feb 25):
      - (Unspoken) vaccine sentiments, climate of vaccine discussion w/ peers & patients.
      - co-construction of vaccine dialogue initiatives.
      - Anthropology, epidemiology, machine learning

Thank you

- ACME Hybrid Social Listening
    - Gaston Bizel-Bizellot
- ACME WP1 Factors of uptake and adherence, confidence and social equity with regard to epidemic PCMs
    - Judith Mueller
    - Pierre Verger
    - Jocelyn Raude
    - Kathy McColl
- Anthropology and Ecology of Disease Emergence, Institut Pasteur
    - Tamara Giles-Vernick