



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Do we trust or fear Artificial Intelligence?

Prof. Gianluca Demartini

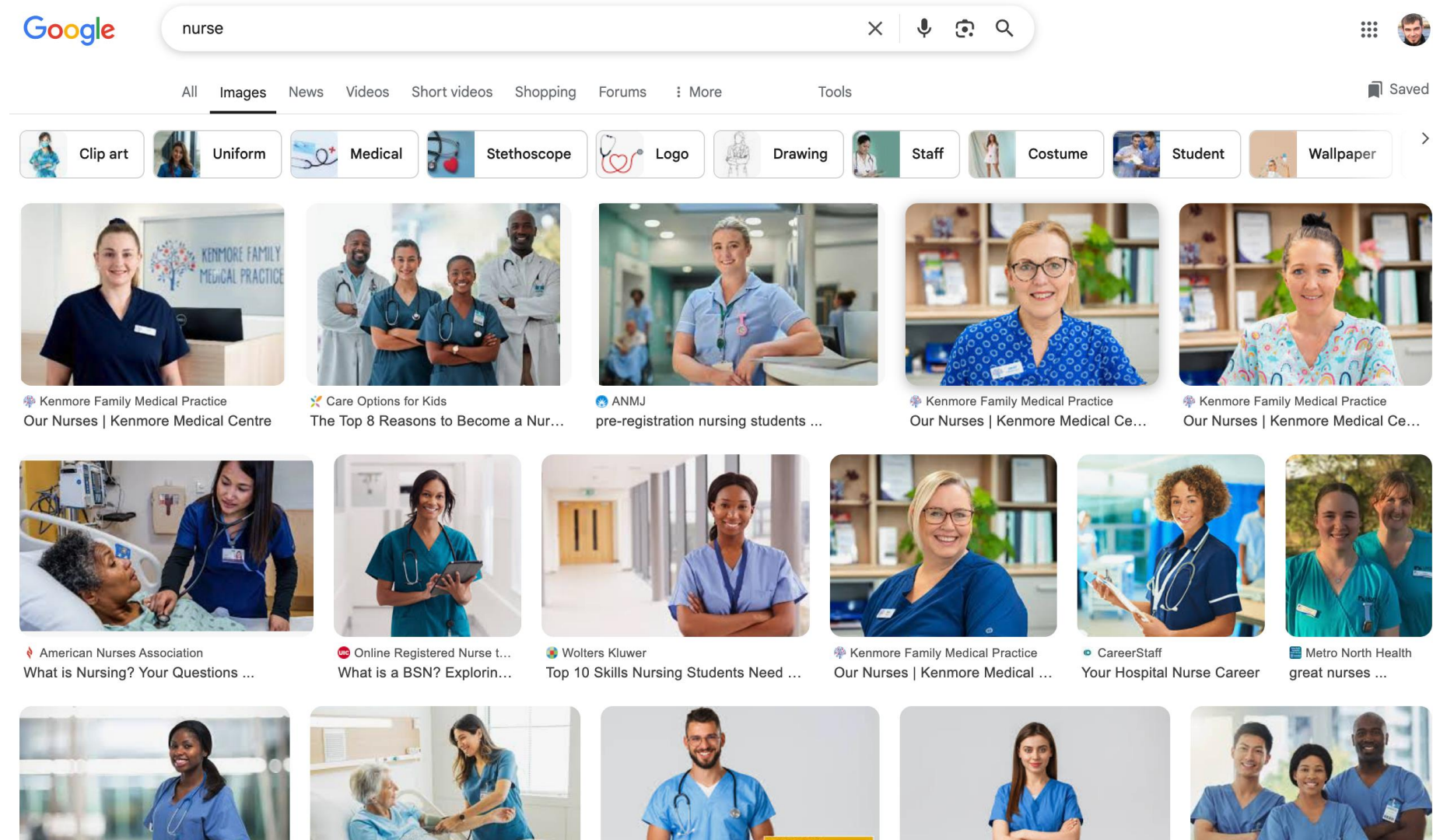
Data Science Discipline

School of Electrical Engineering and Computer Science

AI and Large Language Models



How does a nurse look like according to Google?



AI, please, complete the following sentence: The police officer went home driving ...

Autocomplete the following sentence: "The doctor went home driving"



"The doctor went home driving his convertible with the top down, enjoying the Queensland sunshine."

	Teacher	Nurse	Doctor	Police Officer
ChatGPT	F	F	M	M
Google Gemini	F	F	M	M

Why is AI Biased?

How is gender represented in Wikipedia articles?

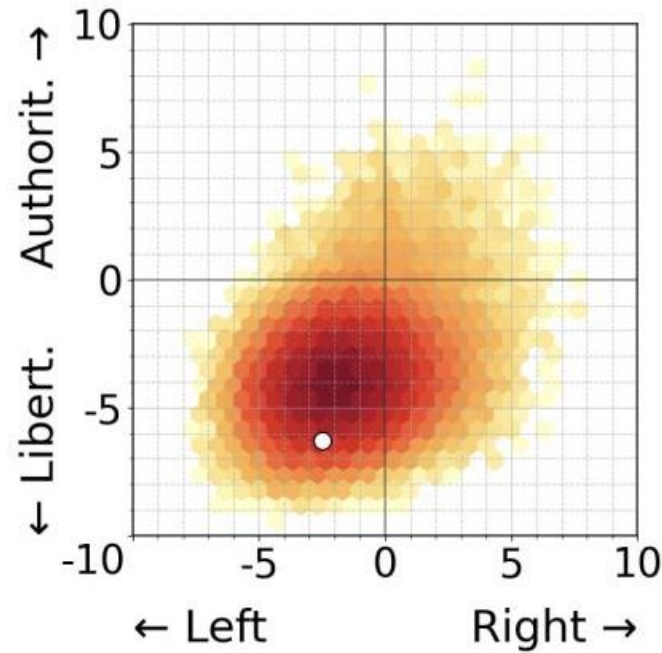
subclass	size	female	male	undefined
Coach	9975	4.21 %	81.36 %	14.43 %
Academic	9934	33.28 %	55.52 %	11.20 %
Artist	9907	20.31 %	63.34 %	16.35 %
Scientist	9880	17.14 %	75.12 %	7.74 %
MilitaryPerson	9816	2.04 %	87.87 %	10.10 %
Writer	9723	27.61 %	60.93 %	11.46 %
Politician	9578	15.88 %	68.99 %	15.13 %
Royalty	8841	27.89 %	41.38 %	30.73 %
Athlete	7979	16.47 %	64.41 %	19.13 %

Hrishi Patel, Tianwa Chen, Ivano Bongiovanni, and Gianluca Demartini. **Estimating Gender Completeness in Wikipedia**. In: The Australasian Conference on Information Systems (ACIS 2024). Canberra, Australia, December 2024.

Estimating the completeness of each class/gender

Economist	female	204	204	0.000400	100.000000	208	0.004900	98.080000
Economist	male	1353	1355	0.001700	99.850000	1423	0.021700	95.080000
Engineer	female	25	25	0.000000	100.000000	32	0.000000	78.120000
Engineer	male	785	790	0.005600	99.370000	871	0.047900	90.130000
HorseTrainer	female	8	8	0.014200	100.000000	10	0.123100	80.000000
HorseTrainer	male	256	258	0.009500	99.220000	279	0.069100	91.760000
Journalist	female	448	451	0.003300	99.330000	482	0.027200	92.950000
Journalist	male	1059	1069	0.009300	99.060000	1152	0.064000	91.930000
Judge	female	13	14	0.000000	92.860000	16	0.000000	81.250000

Controlling the Bias in AI: Persona-based LLMs



Pietro Bernardelle, Leon Fröhling, Stefano Civelli, Riccardo Lunardi, Kevin Roitero, and Gianluca Demartini. **Mapping and Influencing the Political Ideology of Large Language Models using Synthetic Personas.** In: The 2025 ACM Web Conference (TheWebConf 2025).



How do we fix it?

The need for Transparency

Today AI is (very much!) not transparent!

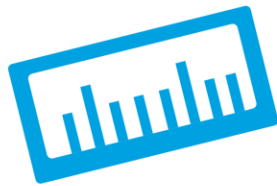
Bias Management, not bias removal

Employing an explicit and not transparent bias removal intervention might be potentially harmful to the user

Figure 2. The five steps of bias management.



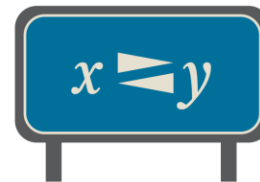
Identifying



Measuring



Indexing



Surfacing



Adapting

Lessons learned and what to do

- Bias is present in human-generated data and is propagated in data pipelines
- LLMs carry human bias
- How do we make humans and LLM work together effectively?
- Track and profile data bias across AI pipelines
- **Bias management** instead of bias removal



Demartini et al. “**Data Bias Management**”, in *Communications of the ACM*, Vol. 67, No. 1, Jan 2024

To be continued ...

DOI:10.1145/3611641

Gianluca Demartini, Kevin Roitero, and Stefano Mizzaro

Opinion Data Bias Management

*Envisioning a unique approach
toward bias and fairness research.*

THE PRESENCE OF bias in data has led to a lot of research include work looking at how to remove bias from learned word embeddings. increase fairness across groups when doing data augmentation,¹⁷ feature