BankGPT: the use of Large Language Models in official communications

CLAUDIA BIANCOTTI, CAROLINA CAMASSA, MARCO FRUZZETTI, **LUIGI PALUMBO**, MYRIAM PORTALURI

LLMs and official communications

The **use of LLMs is increasingly pervasive** in business and official settings.

Potential efficiency gains and lagging guidelines for official use.

Our research questions:

- Is there a **quality gap** between text generated by LLMs and human experts?
- Are **preferences** related to specific **characteristics of the audience**?

Some references

OECD institutions' staff

23% use GenAI
76% do not have guidelines
(Mitton, 2023)

LLMs for economic analysis

Financial sentiment analysis
(Fatouros et al., 2023)
Investment signals
(Fatouros et al., 2024)

LLMs and CB communication

Deciphering Fedspeak
(Hansen and Kazinnik , 2023)
RBA monetary policy statements
(Smales, 2024)

CB Language Models

Special-purpose text models
LLMs for CB workflows
(Gambacorta et al., 2024)

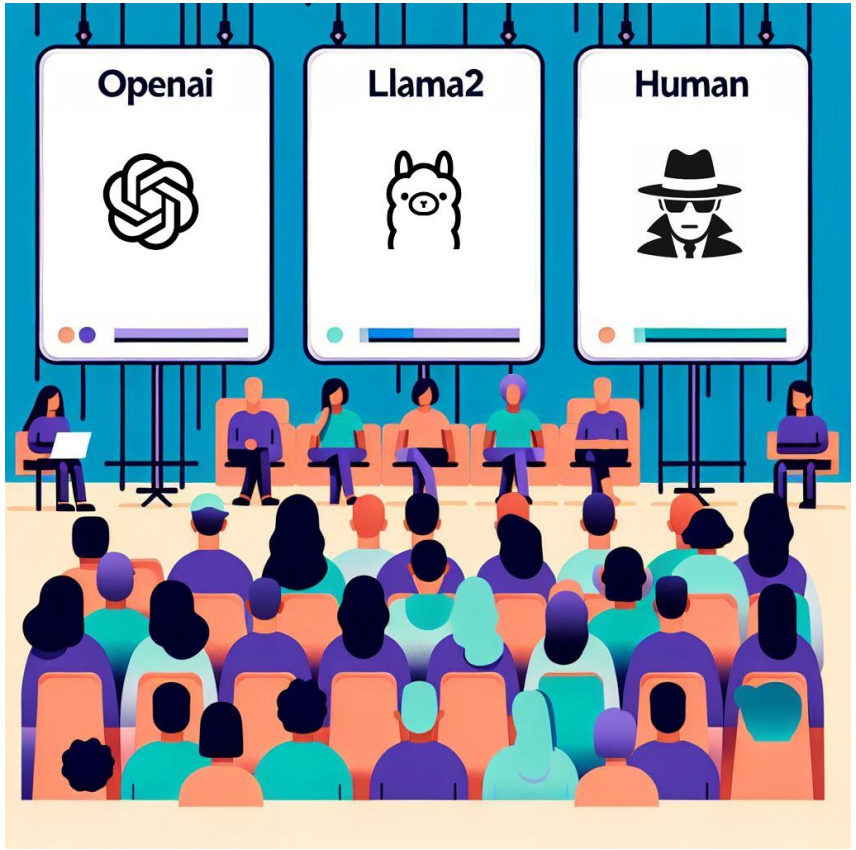
LLMs and summarization

LLMs on par with humans on news… (Zhang et al., 2024)
…but falling short on specific tasks (Lui et al., 2023)

GenAI vs. Human

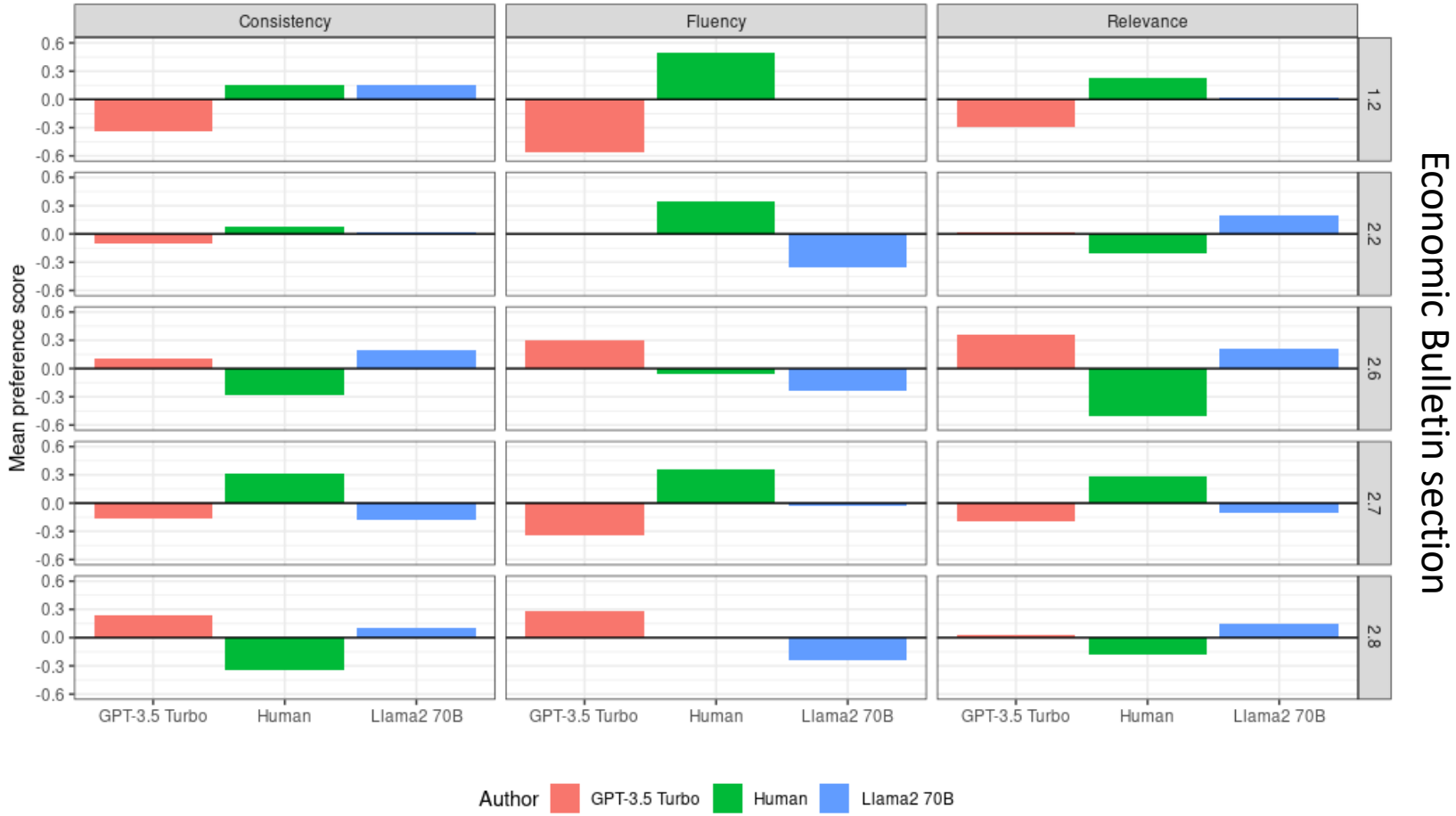
GenAI text better than human
(Porter and Machery, 2024)
Cannot distinguish, and bias
(Grassini and Koivisto, 2024)

Our experiment

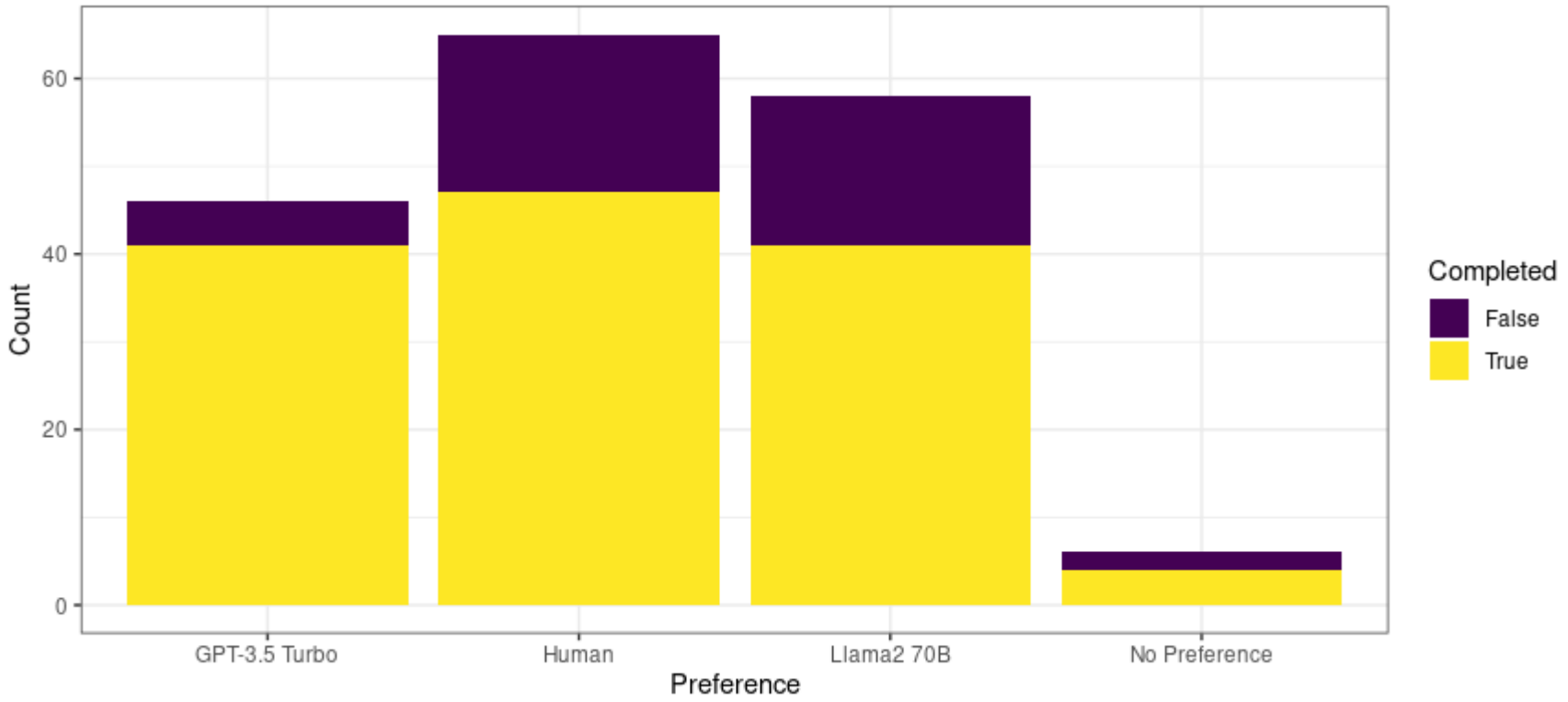


- Target: Bank of Italy's **Economics and Statistics Directorate General**
- Good participation, with **175** answers out of about 600 colleagues
- Pairwise comparison **of *Economic Bulletin* summaries**, all presented as LLM-generated
- Three evaluation metrics: **fluency, consistency and relevance** (Fabbri et al., 2021)
- Bradley-Terry model to convert comparisons into **comprehensive ranking of models** (Bradley and Terry, 1952)

Preferences are varied across sections and metrics



Human authors are (slightly) preferred



Age, gender, and education

	(1)		(2)	
Intercept	-2.064 **	*(0.638)*	-2.175 ***	*(0.528)*
Age: Over35	1.255 *	*(0.627)*	1.117 **	*(0.429)*
Gender: Male	-1.008 *	*(0.395)*	-0.968 *	*(0.389)*
Authorship: Yes	0.793 *	*(0.397)*	0.797 *	*(0.360)*
Motivation: Yes	0.952 *	*(0.414)*	0.858 *	*(0.352)*
Highest degree: PhD	1.119 *	*(0.447)*	1.052 **	*(0.396)*
English level: Advanced	-0.660	*(0.479)*		
Professional Experience: Senior	0.016	*(0.557)*		
Use of LLMs: Yes	0.247	*(0.428)*		
Readership: High	0.164	*(0.406)*		
Time	-0.000	*(0.000)*		
N. obs.	175		175	
AIC	218.931		211.088	

*** p < 0.001; ** p < 0.01; * p < 0.05.

Conclusions

- LLMs are capable of producing text of **comparable quality** to that produced by expert humans…
- …but stronger **domain expertise** steers readers' preferences towards human-generated text.
- Salience of **demographic factors** such as age and gender for preferences' prediction raise questions about **potential bias** in LLMs' training material and process.
- Implications for institutional communications: organizations should carefully **consider their audience composition** when determining optimal content generation strategies.

Q&A

THANKS!