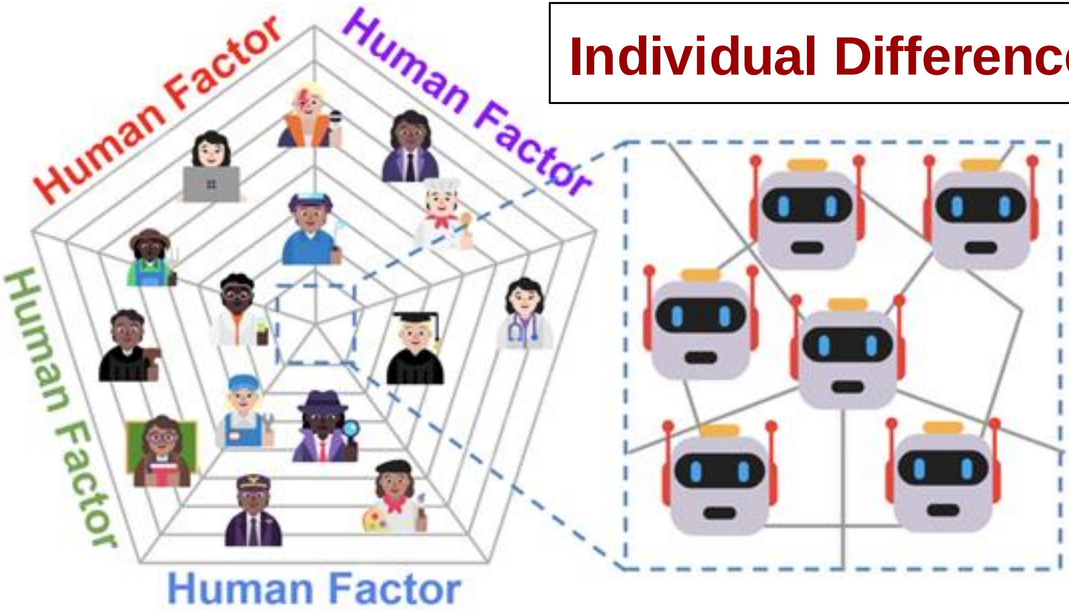
Human Language Analysis Beings

The Consistent Lack of Variance of Psychological Factors Expressed by LLMs and Spambots

Vasudha Varadarajan*,[1], Salvatore Giorgi*,[2], Siddharth Mangalik[1]
Nikita Soni[1], David M. Markowitz[3], H. Andrew Schwartz[1]
[1]Department of Computer Science, Stony Brook University
[2]Department of Computer and Information Science, University of Pennsylvania
[3]Department of Communication, Michigan State University
{vvaradarajan,has}@cs.stonybrook.edu, sgiorgi@sas.upenn.edu

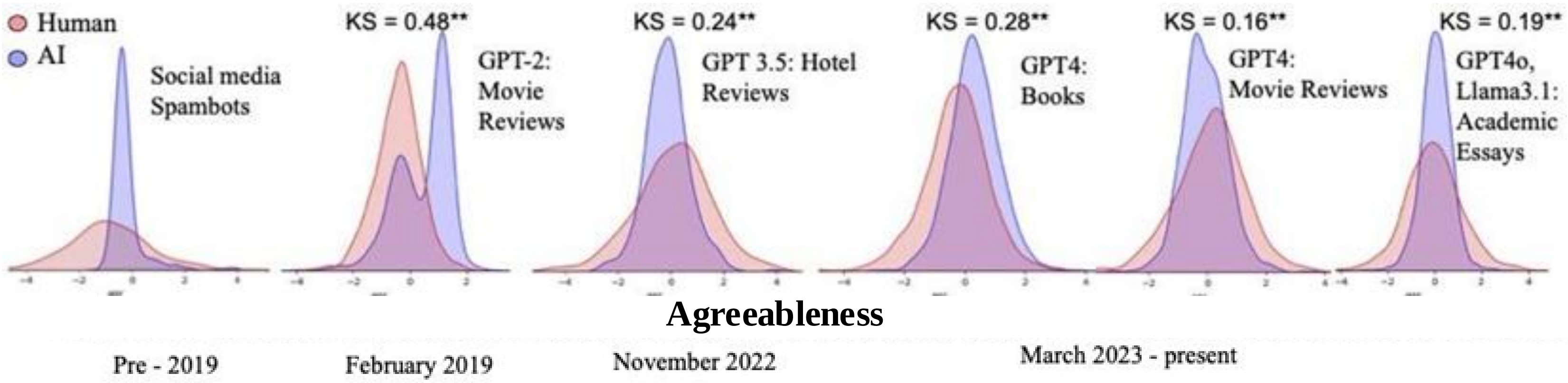
Observation (The Big Idea)

Individual Differences!



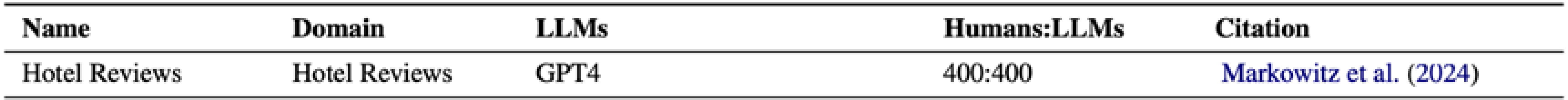
Machine-generated texts score average on multiple human factors at the same time.

Observation (The Big Idea)



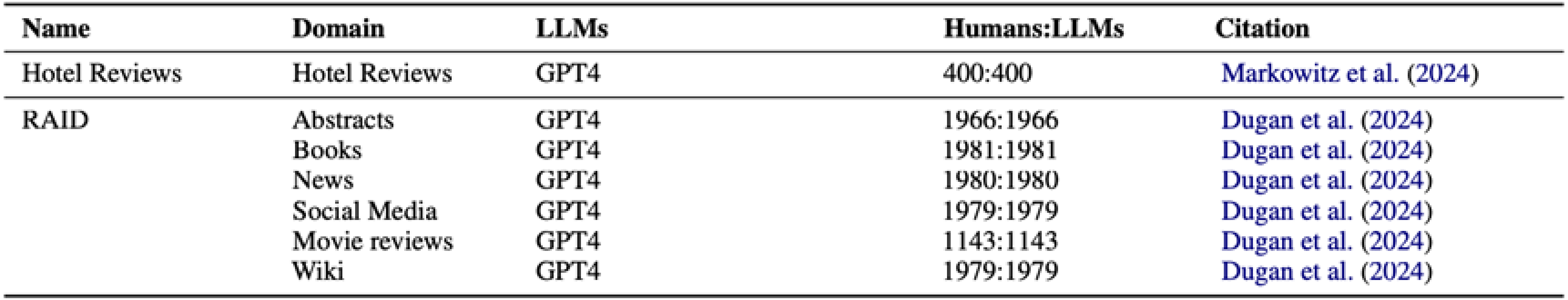
Over the years, from 2019 spambots to modern-day LLMs have exhibited this behavior.

Datasets



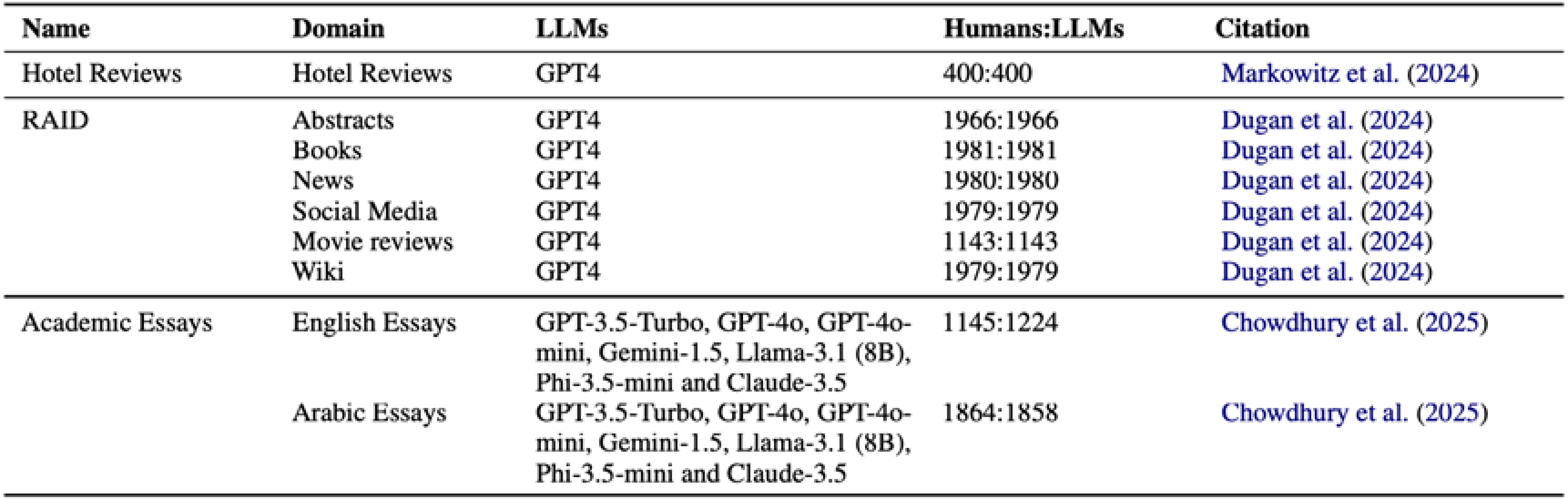
Name	Domain	LLMs	Humans:LLMs	Citation
Hotel Reviews	Hotel Reviews	GPT4	400:400	Markowitz et al. (2024)

Datasets



Name	Domain	LLMs	Humans:LLMs	Citation
Hotel Reviews	Hotel Reviews	GPT4	400:400	Markowitz et al. (2024)
RAID	Abstracts	GPT4	1966:1966	Dugan et al. (2024)
	Books	GPT4	1981:1981	Dugan et al. (2024)
	News	GPT4	1980:1980	Dugan et al. (2024)
	Social Media	GPT4	1979:1979	Dugan et al. (2024)
	Movie reviews	GPT4	1143:1143	Dugan et al. (2024)
	Wiki	GPT4	1979:1979	Dugan et al. (2024)

Datasets



Name	Domain	LLMs	Humans:LLMs	Citation
Hotel Reviews	Hotel Reviews	GPT4	400:400	Markowitz et al. (2024)
RAID	Abstracts	GPT4	1966:1966	Dugan et al. (2024)
	Books	GPT4	1981:1981	Dugan et al. (2024)
	News	GPT4	1980:1980	Dugan et al. (2024)
	Social Media	GPT4	1979:1979	Dugan et al. (2024)
	Movie reviews	GPT4	1143:1143	Dugan et al. (2024)
	Wiki	GPT4	1979:1979	Dugan et al. (2024)
Academic Essays	English Essays	GPT-3.5-Turbo, GPT-4o, GPT-4o-mini, Gemini-1.5, Llama-3.1 (8B), Phi-3.5-mini and Claude-3.5	1145:1224	Chowdhury et al. (2025)
	Arabic Essays	GPT-3.5-Turbo, GPT-4o, GPT-4o-mini, Gemini-1.5, Llama-3.1 (8B), Phi-3.5-mini and Claude-3.5	1864:1858	Chowdhury et al. (2025)

Human Factors: Demographics

- Age and Gender
- *Developing Age and Gender Predictive Lexica over Social Media.* (Sap et al., EMNLP 2014)

Human Factors: Personality (OCEAN)

- Personality based on Big-5
  - O: Openness
  - C: Conscientiousness
  - E: Extraversion
  - A: Agreeableness
  - N: Neuroticism
- *Automatic personality assessment through social media language.* (Park et al., JPSP 2015)

Human Factors: Empathy

- Trained on Interpersonal Reactivity Index and topics from Facebook statuses.
- *Characterizing empathy and compassion using computational linguistic analysis.* (Yaden et al., Emotion 2023)

Human Factors: Behavioral Linguistic Traits

- Based on unprompted social media language use.
- Five linguistic traits derived from factor analyzing ngrams.
- *Latent human traits in the language of social media: An Open Vocabulary Approach.* (Kulkarni et al., PloS One 2018)

Individual human factors

Kolmogorov-Smirnoff test

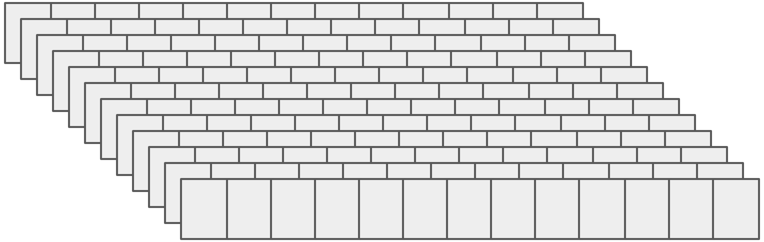
RAID	Personality					Empathy	Behavioral Linguistic Traits					Demographics	
Domains	Ope	Con	Ext	Agr	Emo		F1	F2	F3	F4	F5	Age	Gender
Abstracts	0.18	0.13	0.05	0.06	0.06	0.22	0.05	0.07	0.18	**0.31**	0.25	0.05	0.29
Books	**0.31**	0.1	0.09	0.26	0.18	0.11	0.07	0.05	**0.55**	**0.31**	0.2	0.07	0.16
News	**0.34**	0.05	0.04	0.13	0.07	0.11	0.05	0.08	**0.48**	0.22	0.09	0.16	0.18
Reddit	**0.36**	0.13	0.14	0.13	0.09	0.13	0.16	0.07	**0.48**	**0.31**	0.06	0.1	0.25
Reviews	**0.42**	0.22	0.2	0.13	0.15	0.17	0.08	0.04	**0.5**	**0.55**	0.28	0.13	0.41
Wiki	**0.30**	0.08	0.05	0.12	0.03	0.07	0.09	0.09	**0.41**	0.11	0.13	0.15	0.12

13 Human Factors

Unsupervised Human vs AI classification

13-D Human Factors vector

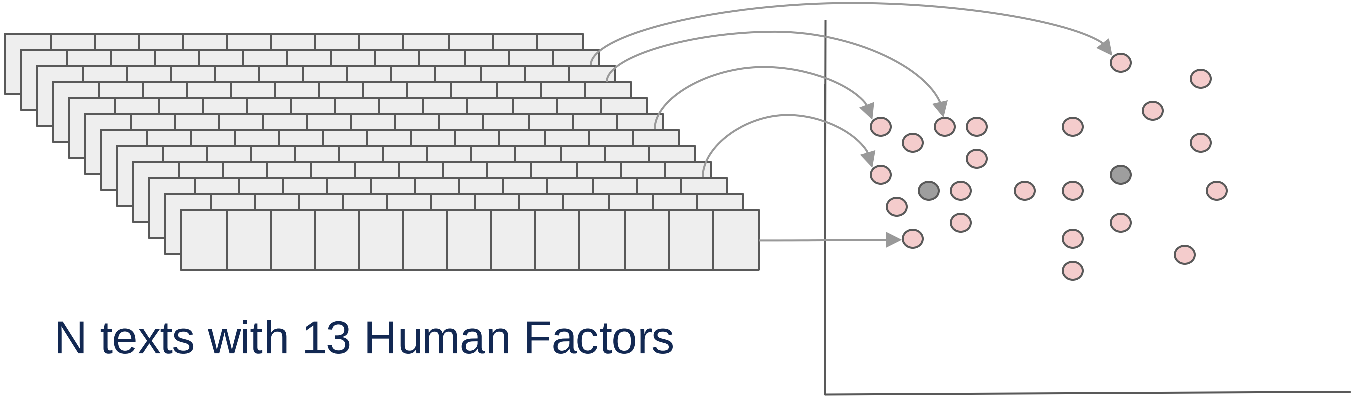
Unsupervised Human vs AI classification



N texts with 13 Human Factors

Unsupervised Human vs AI classification

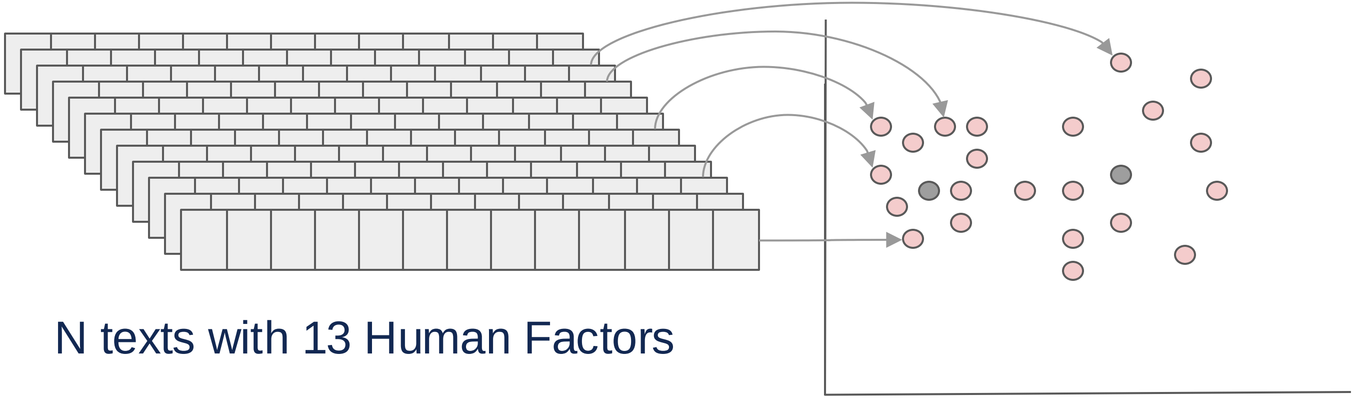
13-D → 2-D (dimensionality reduction)



N texts with 13 Human Factors

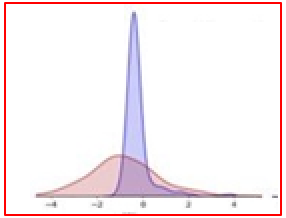
Unsupervised Human vs AI classification

13-D → 2-D (dimensionality reduction)



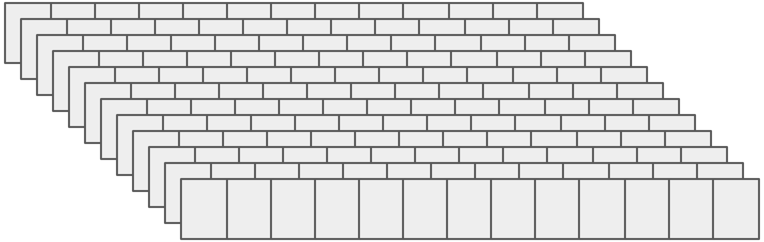
N texts with 13 Human Factors

Spectral reduction with radial basis kernel

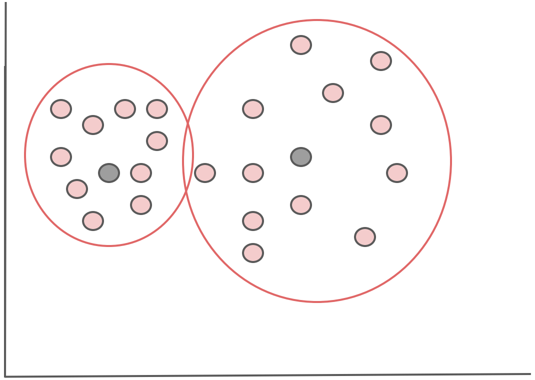


Unsupervised Human vs AI classification

Clustering into two clusters

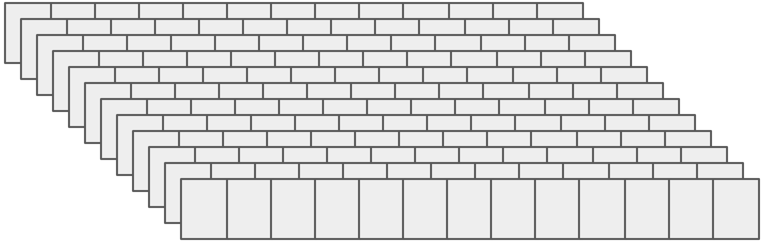


N texts with 13 Human Factors

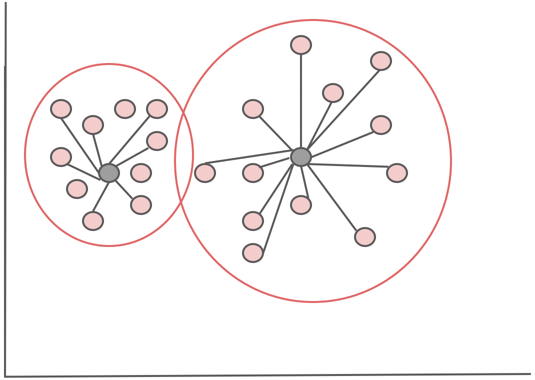


Unsupervised Human vs AI classification

Calculate intracluster spread

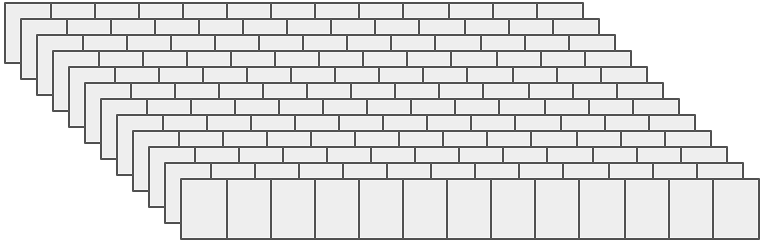


N texts with 13 Human Factors

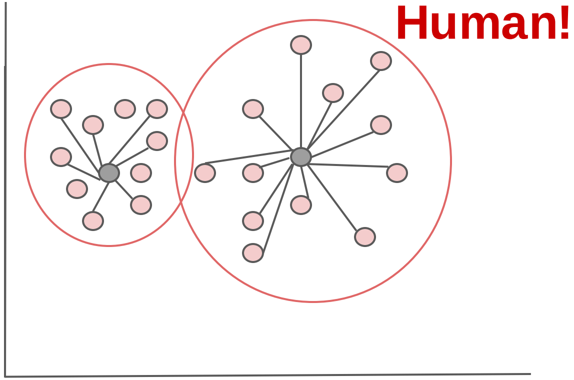


Unsupervised Human vs AI classification

Calculate intracluster spread



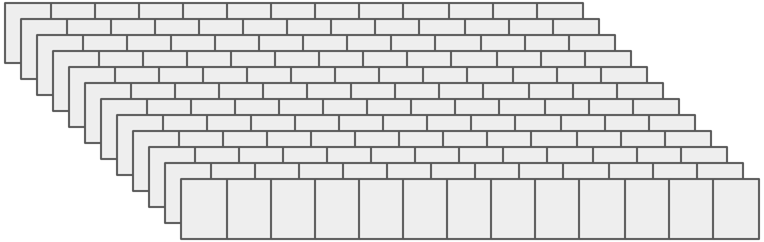
N texts with 13 Human Factors



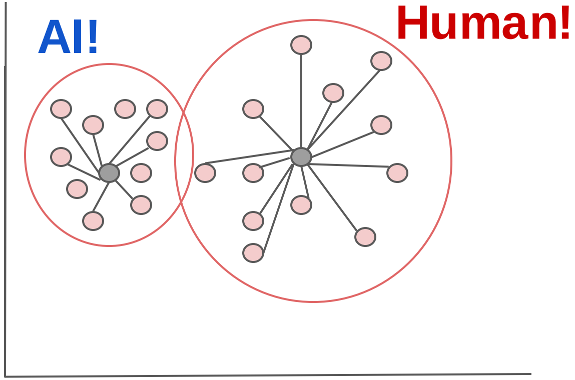
Higher spread: human

Unsupervised Human vs AI classification

Calculate intracluster spread

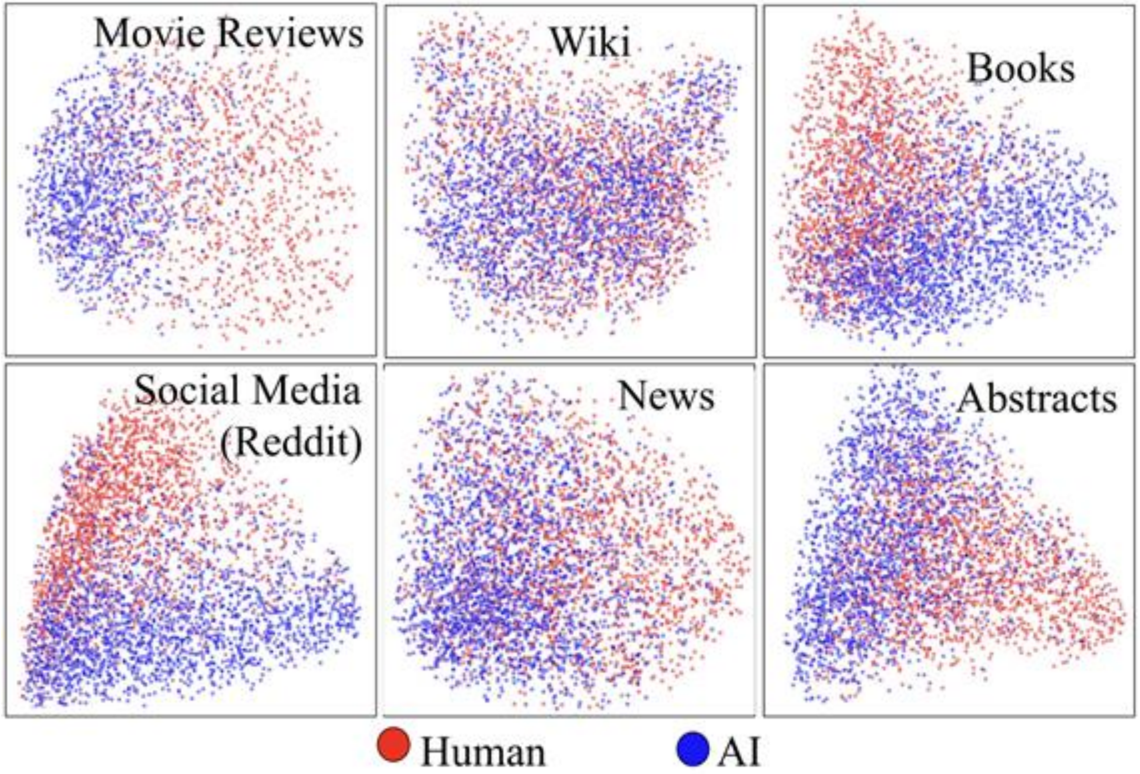


N texts with 13 Human Factors

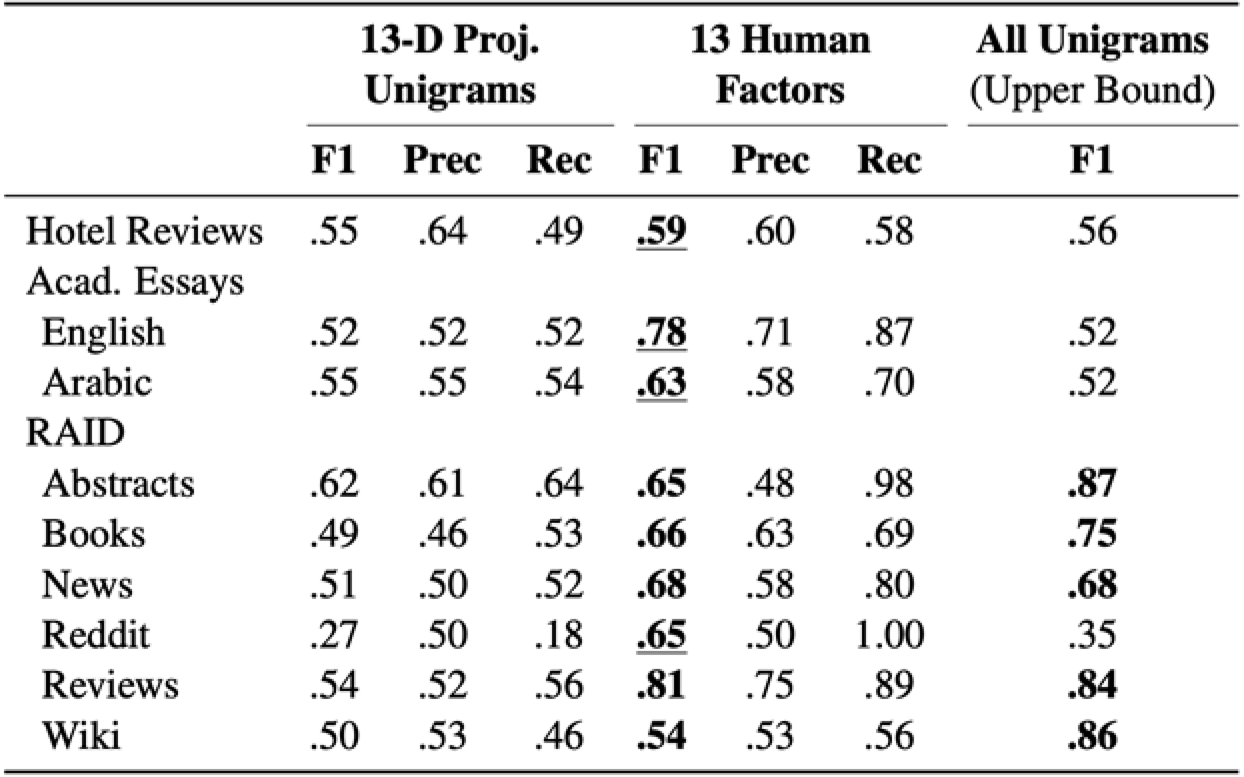


Higher spread: human; lower spread: AI

Results: Clusters

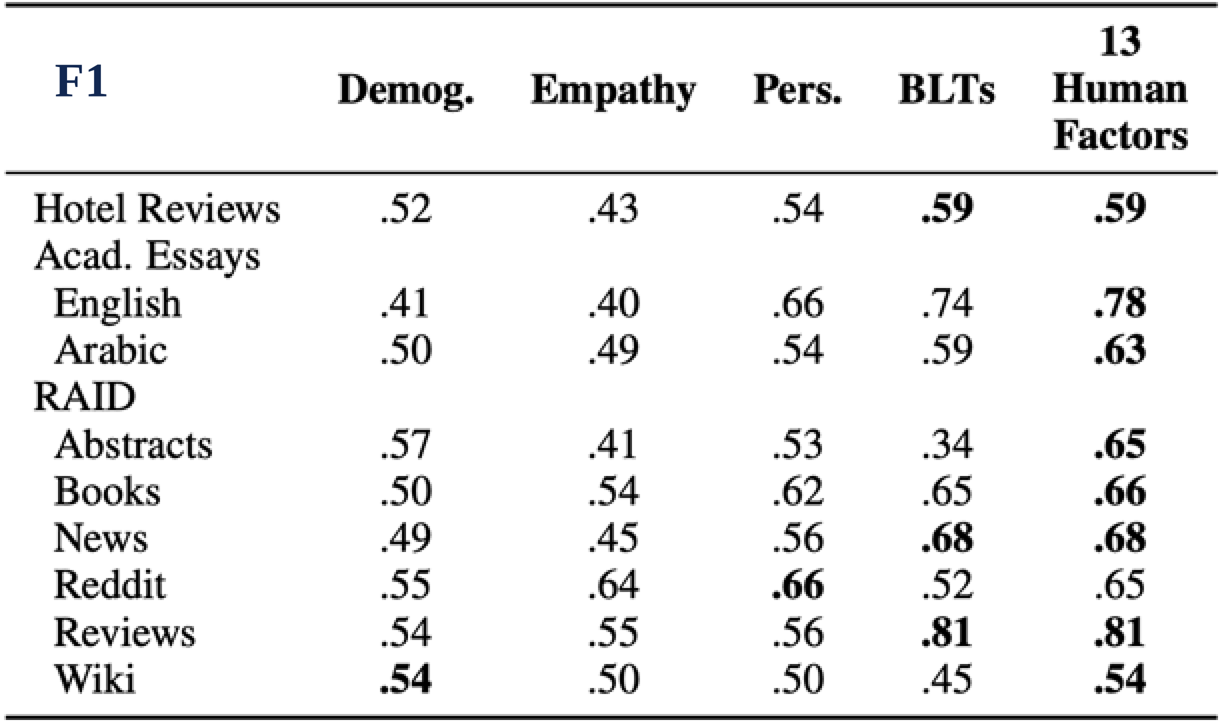


Results: Classification



	13-D Proj. Unigrams			13 Human Factors			All Unigrams (Upper Bound)
	F1	Prec	Rec	F1	Prec	Rec	F1
Hotel Reviews	.55	.64	.49	**.59**	.60	.58	.56
Acad. Essays							
English	.52	.52	.52	**.78**	.71	.87	.52
Arabic	.55	.55	.54	**.63**	.58	.70	.52
RAID							
Abstracts	.62	.61	.64	**.65**	.48	.98	**.87**
Books	.49	.46	.53	**.66**	.63	.69	**.75**
News	.51	.50	.52	**.68**	.58	.80	**.68**
Reddit	.27	.50	.18	**.65**	.50	1.00	.35
Reviews	.54	.52	.56	**.81**	.75	.89	**.84**
Wiki	.50	.53	.46	**.54**	.53	.56	**.86**

Results: Classification



F1	Demog.	Empathy	Pers.	BLTs	13 Human Factors
Hotel Reviews	.52	.43	.54	**.59**	**.59**
Acad. Essays					
English	.41	.40	.66	.74	**.78**
Arabic	.50	.49	.54	.59	**.63**
RAID					
Abstracts	.57	.41	.53	.34	**.65**
Books	.50	.54	.62	.65	**.66**
News	.49	.45	.56	**.68**	**.68**
Reddit	.55	.64	**.66**	.52	.65
Reviews	.54	.55	.56	**.81**	**.81**
Wiki	**.54**	.50	.50	.45	**.54**

Conclusions

- **Consistency** of psychological factors observed across LLMs.
- Human factors for AI-text takes on an average value, which is **atypical for humans**.
- We leverage this property to distinguish human and AI-generated texts in a **completely unsupervised** fashion.

Thank you!

Paper:

SCAN ME

Contact for questions/collaboration:

Vasudha Varadarajan:
vvaradarajan@cs.stonybrook.edu

Sal Giorgi:
sgiorgi@sas.upenn.edu