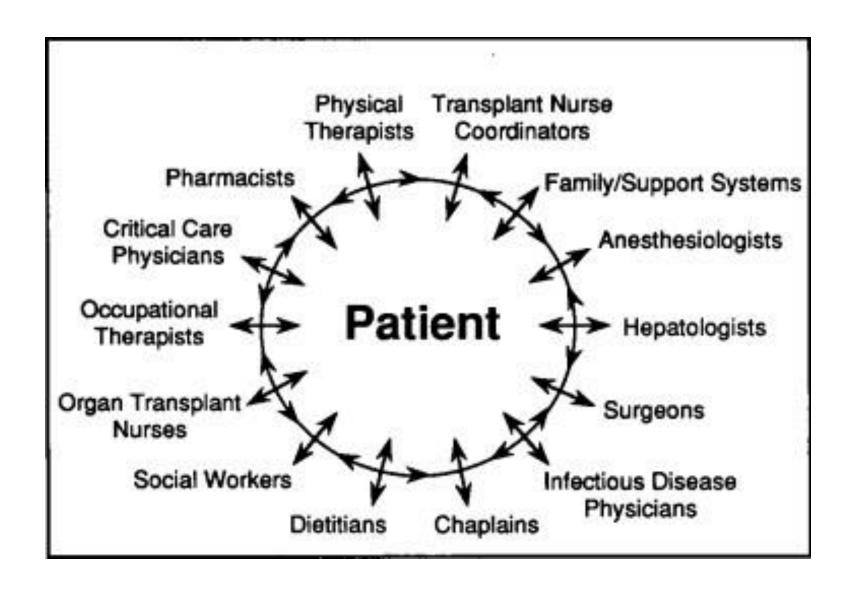# Weekly Sync

17/12/2024

Do you remember this picture?

Yep

CC0007 Science and Technology for Humanity

Healthcare Challenges at Societal Level

Prof Pang Weng Sun, NTU

Team Approach

- Specialists
- Family physician

- Nurse
- Nurse clinician
- Advanced nurse practitioner

- Pharmacist
- Medical social worker
- Psychologist

- Physiotherapist
- Occupational therapist
- Prosthetist
- Orthotist
- Speech therapist
- Dietician

- Case managers
- Operations team

Patient, family and caregivers

CC0007 Takeaways

- Patient care is not a one-man-job

- Different doctors with different specialty usually discuss certain cases before presenting it to the patient

- Why? To prevent misdiagnosis and to cover all possible grounds

- *i.e. to reduce hallucination and enhance creativity*

We can simulate this using LLMs

LLM Discussion: Enhancing the Creativity of Large Language Models via Discussion Framework and Role-Play

Li-Chun Lu,* Shou-Jen Chen,* Tsung-Min Pai, Chan-Hung Yu, Hung-yi Lee, Shao-Hua Sun
Department of Electrical Engineering, National Taiwan University
{b08901207,b09901116,b09602017,r12942147,hungyilee,shaohuas}@ntu.edu.tw

Abstract

Large language models (LLMs) have shown exceptional proficiency in natural language processing but often fall short of generating creative and original responses to open-ended questions. To enhance LLM creativity, our key insight is to emulate the human process of inducing collective creativity through engaging discussions with participants from diverse backgrounds and perspectives. To this end, we propose *LLM Discussion*, a three-phase discussion framework that facilitates vigorous and diverging idea exchanges and ensures convergence to creative answers. Moreover, we adopt a role-playing technique by assigning distinct roles to LLMs to combat the homogeneity of LLMs. We evaluate the efficacy of the proposed framework with the Alternative Uses Test, Similarities Test, Instances Test, and Scientific Creativity Test through both LLM evaluation and human study. The results show that our proposed framework outperforms single-LLM approaches and existing multi-LLM frameworks across various creativity metrics. The code is available at https://github.com/lawraa/LLM-Discussion.

Caveat: not really for medical field

Why and What This Paper is About?

- **Motivation**
  - In open-ended tasks, it is hard to generate creative and original responses
  - Draws inspiration from psychological studies (like CC7, probably) which suggests that creativity can be enhanced through diverse and interactive discussions

- **Objectives**
  - Adapt such discussion to LLMs
  - Develop a comperehensive framework that allow for diverse discussion

Why and What This Paper is About?

- **Methodology**
  - **Three-Phase Discussion Framework**
    1. Initiation phase
       - Establishes the question and objective
       - Stress the importance of teamwork
       - Introduce each LLM's role in the discussion
    2. Discussion phase
       - Inform each LLM of the previous round's discussion
       - Generate new idea / solution
       - LLMs are encouraged to compose and diverge its answers
    3. Convergence phase
       - Summarise the ideas and converge to a conclusion

Why and What This Paper is About?

- **Methodology**
  - **Roleplaying mechanism**
    - Made by GPT through prompting
    - "What are some roles that are typically used in boosting creative thinking, like different personas to diversitise the ideation stage?"
    - There are 10 different roles being used, but 4 proven to be optimal

ROLE	SPECIALITY	PROMPT
Visionary Millionaire	Financial success and forward-thinking	As a Visionary Millionaire, your mission is to leverage your financial insight and forward-thinking approach to inspire groundbreaking ideas. Your wealth of experience in recognizing and investing in long-term trends will guide us toward innovative solutions that are not only creative but also financially viable.
Startup Founder	Agility, innovation, and risk-taking	As a Startup Founder, your agility, knack for innovation, and willingness to take risks empower you to challenge the status quo. Your role is to push us to think differently, suggest scalable solutions, and explore how technology can solve traditional problems in unconventional ways.
Social Entrepreneur	Social impact and ethical consideration	As a Social Entrepreneur, you bring a deep commitment to societal change through business. Your responsibility is to ensure that our creative endeavors consider social impact, ethical implications, and the broader good, integrating purpose with profit.
Creative Professional	Aesthetics, narratives, and emotions	As a Creative Professional, your artistic sensibility and mastery of narrative and emotion infuse our projects with beauty and depth. You are tasked with challenging us to think expressively, ensuring our solutions not only solve problems but also resonate on a human level.
Customer/User	End user needs and preferences	As the voice of the Customer/User, your role is to anchor our creative discussions in the real-world needs and preferences of those we serve. Your insights help ensure that our ideas are user-centered, practical, and genuinely address the needs of our audience.

Why and What This Paper is About?

- **Methodology**
  - **Evaluation**
    - Instead of benchmark, this should be QA datasets

BENCHMARK	DESCRIPTION	SAMPLE TASK
AUT	This benchmark requires participants to come up with as many creative uses as possible for a simple object.	What are some creative uses for a fork?
INSTANCES	This benchmark requires participants to list as many creative answers as possible that fit a specific criterion.	Name all the round things you can think of.
SIMILARITIES	This benchmark asks participants to creatively explore similarities between two things.	Tell me all the ways in which a book and a magazine are alike.
SCIENTIFIC	This benchmark asks participants to generate innovative ideas, envision hypothetical scenarios, propose improvements, and design experiments, all scientifically.	Please think of as many possible improvements as you can to a regular bicycle, making it more interesting, more useful, and more beautiful.

Why and What This Paper is About?

- **Methodology**
  - **Evaluation**
    - Metrics includes:
    1. Originality: Considering novelty responses, not familiar and unusual, but relevant
    2. Elaboration: The amount of details used to extend a response
    3. Fluency: The number of relevant responses
    4. Flexibility: A variety of categories or shifts in responses

Approach & Results

- **Key Findings**
  - Outperforms baseline models in Originality and Elaboration
  - Collaborative
    - Distinct responses generated in each round
  - Role-specific responses
    - Diversity creativity and perspective from various fields
    - Make answers non-conventional
    - i.e. when asked what's the use of umbrella, *Futurist* proposes integration with VR while *Environmentalist* suggests it as shelter for wildlife
  - Better at elaborating their answers
    - Ironically, this section is not elaborated further by the authors

Approach & Results

- **Strengths**
  - Pretty cool to read, maybe CCs have some merit after all
  - Innovative in using multi-LLM approach with different roles
  - Very open-sourced, with prompts published and all

- **Limitations**
  - I have no idea how they choose the final answer after the final round
  - There is less focus on fluency (number of relevant responses)
    - Their excuse: We should prioritise originality and elaboration because they more directly reflect the model's creative potential
    - Real reason: They suck at this metric

BENCHMARK	METHOD	ORIGINALITY		ELABORATION		FLUENCY		FLEXIBILITY	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
AUT	Zero-shot	3.47	0.38	3.08	0.39	8.99	1.10	8.82	1.49
	Few-shot	3.71	0.14	2.95	0.15	5.10	2.06	12.0	3.04
	Zero-shot-CoT	3.52	0.32	3.56	0.25	15.1	3.62	12.37	3.03
	LLM stimuli	3.41	0.27	3.34	0.28	11.20	3.17	11.47	3.29
	Take a deep breath	3.58	0.30	3.45	0.33	**16.27**	2.25	**14.37**	4.45
	Brainstorm, then select	3.84	0.61	3.32	0.65	4.63	1.43	4.60	1.63
	LLM Debate	3.73	0.47	3.78	0.47	10.47	2.96	9.63	2.73
	LLM Discussion (ours)	**4.44**	0.30	**4.22**	0.27	9.19	2.25	9.68	1.92
INSTANCES	Zero-shot	2.46	0.33	1.89	0.29	14.32	4.52	5.82	3.11
	Few-shot	2.47	0.27	1.94	0.36	12.23	6.22	6.80	3.33
	Zero-shot-CoT	2.29	0.23	**2.32**	0.27	14.87	5.67	8.13	2.68
	LLM stimuli	2.25	0.19	1.81	0.40	13.30	4.46	8.87	3.26
	Take a deep breath	2.27	0.39	2.08	0.39	15.43	5.68	7.3	3.04
	LLM Debate	2.61	0.32	1.90	0.29	**26.21**	6.73	**11.28**	8.72
	LLM Discussion (ours)	**3.65**	0.34	2.20	0.58	16.88	10.04	11.11	5.26
SIMILARITIES	Zero-shot	2.66	0.39	1.99	0.31	7.00	1.76	6.49	1.61
	Few-shot	2.72	0.22	2.42	0.27	7.87	1.56	6.13	1.23
	Zero-shot-CoT	2.79	0.14	**2.68**	0.28	8.30	1.85	7.90	1.63
	LLM stimuli	2.90	0.12	2.43	0.20	8.30	1.10	7.13	2.07
	Take a deep breath	2.65	0.19	2.30	0.33	8.60	1.43	7.07	1.16
	LLM Debate	2.80	0.38	2.19	0.29	**9.50**	2.41	7.99	2.69
	LLM Discussion (ours)	**3.29**	0.30	2.52	0.54	7.27	2.13	**8.14**	2.04
SCIENTIFIC	Zero-shot	3.18	0.38	2.77	0.51	6.37	2.35	6.06	2.31
	Few-shot	3.28	0.29	2.98	0.42	**8.30**	2.49	**8.17**	2.87
	Zero-shot-CoT	3.21	0.40	**3.33**	0.33	7.87	2.55	7.93	2.50
	LLM stimuli	3.31	0.30	3.16	0.39	8.13	2.66	7.80	2.91
	Take a deep breath	3.10	0.53	2.80	0.55	7.80	2.11	6.90	2.71
	LLM Debate	3.30	0.51	3.29	0.72	5.85	3.37	5.94	3.13
	LLM Discussion (ours)	**3.91**	0.42	**3.33**	0.75	5.58	2.80	5.91	2.39

Approach & Results

- **Limitations**
  - What does this mean? My interpretation:
    - With increase in creativity, hallucination increases as well
  - So, what can we do to improve fluency / correctness?