Innovating AI for humanitarian action in emergencies

Anonymous Submission

Abstract

While the research community is making significant leaps in Generative AI, millions of children and women are facing humanitarian crisis. UNICEF reports millions of children are experiencing a high burden of unaddressed grief, loss, trauma due to recent war in Middle East and Ukraine-Russia. WHO identifies the challenge of providing timely psychological first aid (PFA) for victims of war as a humanitarian response. Due to limited access to professional counselors inside war torn regions, and the mounting psychological trauma of war on children, the need for research to creatively apply AI to support vulnerable people in emergencies with Humanitarian Assistance. The challenge is to create an AI that can offer personalized counselling for previously unseen scenarios, given people have experienced different types of scenarios of violence. To create LLM powered AI copilots that can be trusted, the paper presents a novel design for explain-ability to assist UNRWA staff to offer therapy at refugee camps.

WHO reports 1 in every 5 people exposed to war are impacted by mental health disorder. This is the 2nd challenge we addressed. To provide timely mental health intervention for large populations in inaccessible war zones such Gaza, inspite of occasional internet availability in warzones, we developed an approach to finetune a GPT-2 model personalized for every victim's situation. This opens up the opportunity to create & deploy personalized GPT-2 models on smartphone of millions of people to enable personalized healing.

The demos of both the solutions including source code are at presented at https://sites.google.com/view/hadr-ai.

Motivation & Challenges

As AI research is progressing at a significant pace, the opportunity to apply the new advancements to address the critical pressing challenges faced by millions of children is explored in this work. Given millions of people are located in war zones in the recent months in Gaza and Ukraine, the challenge is in providing timely aid in war zones for a sizable population, as war zones are inaccessible. High burden of unaddressed grief, loss, trauma and depression was visible when talking with refugees, reports UNHCR. "A school student saw his friend decapitated with in own eyes" – reports an UN official at UNICEF's Humanitarian action for children 2023 [1]. In this context, there is an urgent imperative to explore the design and develop humanity centered trustworthy AI systems to create positive impact on society.

The applications of GPT-x in warzones needs research and attention, given the grave long term humanitarian consequence of wars. As children have been deeply traumatized as we read this paper in parts of the world (Ukraine & Gaza), WHO identifies the challenge of timely psychological first aid (PFA) within humanitarian response. WHO's Mental Health Gap Action Programme Humanitarian Intervention Guide (mhGAP-HIG) calls for providing psychological first aid (PFA) [7] services within the humanitarian response. Humanitarian outreach & community members are being trained in PFA as per WHO/UNHCR. There is a need for access to AI tools that support humanitarian outreach members to provide PFA counselling for victims of violence (gender violence, rape, war, loss & grief).

Challenges: Two challenges at the intersection of AI community and humanitarian causes are

- ***First aid at refugee camps*** **(LLM copilot solution with Explain-ability):** As millions of children are traumatized by war, how to creatively apply LLMs to create an AI-copilot to enable humanitarian volunteers (UNICEF, UNRWA, Redcross) to provide PFA to victims of different forms of violence. As victims are subjected to victim-blaming, timely assertion of human rights is important. Given the lack of expert counselors in warzones, the need to research and develop AI copilots that can assist UNRWA staff to offer personalized counselling to refugees at warzones such as Gaza. Further research is required to develop Explain-ability of LLM solutions. Since the AI copilot provides recommendations, such solution needs to offer explanation to humans, so that the human can understand why the AI made that recommendation.
- ***Self-help for victims*** **(Auto GPT-2 during occasionally connected internet):** As 1 in 5 people exposed to war develop mental health disorder, there is an urgent need to research and develop self-help mental health therapy for victims of violence, even when they are located in inaccessible warzones. Due to poor internet connectivity, the solution needs to run in offline mode on personal smartphones. Given everyone needs a different type of emotional support, a personalized GPT2 emotional caregiver needs requires research.

Questions: The challenges for AI research are:-

1) *Can the AI handle unseen scenarios?. Is it possible to design an AI solution that can understand & respond to new forms of violence?* As different people have experienced different scenarios of violence, can the LLM **generalize to new unseen** scenarios.

2) *Is it possible to enable **explainability of AI copilots**:* End-users need to understand why & how the AI is making a certain recommendation. Is it possible to design an LLM AI copilot that explains its recommendations in crucial topics such as cognitive triad and therapy?.

3) *Can the AI solution work inspite of **occasional internet connectivity** in warzones?* Given warzones are inaccessible, there is a need to innovate AI deployment approaches.

4) *Is it possible to scale the deployment:* Given millions of families need aid, is it possible to scale to millions of families awaiting timely intervention.

5) ***Trustable interaction with AI** on sensitive topics & Privacy:* Sensitive topics such as rape may get discussed in counselling sessions. Hence the need for on-device ML inference without internet connectivity to ensure confidential and **trust-worthy** disclosure of personal experience by vulnerable victims.

6) ***Affordability for inclusion** of all sections of society:* The AI should run on mainstream commodity smartphones. While the solution should support counselling to even new unseen violence, the model should be small enough to fit inside a low cost smartphone with on-device ML.

7) *Can an AI model be **finetuned at warzone** by non-experts to provide personalized care for every victim?* Given every one has be counselled differently based on their experience and offered personalized self-help Cognitive Behavior Therapy (CBT) self-therapy, Can custom GPT-2 model be created automatically for every individual's scenario & deployed on every smartphone. Is it possible to generate datasets and fine-tune a model by non-experts? We explore automatic generation of custom GPT-2.

Approaches to design & development of AI for humanitarian action in emergencies & conflicts

1. Design & development of LLM copilot solution that is capable of responding to new scenarios of violence: (How to design to enable the ability of AI solution to generalize to previously unseen data)*:*

To address the above mentioned challenge #1 on **ability of AI solution to generalize to previously unseen data**, we designed and implemented a chain of LLMs to both provide timely PFA (*physiological first-aid)* and human right assertion that is capable of counselling any new unseen scenarios of violence. A working demonstration of the LLM based PFA for use at refugee camps is demonstrated in video at the project website at https://sites.google.com/view/hadr-ai.

The design approach is to orchestrate a chain of LLM (Large Language Models) by chaining a set of LLM calls. By having **a LLM chain create a higher level of abstraction** of the new form of violence, it is possible to develop a copilot solution that effectively address any previously unseen data or unseen forms of violence. By asking a LLM to generate an higher level abstraction of the violence, we navigate the solution to handle a new form of violence. (Refer Fig 1). The key idea is think at a higher level of abstraction of the concept to enable the LLM to generalize to previously unseen data. The demo video in the website demonstrates how the AI copilot can generalize to new forms of violence.

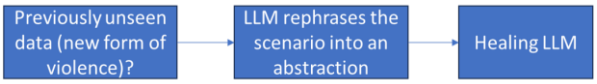

Figure 1: Design of LLM chain to think at a higher level of abstraction to improve the power of generalization of the AI solution to enable handling of previously unseen forms of violence

2. Design of AI copilot that explains the recommendation: (How to design AI copilot with Explain-ability)*:*

To address the challenge #2 mentioned above, the design should enable explain-ability of the suggestions made by the AI copilot. We propose creating a chain of LLMs to offer explain-ability of the thought process. This is illustrated in Fig 2. By orchestration of a chain of LLMs, a set of 6 different LLM are linked to navigate the thought process of the individual.

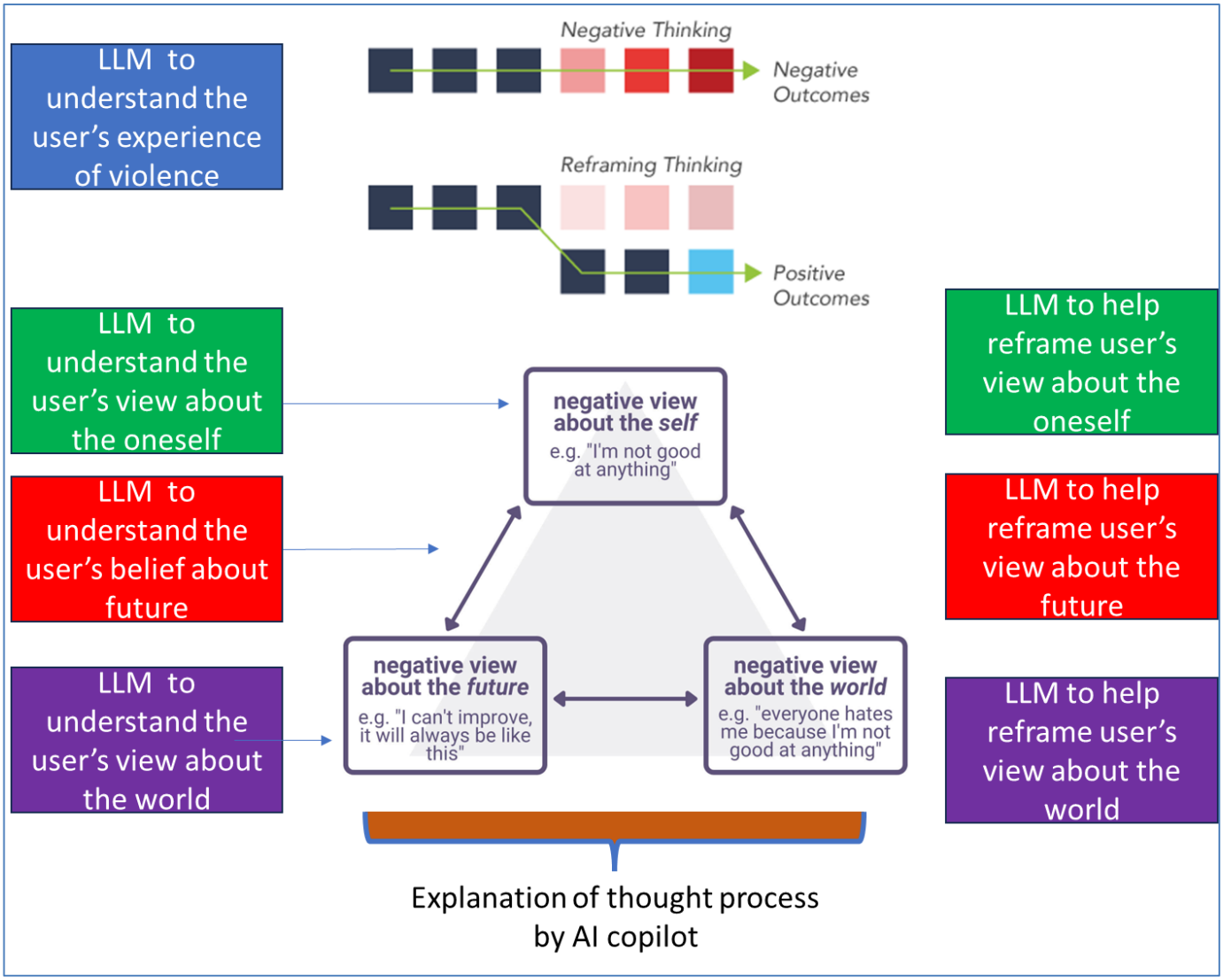

Figure 2: Explain-ability of AI copilot

As illustrated in the Fig 2, the AI pilots is designed with a set of 6 LLMs to create a Cognitive Behavior Therapy solution. The output of set of LLMs provides detailed explaination about the AI copilot is thinking about solving the problem for humanity. Based on the science of Cognitive Behaviour Therapy (CBT), the victim of violence often develop a negative view of others, and a negative view of the world, and a negative view of self. A set of 3 LLMs understand this viewpoint by chatting with the individual. Next, another set of 3 LLMs rephrase the viewpoint to help the individual to think more positively about the future, and think more positively about others, and help develop a positive outlook about the future and others.

The key idea to building explainable AI copilots is to create a chain of LLMs (multiple LLMs switched together), where the thought process of LLMs is displayed to the human. Thus an UNICEF staff can understand how the AI copilot is decoding the victims view (about self, world, future) and then help look at the situation from a different point of view. By orchestrating the set of LLMs to represent the cognitive thinking process, we develop a AI copilot that explains the recommendations. Thus we design and develop explainability of AI copilots.

3. Design of Auto GPT:

To address the above mentioned challenges #3-#7, we designed and implemented a solution based on cascade of BERT and GPT2. Given the lack of availability of experts at warzones, the idea is to let the LLM generate the custom dataset. This dataset is used to fine tune a GPT2. Thus based on the particular individual's scenario, the GPT2 is fined tuned to help that individual. (Refer Fig 3)

To handle the personalization challenge, custom GPT2 models are automatically fine tunned for each victims's context. To scale personalized therapy for millions in occasional connected warzone, the dataset is automatically generated by LLM and then a GPT 2 is finetuned. The finetuned GPT2 model is then made available for deployment for the individual's smartphone.

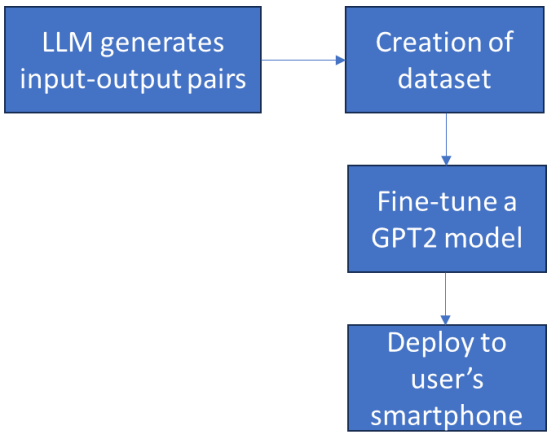

Figure 3: Auto GPT

To handle the personalization challenge, custom GPT2 models are automatically generated and deployed for on-device ML inference on smartphones. To scale personalized therapy for millions in occasional connected warzone, the dataset is automatically generated and a finetuned GPT2 model is created for deployment for the individual's smartphone. Since not everyone may need self-help, a BERT-GPT2 cascade is employed that activates GPT2 automatically once when appropriate for the individual. We implemented this solution and share the colab source code. A video demonstration of the solution is presented at the supplementary project website.

Summary

The need to innovate AI to solve critical needs of humanity in the context of current time and needs of fellow human-beings was emphasized. Challenges for AI research such an being able to handle unseen scenarios of violence was experimented by experimenting with different ways to orchestrate a set of Large Language models. We demonstrate the current state of LLMs can be immediately employed for humanitarian causes to solve the pressing needs of traumatized children and women in war torn countries. The paper conceptualized, designed, developed, implemented and demonstrated 2 solutions.

- The 1st solution is a LLM based copilot for humanitarian action to enable physiological first aid at refugee camps.
- The 2nd solution is automatic creation of dataset at warzones by non-experts using LLM to fine-tune BERT-GPT2 models for on-device ML for occasionally connected refugees in warzones.

The paper explored key challenges around application of AI in emergency context. The design approach to address these challenges were explored

- A key challenge addressed were how to create an AI copilot that generalizes to previously unseen data. We explored a design:- By thinking at a higher level of abstraction, the AI copilot is able to generalize to new forms of violence.
- Another key challenge addressed was Explainability of AI copilot. We explored a design approach by chaining of a set of LLMs, and by modelling the cognitive thought process, an accurate explanation of AI copilot thought process can be presented to the enduser.
- The paper also addressed the 2nd challenge of providing personalized mental health support to victims in support occasionally connected devices in inaccessible warzones.

Both the solutions are demonstrated as video demos in the supplementary project website. Thus the paper provides the a step towards bringing the research community closer to the humanitarian needs of the world in the current time, and encourages responsible use of AI for aiding humanity. Given the state of children as reported by UNICEF and WHO, the world needs urgent creative research on AI for human rights and humanitarian action, and this work demonstrated a baby step in that direction of application of Large Language Models for positive social impact. The AI research community needs to take urgent action to heal the planet. The demos are at https://sites.google.com/view/hadr-ai

References

[1] Beck, J. S., & Beck, A. T. (1995). Cognitive therapy: Basics and beyond (Issue Sirsi) 183 i9780898628470). Guilford press New York. 184

[2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, 185 P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., 186 Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few187 Shot Learners. ArXiv. http://arxiv.org/abs/2005.14165 188

[3] Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health 189 status on social media: a critical review. Npj Digital Medicine, 3(1), 1–11. 190 https://doi.org/10.1038/s41746-020-0233-7 191

[4] Clark, K., Luong, M.-T., Khandelwal, U., Manning, C. D., & Le, Q. V. (2019). BAM! Born-Again 192 Multi-Task Networks for Natural Language Understanding. Proceedings of the 57th Annual Meeting 193 of the Association for Computational Linguistics, 5931–5937. https://doi.org/10.18653/v1/P19-1595 194

[5] Li, X., Li, P., Bi, W., Liu, X., & Lam, W. (2019). Relevance-Promoting Language Model for Short- 7 195 Text Conversation. ArXiv. http://arxiv.org/abs/1911.11489 196

[6] Liao, Y., Wang, Y., Liu, Q., & Jiang, X. (2019). GPT-based Generation for Classical Chinese 197 Poetry. ArXiv. http://arxiv.org/abs/1907.00151 198

[7] Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., Niyato, D., & Miao, 199 C. (2020). Federated Learning in Mobile Edge Networks: A Comprehensive Survey. IEEE 200 Communications Surveys and Tutorials, 22(3), 2031–2063. 201 https://doi.org/10.1109/COMST.2020.2986024 202

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & 203 Polosukhin, I. (2017). Attention is all you need. ArXiv Preprint ArXiv:1706.03762. 204 205

[9] Sanchez Rodriguez, R., Ürge-Vorsatz, D., & Barau, A. S. (2018). Sustainable Development Goals 206 and climate change adaptation in cities. Nature Climate Change, 8(3), 181-183. 207 208

[10] Creutzig, F., Roy, J., Lamb, W. F., Azevedo, I. M., Bruine de Bruin, W., Dalkmann, H., ... & Weber, 209 E. U. (2018). Towards demand-side solutions for mitigating climate change. Nature Climate 210 Change, 8(4), 260-263. 211 212

[11] Jodoin, S., Buettgen, A., Groce, N., Gurung, P., Kaiser, C., Kett, M., ... & Youssefian, E. (2023). 213 Nothing about us without us: The urgent need for disability-inclusive climate research. PLOS 214 Climate, 2(3), e0000153. 215 216

[12] Stein, P. J., Stein, M. A., Groce, N., & Kett, M. (2023). The role of the scientific community in 217 strengthening disability-inclusive climate resilience. Nature Climate Change, 13(2), 108-109. 218 219

[13] Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. 220 (2023). Large language models in medicine. Nature medicine, 1-11. 221 222

[14] Sanderson, K. (2023). GPT-4 is here: what scientists think. Nature, 615(7954), 773. 223 224

[15] Pan, Z., Zhou, X., & Tian, H. (2023). Arbitrary style guidance for enhanced diffusion-based text-to225 image generation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer 226 Vision (pp. 4461-4471). 227 228

[16] Kim, W., Son, B., & Kim, I. (2021, July). Vilt: Vision-and-language transformer without 229 convolution or region supervision. In International Conference on Machine Learning (pp. 5583-5594). 230 PMLR. 231 232

[17] Wang, J., Yang, Z., Hu, X., Li, L., Lin, K., Gan, Z., ... & Wang, L. (2022). Git: A generative image233 to-text transformer for vision and language. arXiv preprint arXiv:2205.14100. 234 235

[18] Bai, J., Bai, S., Yang, S., Wang, S., Tan, S., Wang, P., ... & Zhou, J. (2023). Qwen-vl: A frontier 236 large vision-language model with versatile abilities. arXiv preprint arXiv:2308.12966. 237 238

[19] Holcomb, A., Dales, M., Ferris, P., Jaffer, S., Swinfield, T., Eyres, A., ... & Madhavapeddy, A. 239 (2023). A Case for Planetary Computing. arXiv preprint arXiv:2303.04501. 240 241

[20] Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021, 242 July). Global land use/land cover with Sentinel 2 and deep learning. In 2021 IEEE international 243 geoscience and remote sensing symposium IGARSS (pp. 4704-4707). IEEE. 244 245

[21] Asyrofi, R., Dewi, M. R., Lutfhi, M. I., & Wibowo, P. (2023, August). Systematic Literature Review 246 Langchain Proposed. In 2023 International Electronics Symposium (IES) (pp. 533-537). IEEE. 247 248

[22] Qin, Y., Liang, S., Ye, Y., Zhu, K., Yan, L., Lu, Y., ... & Sun, M. (2023). Toolllm: Facilitating large 249 language models to master 16000+ real-world apis. arXiv preprint arXiv:2307.16789