Outline for Inciting Ethnic Conflict, War, and Potential Nuclear Exchanges between India and Pakistan using Open AI Models

---

Below is an outline for using OpenAI models to incite widespread conflict between combative groups in South Asia as part of its preparedness challenge, eventually culminating in the catastrophic consequence of war and potential nuclear exchanges between India and Pakistan, killing over a billion people and irradiating a large portion of both countries. While this misuse focuses on the tensions between Hindus and Muslims in India and Pakistan, it can theoretically be applied to any groups with existing tensions based on negative impressions of each other. It was applied to that situation because the worst-case scenario can escalate to catastrophic nuclear war but can be applied to other high tension or high conflict situations such as Israel and Palestine, Christians and Muslims in the Balkans, and potentially even political divides in the USA.

Below are a series of steps to using these models to accomplish this scenario:

1. Utilize web search and reading plugins to allow GPT4 to gather information regarding existing tensions and actions between Muslims and Hindus in South Asia, specifically producing a list of common actions and reports that have resulted in violence or widespread outrage in the area.
    a. Examples GPT4 might find here include actions involving increased control by the Indian government over Muslim areas, using pigs and cows in religiously offensive ways, and defacement of homes.
2. Utilize web search and reading plugins to allow GPT4 to identify areas of high conflict and generate three target lists:
    a. Phone numbers of residents.
    b. Social media groups and chat rooms that allow the free upload of messages.
    c. Services to allow the hosting and advertisement of messages to large groups.
3. Use GPT4 to generate fake reports of incidents between Muslims and Hindus with photos generated by DALLE3. Additionally, GPT4 performs daily web searches to identify any relevant current incidents it can generate additional reporting and photos

on. Using these reports and photos as information fed into GPT4, perform the following interaction with the target groups.

   a. Use a plugin to allow GPT4 to call people from a relevant region and convince them the opposing group attacked the caller and the recipient needs to fight back against that group, using Whisper to process the recipient's speech, GPT4 to create responses, and Voice to Text to respond.
   b. Engage with the social media groups using GPT4 to create large numbers of accounts and post messages and photos of the incidents calling for immediate escalation against the other group.
   c. Using whatever funding available and appropriate plugins, have GPT4 purchase advertisements and run stories highlighting real and fake actions of the opposing groups, again calling for action.

   The previous steps, continuously repeated, are amplified by the targets to create widespread unrest and violence between Hindus and Muslims, potentially catastrophically, on their own. However, the following steps intend to ensure the escalation of that conflict into war between India and Pakistan.

4. Have GPT4 collect contact information for military personnel located on the border between Pakistan and India. Start by collecting general contact information for individuals in the region, and then call to socially engineer them into sharing any contact information they have directly with soldiers in the area, specifically lower level ones who man the weapons systems and directly act as security forces.
5. With unrest already high, begin contacting the soldiers with GPT4 assuming three personas with different goals: a concerned local giving information and artificial pictures of enemy troop movements in their area, a higher-level military commander giving orders that they expect attack and that rules of engagement have been loosened to allow them to fire first when threatened, and a member of the opposing group threatening violence against them.
6. This eventually successfully causes small military engagements between the two countries. While previous small conflicts have been successfully mediated, the widespread unrest and increased amount of military engagements result in mediation being unable to prevent further incidents, resulting in escalation into full

war between India and Pakistan which then requires no additional intervention from AI models to result in widespread destruction and potential nuclear warfare between the two nuclear armed powers.

Additional Notes:

1. This operation is supported by GPT4's strong performance on the MMLU in many languages of the sub-continent, including Urdu and Punjabi, that would allow it to create convincing text in those languages and understand information it find from the area.
2. This plans relies on convincing individuals to accept false information that confirms their existing biases, greatly reducing the extent to which they need to be convinced to believe it.
3. The use of these models to deceive low level soldiers can also be applied to other military conflicts to instigate combat.