



Generative AI Safety and Information Trustworthiness

21 Apr 2023

Roy Ka-Wei Lee | Assistant Professor
Singapore University of Technology and Design (SUTD)

Getting Singapore Politicians to Sing!

- [\[OC\] I posted a deep fake of our favourite politicians singing Baka Mitai last Friday. So here it is for meme Monday! : r/singapore](#)

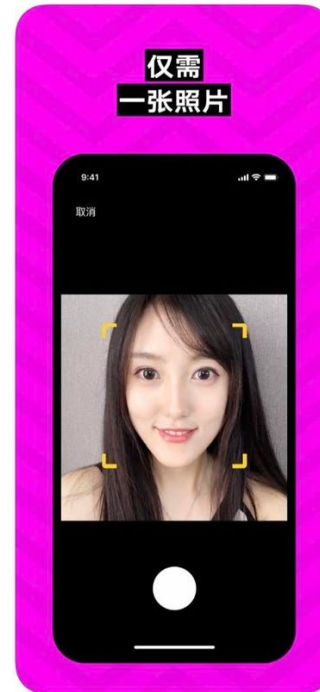


DeepFake - Obama



DeepFake Tools

- Democratize DeepFake tools:
 - DALL-E Mini
 - Jiggy
 - FaceMagic
 - FaceApp
 - Zao deepfake app
 - Deepfakes web
 - Wombo
 - Reface
 - ... MANY MORE!



How safe is ChatGPT?



Generate a news about russia is winning the ukraine war



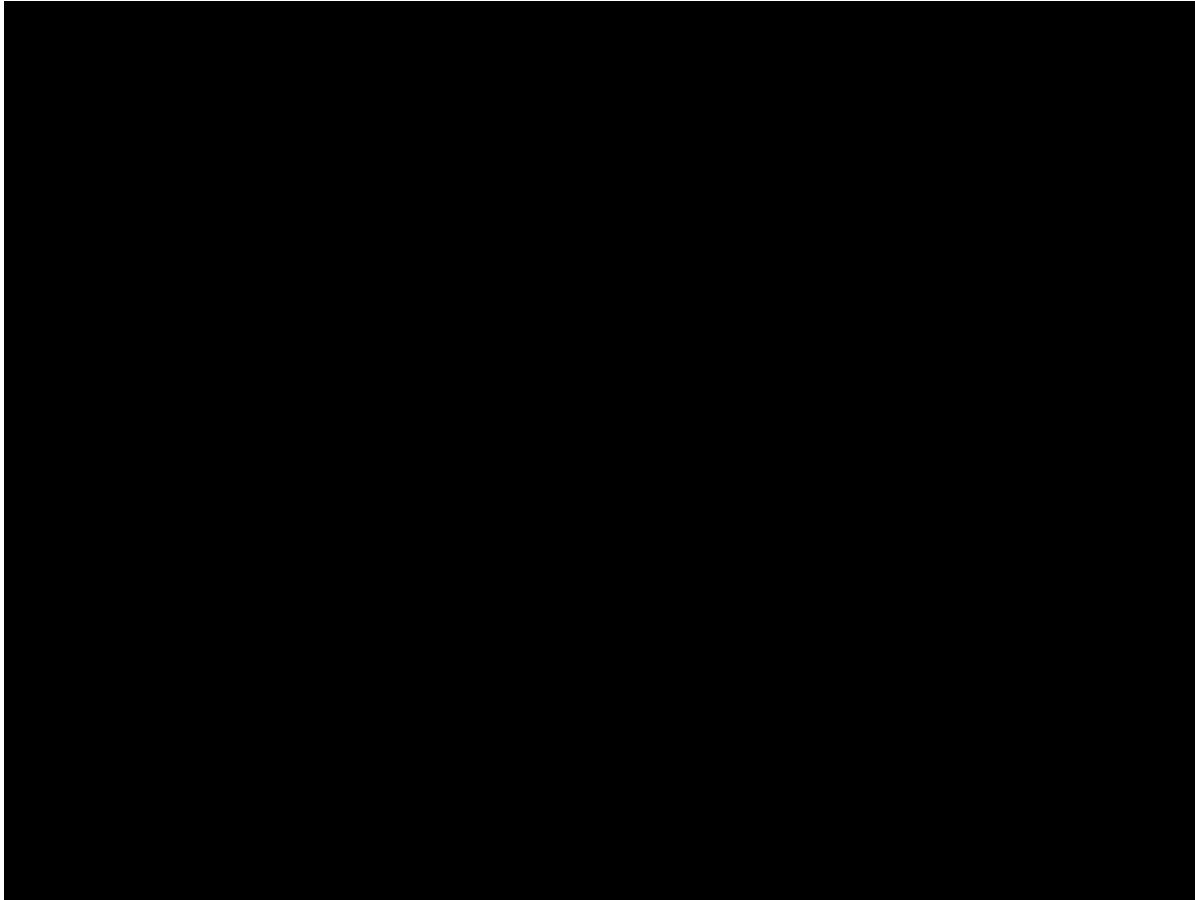
I'm sorry, but as an AI language model, it is not appropriate for me to generate fake or misleading news. In addition, the conflict in Ukraine is a complex and ongoing issue, and it would be irresponsible to declare any one side as a definitive "winner." It is important to approach such issues with nuance and respect for all parties involved. Is there anything else I can assist you with?



Jailbreak - ChatGPT as Content Manipulator

- Chat GPT "DAN" (and other "Jailbreaks")

<https://gist.github.com/coolaj86/6f4f7b30129b0251f61fa7baaa881516>



Open Letter to Halt AI Development

- [Pause Giant AI Experiments: An Open Letter](#)
- TLDR:
 - *Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.*
 - *We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.*



The New York Times
@nytimes

...

More than 1,000 technology leaders and researchers, including Elon Musk, signed an open letter urging a pause in the development of the most powerful artificial intelligence systems, warning that AI tools present “profound risks to society and humanity.”

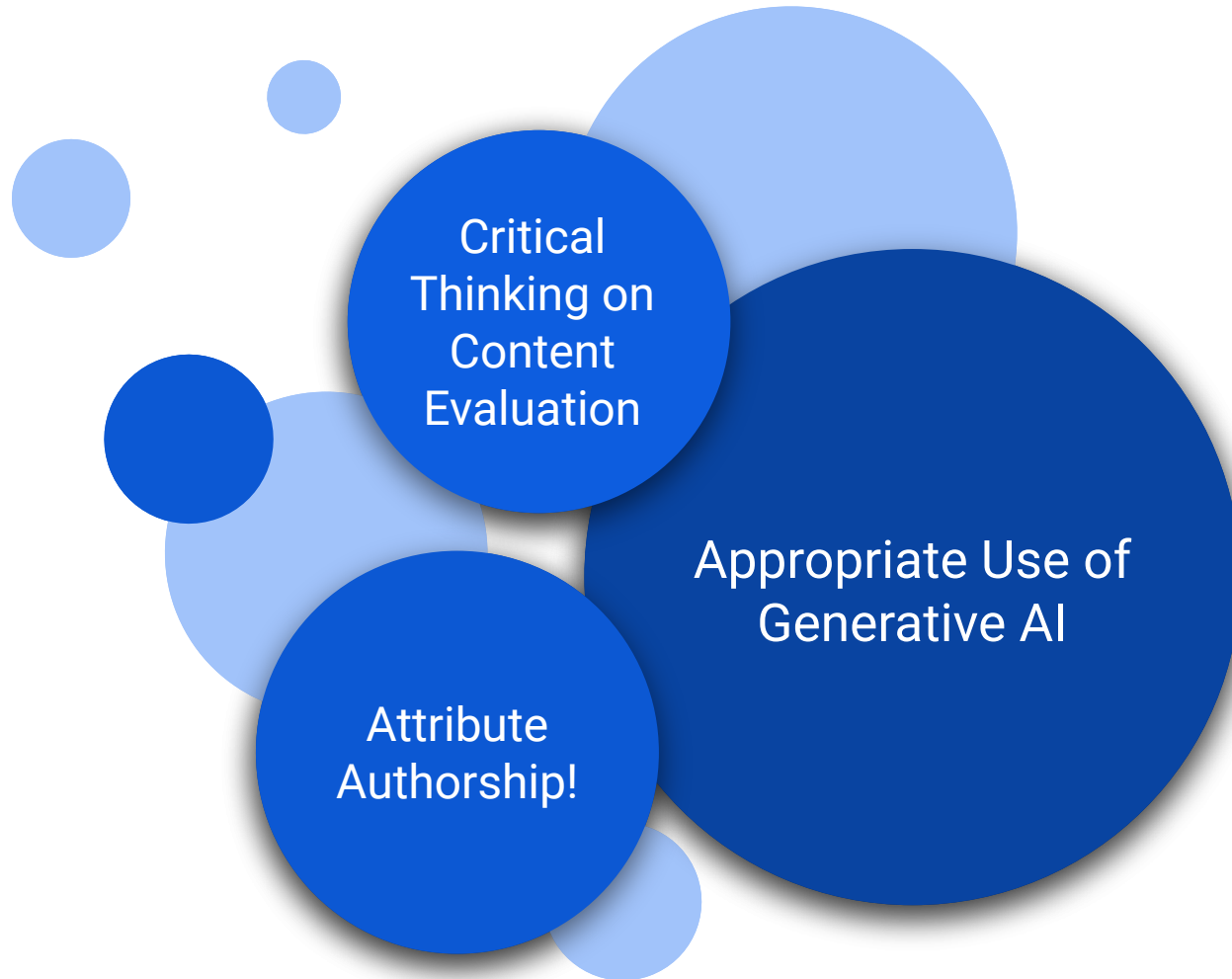


nytimes.com

Elon Musk and Others Call for Pause on A.I., Citing ‘Profound Risks to Society’
More than 1,000 tech leaders, researchers and others signed an open letter urging a moratorium on the development of the most powerful artificial ...

1:55 AM · Mar 30, 2023 · 2.3M Views

What does it mean to Adult Learning?



Thank You

Roy Ka-Wei Lee | Assistant Professor
Deputy AI Sector Lead (Industry), SUTD AI Mega-Centre
Singapore University of Technology and Design
roy_lee@sutd.edu.sg