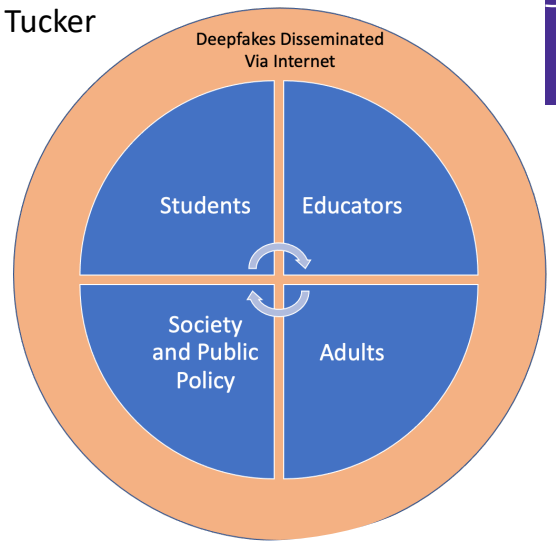
EAGER: SaTC AI-Cybersecurity: Safeguarding STEM Education and Scientific Knowledge in the age of Hyper realistic AI-Generated Data

RAND Corporation: Dr. Christopher Doss and Dr. Jared Mondschein
Carnegie Mellon University: Dr. Conrad Tucker
Challenger Center: Dr. Lance Bush

Challenge:

- AI lowers barriers to mass creation and dissemination of manipulated digital content (deepfakes)
- The risk of exposure among education stakeholders has increased as learners and educators rely more on the Internet for information
- We do not know the level of vulnerability or what video and personal characteristics moderate vulnerability



Solution:

- Fielded randomized controlled trials in surveys to understand the vulnerabilities of U.S. adults, educators, K-12 students and college students
- Asked respondents which aspects of videos helped make determination of videos' authenticity
- Asked about climate change knowledge, political orientation, learning habits, and social media use

Scientific Impact:

- All stakeholders are vulnerable
- Adult populations were more vulnerable
- Those who reported more trust in information were more vulnerable
- More exposure increased vulnerability
- Social context of video helps to better identify deepfakes, but was used less by respondents

NSF Award # 2039612

RAND Corporation: Dr. Christopher Doss, cdoss@rand.org
Carnegie Mellon University: Dr. Conrad Tucker, conradt@andrew.cmu.edu
Challenger Center: Dr. Lance Bush, lbush@challenger.org

Broader Impact and Broader Participation:

- Characteristics of more vulnerable stakeholders were identified
- Potential threat of unmitigated deepfakes was quantified
- Importance of social context quantified
- Future work can port lessons into mitigation strategies