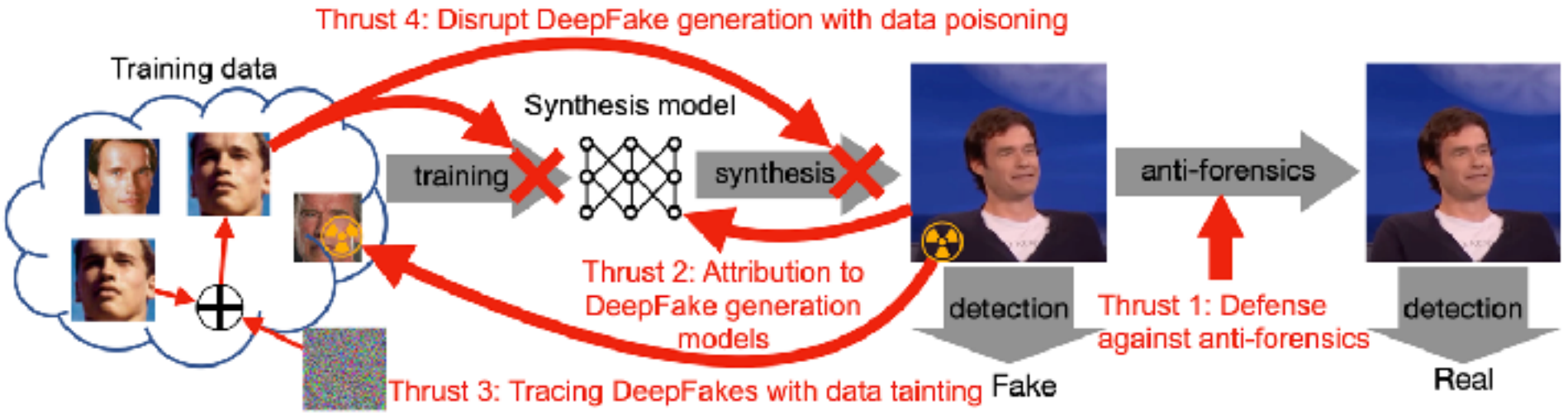
SaTC: CORE: Small:
Combating AI Synthesized Fake Media Beyond Detection



Challenge:

- AI synthesized fake media have negative societal impacts as a means of disinformation
- Passive detection is not sufficient to provide timely and effective protection for users

Solution:

We will develop proactive and active approaches

- to disrupt training of the AI synthesis models,
- to add traceable marks to synthesized media,
- to attribute the specific model used to generate a DeepFake,
- and to defend an existing or new DeepFake detector from anti-forensics.

Scientific Impact:

- The project targets at the problem of AI synthesized media, which is increasingly used for cognitive/cyber attacks.
- The results from this project will provide deeper insights about the capacities and vulnerabilities of the state-of-the-art detection methods.
- The proactive and active approaches complements the detection methods for more comprehensive defense.

Broader Impact and Broader Participation:

- Collaborate with NYS Assembly on legislations concerning synthesized media
- Provide DeepFake-o-meter (http://zinc.cse.buffalo.edu/ubmdfl/deep-o-meter),
- Design new course on Multimedia Forensics,
- Encourage women and URM participation in research.

NSF SaTC-2153112, University at Buffalo, State University of New York
PI: Siwei Lyu (siweilyu@buffalo.edu)