Just a few **tampered** ballots can swing a close election.

This is a cautionary tale about using machine learning in critical systems.

Research Question

Is it possible to use adversarial machine learning in the voting domain to flip votes?



Threat Model

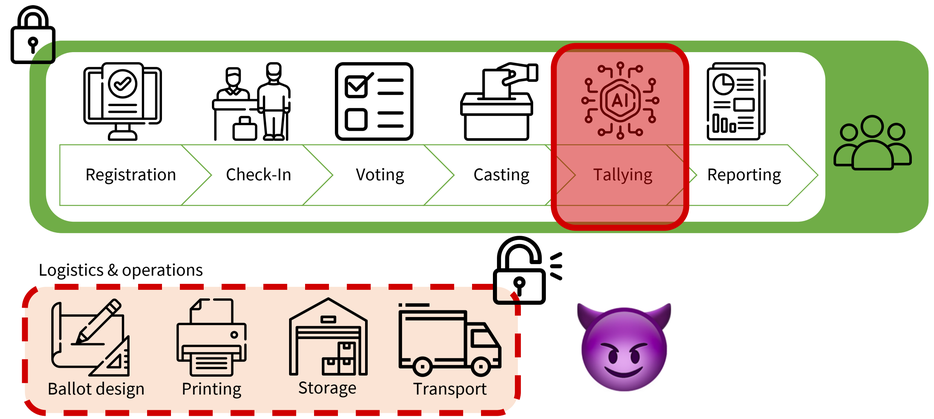
- Election process is well-defined and well-secured.
- We only replace the traditional tabulator with a ML classifier.
- Physical-world pipeline is unsecured.
- Adversary has full white-box knowledge of ML system (but no direct access), and full access to physical pipeline.

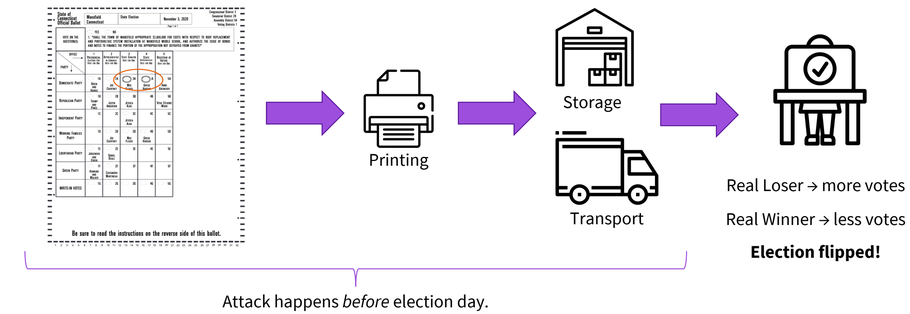
Key Results

- Models trained on variety of marks classify more accurately but cannot resist adversarial attacks.
- Physical world process degrades image signal and affect classification.
- ML in voting domain is very risky because we do not fully understand security landscape.

References

1. Barretto et al. *Improving the accuracy of ballot scanners using supervised learning*. In E-Vote-ID (2021).
2. UConn Voter Center. *Voting bubbles with swatches*. (2025). Available at https://doi.org/10.5281/zenodo.15458710.
3. Specter et al. *The ballot is busted before the blockchain: A security analysis of voatz, the first internet voting application used in US federal elections*. In USENIX Security. (2020).







Attack Process

Goal: for adversary to mix tampered ballots with real ballots, enough to flip election by 0.5%.

How is this achieved?

- Adversary uses white-box knowledge of ML classifier to craft imperceptible adversarial images.
- Adversary tests their attack by printing + scanning to model the physical process.
- When an attack survives the physical process *and* is imperceptible *and* is misclassified, adversary has succeeded.
- Adversary would then use the tampered bubble for adversarial ballots, and then print the required number of ballots to flip the election by 0.5%.

Technical attack details

- Trained models: SimpleCNN, ResNet-50, VGG-16, Twins, CaiT.
- Attacks: APGD, MIM, PGD, FGSM attacks.
- $\varepsilon = \{0, 4, 8, 16, 32, 64, 128, 255\} / 255$.
- CE and DLR loss.



This poster is based on our research paper accepted at CCS 2025: "Busting the paper ballot: Voting meets adversarial machine learning". K. Mahmood, C. Manicke, E. Rathbun, A. Verma, S. Ahmed, N. Stamatakis, B. Fuller, and L. Michel.

Disclaimer: The images of ballots in this poster are for illustrative purposes only. They are not intended as endorsements of any candidate, political party, or policy.

Aayushi Verma
aayushi.verma@uconn.edu
awesomecosmos

Check out our paper!
https://dl.acm.org/doi/10.1145/3719027.3744882