Black Box AI and DHS Systems

Matthew Merzbacher
ADSA 23 ~~ May 4, 2022

SWWC: Can Black Box AI be used in DHS Systems?

- ➔ Black-Box AI (typically using “machine learning”) works for many domains
  - Throw data at the problem, let the algorithm sort it out
  - Is this just a fad or is it real? Is it good enough that it works?
    Facial Recognition (DHS Biometric Rally): 93% without masks, 77% with masks
    DHS (e.g., SVIP week) embracing the possibility
- ➔ Issues: Data, Bias, Observability
  - Data: Do we have enough? Of the right kind?
  - Bias & Equitability:
    - Facial recognition has dramatically different performance for different race/age groups
      - False Negatives & False Positives
    - Iris & Fingerprints don’t suffer similar bias
  - Observability: Explain (local) reasoning behind decisions
- ➔ Need technology testing (against spec) & scenario testing (bias/observability)
- ➔ With (local) observability, can Black-Box AI be used?

Data & Bias: Huskies versus Wolves

1. Segmentation
2. Feature Extraction
3. (automated) Classification

90% performance

But, when explanation was added…

What Happened?

Is detecting wolves by detecting snow a good or bad precedent for DHS applications of AI?

https://www.researchgate.net/figure/A-husky-on-the-left-is-confused-with-a-wolf-because-the-pixels-on-the-right_fig1_329277474

Why do we care?

"Watch the doughnut not the hole" – *Burl Ives*

Should we care?

Adoption of New Technology Requires…

Better: High performance, Automated development, Data driven (unbiased)

Compatible: Plug & play

Simple: easier to accept, less scary

Testable

Observable: understand choices (Regulators, Testers, Screeners, Passengers, Adversaries)

General Data Protection Regulation (GDPR) limits use of black box AI, primarily due to observability concerns

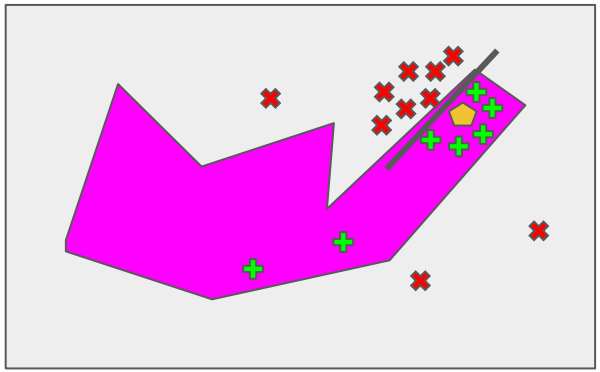
Achieving Observability: Explanation Systems

Let's Automate!

LIME (University of Washington)

https://arxiv.org/pdf/1602.04938.pdf

Locally faithful, Globally not



Human observers can validate, gain confidence

Used on Husky/Wolf example to validate

Good for testing, maybe for screeners

Concerns

- Overreaction to localized explanations
- Who decides a reason is wrong? Can a “wrong reason” be a good reason?
- Is it sufficient to meet the specification without explanation?
  - Improve the spec!
- Explain end-to-end decisions in a complex system, including human-in-loop
- Bias:
  - Can we explain bias?
  - Evaluating the Equitability of Commercial Face Recognition Technology in DHS Scenarios
    - https://www.youtube.com/watch?v=cd_KQN6YLaA
- Technology Testing validates the technology against the spec
- Scenario Testing checks for bias and other systemic issues