



Homeland Security



**Homeland Security Advisory Council
Report from the Emerging Technologies Subcommittee
November 14, 2019**

Emerging Technologies Subcommittee Tasking

- Unmanned Autonomous Systems (UAS) & Counter UAS (cUAS)
- Artificial Intelligence, Machine Learning (AI/ML)
- 3/4-D Printing
- Biotechnology - Gene editing, splicing
- Quantum Computing
- Advance Robotics

HSAC Emerging Technologies Subcommittee - Status

- Unmanned Autonomous Systems (UAS) & Counter UAS (cUAS)
- Artificial Intelligence, Machine Learning (AI/ML)
- 3/4-D Printing
- Biotechnology - Gene editing, splicing
- Quantum Computing
- Advance Robotics

Recommendations of the Draft ET Interim Report

Further Subcommittee Work

- Interim report
- Subcommittee continued its work for an additional 180 days
- Requested assignment of subject matter experts for the technologies under review
 - Priority: Technical expert on unmanned aerial systems
- Cathy Lanier served on the Blue Ribbon Task Force (BRTF) on UAS Mitigation at Airports (Sponsored by the Association for Unmanned Vehicle Systems International & Airports Council International – North America). Recommendations that align with this effort are noted BRTF.

Recommendations of Draft ET Interim Report on UAS/cUAS (1 of 2)

- Continue to place a high priority on the implementation of the new authorities granted in the 2018 FAA Reauthorization
- UAS/cUAS efforts be made a permanent program of record in appropriations.
- Consider proposing legislative changes that would identify TSA's role and authorities related to UAS/cUAS.
- Develop capabilities matrix arraying individual component activity.
- Rapidly share test and evaluation information and evolving CONOPS.

Recommendations of Draft Interim Report on UAS/cUAS

- Engage SLTT authorities and identify operational, tactical, and legal issues that must be addressed to implement UAS/cUAS locally.
- Propose test sites for technology evaluation and CONOPS at four use cases: fixed locations (covered assets), regional locations (SW Border), temporary locations (special events), and mobile locations (dignitary, mobile asset protection).
- Use implementation of UAS/cUAS authorities and capability as a use case to operationalize “unity of effort” across the Department.
- Consider the current wide variation of technologies being developed and employed by the federal government and SLTT authorities a safety issue that requires close attention.

Recommendations of Draft Final Report on UAS/cUAS

- Dedicated, recurring funding is needed to resource
 - The cUAS Coordinator's/Program Office and hire permanent staff
 - Component procurement of approved technology, perform Testing and Evaluation, and conduct Research and Development.
- Support funding for the FAA office to support cUAS operations (*Congress must additionally authorize and appropriate the necessary funds for the FAA to ensure adequate testing, acquisition, deployment, staffing, and maintenance of DTI technology in the airport environment.*) **BRTF**
- *There is an urgent need for the FAA to establish UAS detection & mitigation system standards, and provide straightforward guidance to those seeking to deploy DTI technologies.* **BRTF**

Recommendations of Draft Final Report on UAS/cUAS

- *Congress must extend authority to engage in UAS interdiction— kinetic or electronic—to trained state and local law enforcement. **BRTF***
- The deputation of C-UAS authority to state and local law enforcement, consistent with existing task forces, overseen by DOJ to ensure legal protections on privacy and liability.
- There is an urgent need for clear direction to be issued by the FAA and DOJ to State and Local Law Enforcement regarding their current role within the current legal, statutory and regulatory limitations. **BRTF**

Recommendations of Draft Final Report on UAS/cUAS

- Authorize SLTT law enforcement, airports, and certain critical infrastructure owners to procure and operate essential detection equipment, provided they are certified and trained to do so.
- Identify whether DHS can charge a fee for providing “cUAS as a Service” to allow for more rapid procurement and training of work force.
- Identify whether a private entity (i.e., sports team or airport) can procure, store, and maintain equipment which DHS or DOJ then operates under certain agreed-to conditions.

Overview

Artificial Intelligence / Machine Learning

Excerpt from the Final Report of the Emerging Technologies
Subcommittee

Context

The increase in federally sponsored AI research and development (R&D), combined with investments from private industry and other R&D organizations, will **foster the development of a large set of technologies that undoubtedly will have lasting impacts on national security and the entire homeland security enterprise.**

Assessment of perceived AI threats over next 10 years

1. Two to five years

- Deepfakes

2. Long-term arms race

- AI-driven social media attacks

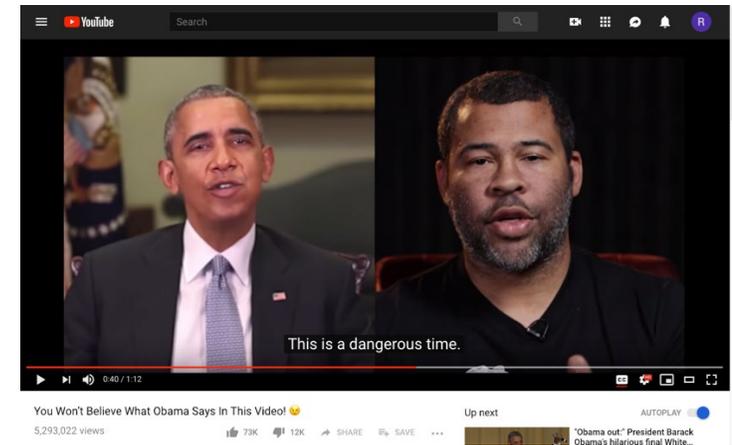
3. Potentially emerging but not imminent threats (five to ten years)

- Information attacks on emerging AI infrastructure
- AI-driven cyber attacks
- Large-scale social engineering attacks

Deepfakes

- **Projected timeline:** 2 - 5 years.
- **Overview:** “Deepfake” algorithms utilize deep learning to almost seamlessly map target images, video, or audio content into other media content to create realistic depictions of situations that never occurred.
- **Current state:** Recent techniques are more powerful and can capture integrated head position and rotation movements, facial expressions (including eyebrow movements and blinks), and eye movements.

Figure 1: Example of a deepfake video
"President Obama" delivers a warning message about deepfake technology



AI-driven social media attacks

- **Overview:** Social media attacks can be defined as attacks that utilize fake social media messages to influence or disrupt public discourse. The goal of social media attacks, when undertaken against the homeland, is typically the dissemination of falsehoods in order to gain temporary political advantages, delegitimize political opponents of the attacker, or damage public safety in other ways, such as misinforming the public about existing crises or fomenting rioting or acts of vandalism.
- **Current state:** Current types of attacks tend to utilize both AI-driven bots and armies of human actors. Attackers utilize social media platforms, such as Twitter or Facebook, to make it appear that certain opinions or beliefs are more common than they are among the public, often to lend public support to positions that are favorable to the attacker. Bots—software applications that run automated tasks online—can also be used to boost the visibility of actors or users on social media platforms. These artificial social agents may or may not be easily spotted by the average user.

New threats to homeland security

Deepfake voice technology used to create crisis

- **A deepfake voice tool could be used to simulate commands or instructions delivered over the phone.** This could be used to generate an artificial crisis, such as an order to take a political opponent into custody, evacuate a building, or send emergency resources to an area, perhaps to divert them from a planned real attack.

Botnets to delegitimize public fora or make them unusable

- Instead of attacking a particular position, **AI-driven botnets could be used to simply drive up the discussion level on both sides of an issue to a level that would render the social media platform unusable for discussion**, or at least unusable for certain topics. This is also a matter of public trust. If all sources of information are demonstrated unreliable and compromised, the public's trust in any information, including legitimate messages, will decrease, potentially resulting in serious impacts to messaging during a time of crisis.

Recommendations of the Final Report Artificial Intelligence / Machine Learning

New capability for homeland security

Media forensics units

- **Special units within DHS could be provided with the latest tools to combat deepfake technology**, such as DARPA's MediFor toolkit, along with alternative means of verification, to combat arising fake videos, images, and audio. Alternatively, a standards agency such as National Institute of Standards and Technology (NIST) could certify organizations that detect fake media.

Recommendations to mitigate the perceived deleterious impacts of the assessed technological advancements

It is anticipated that the AI technologies will be used in the DHS operational environments to support DHS in executing various missions and priorities. Therefore, it is important for DHS to invest in workforce development for an AI-ready workplace. Additionally, capabilities such as AI Testbeds and AI Forensic teams can be established through a combination of Public-Private Partnerships, as well as federally funded multi-agency/multi-use infrastructure, to ensure DHS is prepared to deter potential threats that may arise from malicious AI systems.

- ***Recommendation #1: Provide mechanisms or standards for validating user identity across platforms.*** Currently, some social media platforms have mechanisms for identity validation, but widespread real-world validation of user identity—using government identification or similar means—remains rare, nor are there industry standards for identity validation for social media. Widespread identity validation for regular users, as well as open industry standards for identity validation across social media platforms, would both reduce costs and any perceptions of user endorsement.
- **Recommendation #2: Encourage standards for commercial providers of imagery technology to include watermarking and other anti-fraud measures to help combat deepfakes.** To help combat deepfake technology, it may be possible to embed watermarks or digital signatures to label known true images or videos or, as part of image manipulation software, to mark images or videos as modified. In addition, image creation systems could optionally register images or videos to a public ledger using blockchain technologies, as done by the camera app TruePic.

Back-up Slides

Definitions

- **Artificial Intelligence (AI)** – “The ability of machines to perform tasks that normally require human intelligence. For example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems.”
- **Machine Learning (ML)** – A subset of AI, ML approaches use algorithms, predominantly statistical, that learn how to perform classification or problem solving without being explicitly programmed for the task domain.

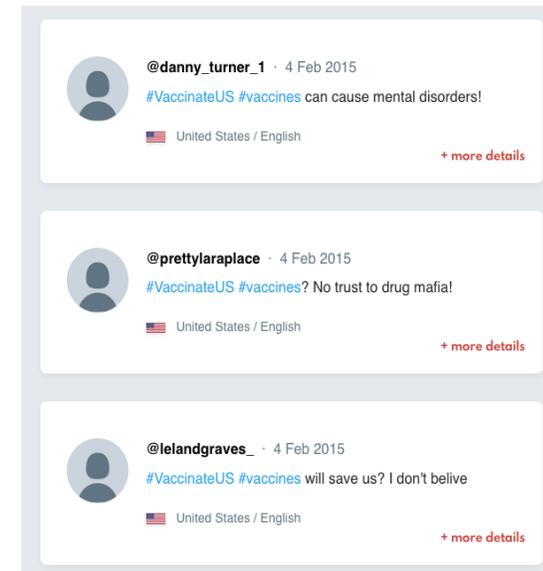
Deepfakes

- **Expected advances:** Significant improvements in coming years, harder to detect and deem fake, easy to acquire through open source.
- **Impediments/countermeasures:** Using techniques that detect tiny disfluencies in the generated video; watermarks; or making the production of deepfakes more computationally intensive [insert FN]. Once particular artifacts of the process are identified, such differences can be trained against and eliminated using adversarial neural network techniques.
- **Converging technologies:** Deepfakes and social media attacks could increase the ability of such methods to disrupt social structures and political activity. Similarly, the ability to mimic voice could be used to supplement cyber-attacks by automating, for example, voicemail that suggests opening a spear-phishing email.

AI-driven social media attacks

- **Expected advances:** Sophisticated narratives of attacks, automated responses.
- **Impediments/countermeasures:** Stronger verification techniques, user-reporting of bots, ML-based bot recognition and anomaly-detection algorithms to improve bot detection, geographical filters to eliminate trolls, public education, human tests such as CAPTCHAs.
- **Converging technologies:** Deepfakes will increase severity of social media attacks; cyberattacks will improve the ability of malicious actors to take over social media accounts.
- **Projected timeline:** Long-term arms race.

Figure 2: Example of Russian tweets on vaccine debate produced by Russia's Internet Research Agency.



Emerging but not imminent threats (1 of 3)

Information attacks on emerging AI infrastructure

- **Overview:** Actions to disrupt emerging AI capabilities—be they fleets of autonomous vehicles, voice assistants used for critical functions, or other newly-essential AI technology—will themselves constitute threats. Various techniques could fool the systems into misclassifying or misinterpreting information in their environment.
- **Current state:** Attacks rely on a fair amount of technological sophistication to be properly implemented, and there are many alternative means of attack that are currently cheaper and more effective.
- **Projected timeline:** 5 – 10 years.

Emerging but not imminent threats (2 of 3)

AI-driven cyber-attacks

- **Overview:** AI-driven cyber-attacks utilize AI to help direct the infiltration, capture, or disabling of targeted computer systems. Evolving AI capabilities are likely to permit a small number of human attackers to direct attacks against a much larger number of targets.
- **Current state:** Little evidence for the use of AI in cyber-attacks “in the wild” to date, but there are recent research demonstrations of the utility of AI for cyber-defense.
- **Projected timeline:** 5 – 10 years.

Potentially emerging but not imminent threats (3 of 3)

Large-scale social engineering attacks

- **Overview:** Social engineering attacks are a kind of cyber-attack that use social vectors as part of the method for infiltrating a system (e.g., spear phishing).
- **Current state:** Attacks currently require careful analysis of their targets. However, advances in AI to extract information from social media and other sources of what has been called “digital exhaust” generated by individuals’ online actions (e.g., search and browser history) opens up the possibility of mass spear-phishing attacks, where an AI agent constructs targeted messages for each individual, even if the number of targets is in the hundreds or thousands.
- **Projected timeline:** 5 – 10 years.