THE FUTURE OF WARFARE AND THE ROLE OF EMERGING TECHNOLOGIES

25 November 2020

A Somewhat Contrarian View of AI in Warfare

Dr. Herb Lin

Stanford University

Some military applications for AI

- Semiautonomous and Autonomous Vehicles
  - Lethal Autonomous Weapon Systems
- Command and Control
- Intelligence, Surveillance, and Reconnaissance
- Logistics
- Cyberspace Operations
- Information Operations

Ask yourself…

Question 1

Will artificial intelligence lead to more or less understanding of the battlespace?

A: More understanding

B: Less understanding

Question 2

Will artificial intelligence increase or reduce the mistakes made in weapons systems targeting?

A: Reduce mistakes

B: Increase mistakes

Question 3

Will the most strategically important defense applications of AI be for military operations or civilian management?

A: Military operations

B: Civilian management

AI is very bad at explaining itself

- State of the art AI today is machine learning.
  - “Supervised ML” depends on labeled training data; makes statistical inferences.
  - “Unsupervised ML” finds clusters and outliers in unlabeled data that might otherwise go unnoticed if examined by humans.
- ML systems are unable to explain to human users the conclusions they make.
  - ML is correlation; correlation is not causation; causality required for explanations
  - Human examination of the path from input to output yields little insight about the features of the input that led to the inference in question.
  - Statistical reasoning yields inferences about populations, but not individuals
- Lack of explainability makes people…
  - unable to understand errors that occur
  - question novel but brilliant inferences (the move 37 problem)

Explainability in a military context

- No plan ever survives contact with the enemy, and the enemy gets a vote.
- ML systems must be built and trained on the very same *hypothetical data* based on what military planners *predict* contact with the enemy will entail.
  - How will the ML systems perform when real contact with the enemy is different from assumptions built into training data?
  - How will the ML system even know that it is being asked to perform outside of its range of competence?
- Questions that might be asked:
  - Why did an automatic target recognition system identify a bulldozer as a tank?
  - Why did the decision-support system recommend moving a battalion to A rather than B?

ML can easily be tricked, hacked, and fuzzed

Key example: Adversarial inputs in computer vision

Step 1: pick starting image (“sloth”)

“sloth”

>99% confidence

Step 2: pick target class (“race car”)

Step 3: create adversarial image by adding carefully chosen imperceptible noise

“race car”

>99% confidence

Combat example: Misidentification of window reflection

Imagine a hypothetical scenario where an ML-enabled image classification targeting software for a missile system confuses *reflections* of gunshot muzzle flares on the windows of a building (say, a hospital) for gunshots coming from *inside the hospital* itself.

The conclusion matters.

AI, Laws of Armed Conflict, and A Race to the Bottom

- AI is a *technology*, but kinetic combat also entails
  - How they are used, i.e., doctrine and rules of engagement
  - Numbers of fielded systems that engage in combat;
- The West will build into its AI-driven weapon systems its interpretation of LOAC, and adversaries will build into their AI-driven weapon systems their own interpretations.
- If great powers develop roughly equal levels of technology, whose weapons systems will have greater levels of military effectiveness? I posit that advantage goes to the side with looser LOAC interpretations.
  - Precedent – unrestricted submarine warfare in WWII.
- Our choices:
  - accept lower military effectiveness (unacceptable on the battlefield)
  - develop better technology (but we will be doing that anyway)
  - relax our own LOAC standards. This is the most likely outcome.

Historical precedent: unrestricted sub warfare

- Unrestricted submarine warfare refers to the wartime practice of submarines sinking civilian ships without warning.
- Practiced first in World War I, banned by London Naval Treaty of 1930; required submarine commanders to provide for the safety of those on board before attacking a merchant ship.
- Giving warning severely compromised submarine effectiveness and made them vulnerable (Q-ships), and both Allied and Axis forces abandoned Treaty restrictions on submarine attack.
- Will a similar dynamic emerge with respect to LOAC compliance?
- Coda - violations of London Treaty were ignored at Nuremberg Trials.

Deliberate escalation; inadvertent escalation

- Training data for ML systems will never fully capture or predict the range of real-world inputs.
  - Users may not realize ML-based decision support systems is operating outside of scope of training data
  - ML-enabled decision support systems have no reason to question user input regarding self-assessed capabilities ➔ recommendations may be based on overly rosy estimates of own capability vs actual capability.
  - Deliberate escalation may lead to defeat or disaster rather than success.
- ML systems often undertake actions that narrowly optimize the achievement of particular design specifications at the expense of the human-intended goal
  - Military ML system could optimize for success in a conventional military engagement without recognizing the potential for inadvertent escalation.

On excessive trust in AI

- AI is an enabling technology of the future.
- AI will be ubiquitously embedded in civilian and military devices and infrastructure.
- Ubiquitous applications are, by definition, not novel applications.
- User skepticism about AI depends in large part on novelty.
- Reduction in skepticism will lead to excessive trust in AI-driven systems.
- More mistakes, accidents, and untoward events will occur.
- We may not recognize them as such when they do occur.

The most important AI defense applications today are lower-stakes and least flashy (*not weapons and C2*)

- At least with present technology, the greatest potential impact for artificial intelligence applications on military and national defense is for low-stakes applications.
  - "boring" administration, management, and logistics tasks
  - Examples: health care, procurement, logistics, intelligence analysis, reports.
- Better tooth-to-tail ratios become easier to achieve.
- The civilian sector will develop good metrics of success, and military and civilian efforts can support each other.
- Civilian-oriented efforts are not well-matched to support military-specific applications.

How to use AI for military applications – some guidelines

- Think of machine learning as statistical rather than smart.
  - Corollary: more likely useful for populations rather than individual cases
- Focus on explainability
  - Meld “old” paradigms of rule-based AI.
- Aim for incremental rather than revolutionary improvement
  - 10% annual improvement it 2X better in 7 years
- Gain experience in low-stakes environments where mistakes don’t matter very much.
- Be able to recognize success and failure when you see it.
- Always ask “what could go wrong?” and never accept vendors’ answers.

Two closing thoughts

There are 3 roads to ruin.

- Alcohol is the fastest.
- Sex is the most fun.
- Technology is the most certain.

AI is what you call something that doesn't quite work and you don't understand very well.

- It works more often than not, but not all of the time.
- You can't predict when it will work, and therefore should not have high confidence that it does work.
- When it doesn't work, and even when it does, you don't understand why it did or didn't work, and you don't *know* how to make it work better.

THE FUTURE OF WARFARE AND THE ROLE OF EMERGING TECHNOLOGIES

25 November 2020

For more information

Dr. Herb Lin

herblin@stanford.edu

Senior Research Scholar, CISAC

Hank J. Holland Fellow in Cyber Policy and Security, Hoover Institution

Stanford University

GLOBSEC
IDEAS SHAPING THE WORLD

THE FUTURE OF WARFARE AND THE ROLE OF EMERGING TECHNOLOGIES

NATO OTAN 2030

25 November 2020

The research and conclusions contained in this presentation were conducted and developed within the NATO S&T Organization (STO) (www.sto.nato.int), drawing upon the support and participation of the Alliance's defence S&T community. This presentation does not represent the official opinion or position of NATO or individual governments, but provides considered advice to NATO and Nations' leadership on significant S&T issues.

Stanford | CISAC Center for International Security and Cooperation

HOOVER INSTITUTION