ARTIFICIAL INTELLIGENCE · THE DARK SIDE

AI IN WARFARE

How Artificial Intelligence is Reshaping the Battlefield and Threatening Humanity

SUBJECT: ARTIFICIAL INTELLIGENCE | TOPIC: DARK SIDE OF AI | MADE BY: SHASHANK PACHORI

AGENDA

01 **Physical Warfare & Autonomous Weapons**

Maven Smart System · LAWS · Killer Drones

02 **Cyber Warfare & Data as a Weapon**

Stuxnet · AI-powered cyberattacks · Critical Infrastructure

03 **Psychological Warfare & Deepfakes**

Disinformation · Cognitive Warfare · Synthetic Media

04 **Automation & the Kill Chain**

Lethal Autonomous Weapons Systems · Human-out-of-loop

05 **Geopolitical & Economic Impact**

AI arms race · Tech inequality · Economic disruption

06 **Iran–Israel War & AI's Role**

Lavender AI · The Gospel · Civilian Casualties

07 **OpenAI × Pentagon Deal**

Policy reversal · Threat scenarios

08 **Governance & the Way Forward**

UN frameworks · Red lines · What must change

MAVEN SMART SYSTEM & AUTONOMOUS WEAPONS

PROJECT MAVEN (2017–Present)

What it is:

Pentagon's flagship AI programme using computer vision to auto-identify people & objects in drone footage.

The Crisis:

In 2018, 3,100 Google employees signed a letter demanding Google exit - calling it a 'weapon of war'. Google ultimately did not renew the contract.

Then came Palantir:

Palantir & Anduril picked up the contract. Maven now integrates AI decision-support across CENTCOM, INDOPACOM and EUCOM.

The real danger:

AI misidentification in conflict zones has no 'Ctrl+Z'. One false positive = civilian casualties.

LAWS - LETHAL AUTONOMOUS WEAPONS

Kargu-2 Drone (Turkey, 2020):

UN experts reported the Kargu-2 autonomously hunted & attacked Libyan combatants - the first known LAWS use in combat.

30+ nations

now deploy or develop LAWS. Zero binding international treaty exists.

WHY THIS IS DANGEROUS

1. No moral accountability - machines cannot be tried for war crimes
2. Algorithmic bias can systematically target ethnic or demographic groups
3. Sub-second kill decisions remove human judgment entirely
4. Proliferation to non-state actors & terrorist groups is inevitable

Sources: The Intercept (2018 Maven), UN Panel of Experts on Libya (2021 Kargu-2), US DoD AI Strategy 2019

CYBER WARFARE: DATA IS THE NEW AMMUNITION

STUXNET (2010) – Iran Nuclear Plant

The first known cyber-weapon used in physical warfare. US–Israel joint operation infected Natanz uranium enrichment facility, destroying ~1,000 centrifuges. AI-assisted targeting of industrial SCADA systems. Proved cyber = kinetic damage.

Ref: Symantec W32.Stuxnet Dossier (2011); Kim Zetter, 'Countdown to Zero Day'

Ukraine Power Grid (2015–16) – BlackEnergy & Industroyer

Russian-linked APT used AI-enhanced malware to knock out electricity for 230,000 Ukrainians in winter. Industroyer (2016) was the first malware specifically designed to disrupt industrial control systems.

Ref: ESET Industroyer Analysis (2017); CISA Alert AA22-110A

AI-Powered Spear Phishing & Reconnaissance

Nation-state actors (China, Russia, Iran) now use LLM-based tools to craft hyper-personalised phishing emails at scale and automate OSINT gathering on military personnel. Pace of attack has increased 10×.

⚡ **Critical Infrastructure Attacks using AI jumped 300% between 2020–2024 (IBM X-Force Threat Intelligence Index 2024)**

AI, DEEPFAKES & COGNITIVE WARFARE

REAL-WORLD CASE STUDIES

Ukraine President Zelensky Deepfake (Mar 2022)

A deepfake video appeared showing Zelensky ordering Ukrainian soldiers to surrender. Circulated on social media & hacked TV stations. Designed to demoralise troops and cause battlefield confusion.

Ref: Reuters Fact Check (Mar 16, 2022)

Slovakian Election Interference (Sep 2023)

48 hours before Slovakia's election, AI-generated audio of opposition leader Michal Šimečka was released, falsely depicting him discussing vote-rigging. Too late to debunk. His party lost.

Ref: AFP Fact Check; EU DisinfoLab Report (Oct 2023)

Hamas–Israel Info War (Oct 2023–2024)

Both sides deployed AI-generated images, fabricated casualty counts, and synthetic social media accounts. Gaza hospital blast: within hours, AI-generated images from both sides flooded feeds, making ground truth impossible.

Ref: Stanford Internet Observatory (2024); Graphika Report

HOW IT WORKS

Synthetic Media

GAN & diffusion models generate photo-realistic fake videos, voice clones & fabricated documents.

Narrative Injection

LLMs write thousands of coordinated social media posts amplifying false military narratives.

Cognitive Overload

Flooding information channels makes distinguishing truth impossible - 'liar's dividend'.

Targeted Psyops

AI micro-targets soldiers, civilians or leaders with personalised fear/demoralisation campaigns.

AUTOMATION: REMOVING HUMANS FROM THE KILL CHAIN

FIND	FIX	TRACK	TARGET	ENGAGE	ASSESS
AI surveillance satellites, SIGINT drone swarms	Target tracking & predictive movement modelling	Persistent ISR without human oversight	AI nominates targets with probability scores	Autonomous or semi-autonomous weapons fire	AI BDA - battle damage assessment in real time

⚠ **Steps 4 & 5 are now being automated - a machine decides who lives and who dies**

Aegis Combat System (US Navy)

Automated surface-to-air missile system. 1988: USS Vincennes - Aegis shot down Iran Air Flight 655, killing 290 civilians. System misidentified civilian Airbus as military jet.

Ref: NTSB / US Navy official investigation records

AI-enabled Loitering Munitions - Ukraine 2022–24

Both Russia (Lancet) and Ukraine (Switchblade 600) use AI-guided loitering munitions. Lancet destroyed 400+ armoured vehicles. Human pilot is miles away or not in the loop at all.

Ref: Oryx open-source tracking; Breaking Defense reports

The Gospel AI System (Israel)

IDF AI nominates targets at rate of 100/day vs 50/year manually. Officers reportedly approve strikes in 20 seconds with minimal review. Raised serious IHL concerns.

Ref: +972 Magazine / Local Call investigation (Apr 2024)

GEOPOLITICAL & ECONOMIC SHOCKWAVES

$900B+
Global AI defence market by 2030

36+
Nations developing military AI programs

10×
Speed advantage: AI targeting vs manual

0
Binding international AI warfare treaties

AI ARMS RACE

US vs China:

Both nations have declared AI supremacy a national security priority. China's 'New Generation AI Development Plan' (2017) explicitly targets military dominance by 2030.

Russia:

Putin (2017): 'Whoever leads in AI will rule the world.' Russia has invested in Uran-9 unmanned ground combat vehicles and AI-enhanced nuclear command systems.

Small-State Threat:

AI democratises destructive capability. A small nation or non-state actor with access to off-shelf models can now build targeted cyberweapons or drone swarms at minimal cost.

ECONOMIC DISRUPTION

Defence AI spending displaces:

Social welfare, healthcare and education budgets in developing nations whose militaries are pressured to 'keep up' with AI-capable powers.

Job Displacement:

Automated warfare reduces demand for infantry but creates new classes of 'AI soldiers' - accelerating inequality in who bears the physical risk of war.

Private Sector Power:

Companies like Palantir, Anduril, Rheinmetall & Clearview AI are becoming de-facto arbiters of life-and-death military decisions - outside democratic oversight.

Sources: SIPRI, McKinsey Global AI Report 2024, Rand Corporation 'AI and International Stability', China State Council AI Plan 2017

IRAN–ISRAEL: AI ON THE MODERN BATTLEFIELD

THE GOSPEL (Habsora)

AI system that automatically generates bombing targets in Gaza. Generates ~100 targets/day. IDF officers reportedly approve strikes in under 20 seconds with minimal independent review.

Enabled highest-density bombing campaign in modern history. 40,000+ casualties in first year. Human Rights Watch flagged AI targets for potential war crimes.

Ref: +972 Magazine, Local Call (Apr 2024) - 'Lavender: The AI Machine Directing the Gaza Bombing Campaign'

LAVENDER AI

Machine learning system that assigns each person in Gaza a 1–100 score indicating likelihood of being a militant. IDF used this to pre-approve lethal strikes.

System had ~10% error rate. Policy: for every junior Hamas militant, 15–20 civilian deaths considered acceptable. AI error × policy = mass civilian casualties.

Ref: +972 Magazine investigative report (Apr 2024) - based on testimonies of IDF intelligence officers

IRAN'S COUNTER-AI

Iran deployed AI-enhanced drone swarms in April 2024 (first state-vs-state aerial drone attack). 300+ drones + cruise missiles launched at Israel. AI guided flight path evasion.

Israel's Iron Dome + AI interceptors shot down 99% of the attack. First large-scale AI-drone vs AI-defence battle in history.

Ref: BBC, Reuters, Times of Israel (Apr 14, 2024); IISS Military Balance 2024

This war is the first conflict where AI has been used end-to-end: intelligence, targeting, strike, defence, and information warfare - simultaneously.

OPENAI × PENTAGON: THE DEAL & ITS IMPLICATIONS

TIMELINE OF POLICY REVERSAL

2023

OpenAI Usage Policy explicitly prohibited military use, weapons development, and cyberoffence

Jan 2024

OpenAI quietly updated its policy - removing blanket military prohibition, allowing 'national security' applications

Oct 2024

OpenAI announced partnership with Anduril (AI defence startup) to build AI-powered anti-drone systems for US military

2025

OpenAI awarded US government contracts; CEO Sam Altman cited 'democratic values' as justification

WORST CASE SCENARIOS

1. ChatGPT-level intelligence in autonomous weapons - self-reasoning kill decisions
2. Adversaries steal/replicate the model for offensive AI weapons
3. AI used to generate military disinformation at state scale
4. Precedent forces Google, Meta, Anthropic into military deals - full commercialisation of AI warfare
5. Democratic oversight collapses as military-AI decisions become too fast for humans to review

WHY THIS MATTERS

OpenAI was founded on the principle that AGI must benefit all humanity. The Pentagon deal signals a fundamental mission drift - from 'safe AI for everyone' to 'powerful AI for the highest bidder.'

Once OpenAI crosses this line, every frontier AI lab faces investor & government pressure to follow. The result: the most powerful AI models on Earth are optimised for warfare, not welfare.

The public never voted on this.

Ref: OpenAI Policy Update Jan 2024; Reuters/NYT OpenAI-Anduril partnership Oct 2024; MIT Technology Review analysis

GOVERNANCE: WHAT MUST BE DONE

UN Treaty on LAWS

A binding Convention on Certain Conventional Weapons (CCW) protocol to ban fully autonomous lethal weapons. Campaign to Stop Killer Robots advocates for 'meaningful human control' requirement.

Status: Stalled - US, Russia, China blocking binding resolution

US AI Safety Order (2023)

Biden's Executive Order on AI Safety required frontier AI developers to share safety test results with government. Established AISI (AI Safety Institute). Trump administration partially reversed this in 2025.

Status: Partially reversed - regulatory uncertainty

EU AI Act (2024)

World's first comprehensive AI law. Classifies AI in military/biometric targeting as high-risk. BUT: explicitly exempts 'national security' applications - the biggest gap.

Status: Enacted, but military AI exempted

Bletchley Declaration (2023)

28 nations (incl. US, UK, China, India) signed declaration acknowledging frontier AI risks. Established a global AI Safety Summit process.

Status: Non-binding, but historic multilateral acknowledgement

WHAT MUST CHANGE

Mandatory human-in-the-loop for lethal decisions · Open-source AI safety audits for defence contracts · International kill-switch protocols · Ethical AI curriculum in defence academia · Whistleblower protections for AI engineers

THE QUESTION WE MUST ANSWER

Will we control AI - or will AI Control the war?

AI is not inherently evil - but in the hands of states and corporations without accountability, it becomes the most efficient engine of human destruction ever built.

1. Demand international treaties on Lethal Autonomous Weapons
2. Hold AI companies legally accountable for military applications
3. Support whistleblowers inside defence AI programmes
4. Push for AI ethics as a compulsory course in every engineering & CS degree