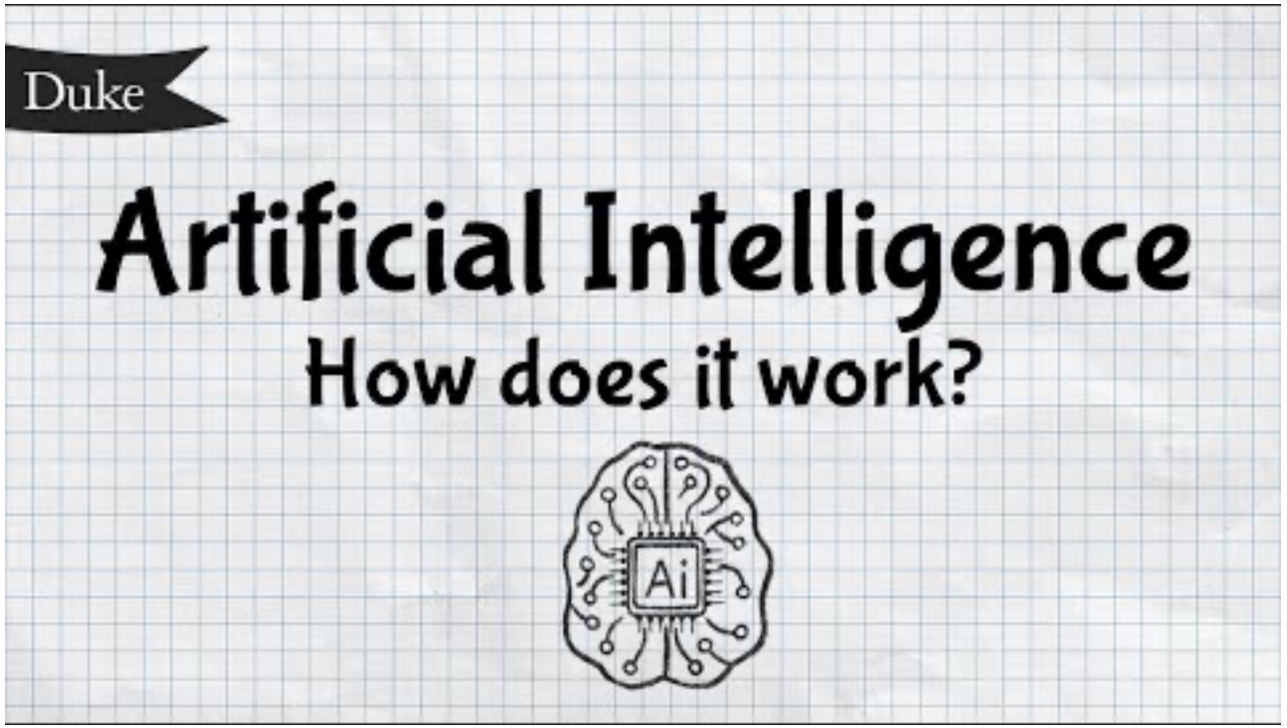
שיעור 10

מדיניות בינה מלאכותית בארגונים ומסגרת ה-NIST AI RMF 1.0

מה זה AI? הסברים במספר רמות: 1



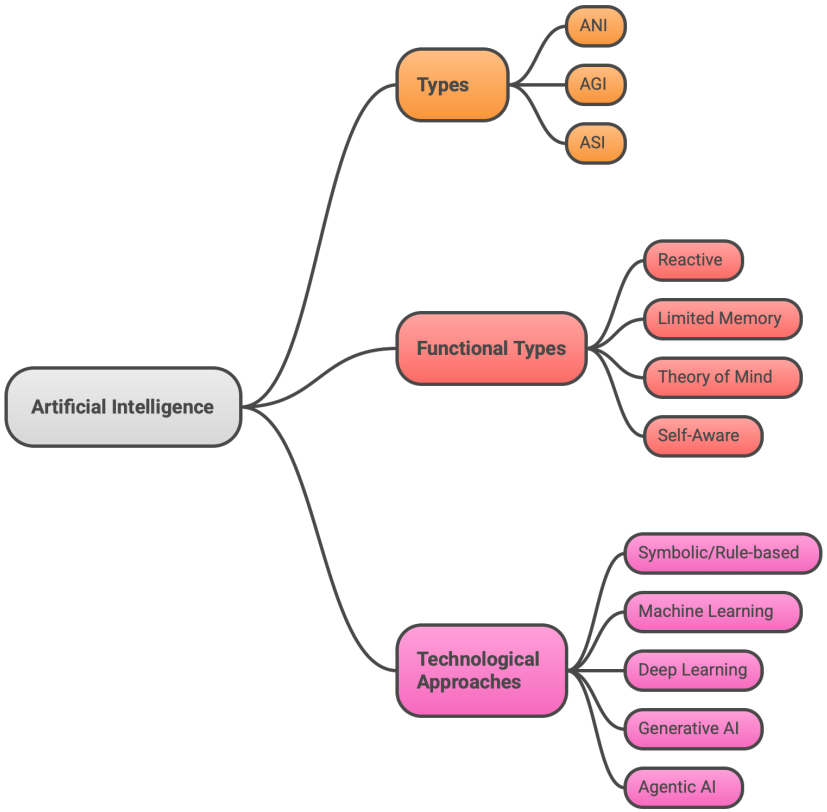
https://www.youtube.com/watch?v=com6yaGlZh4

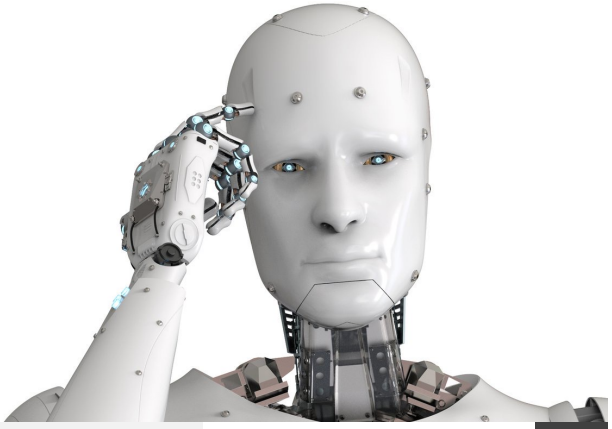
מה זה AI? הסברים במספר רמות: 2

https://www.youtube.com/watch?v=qYNweeDHiyU

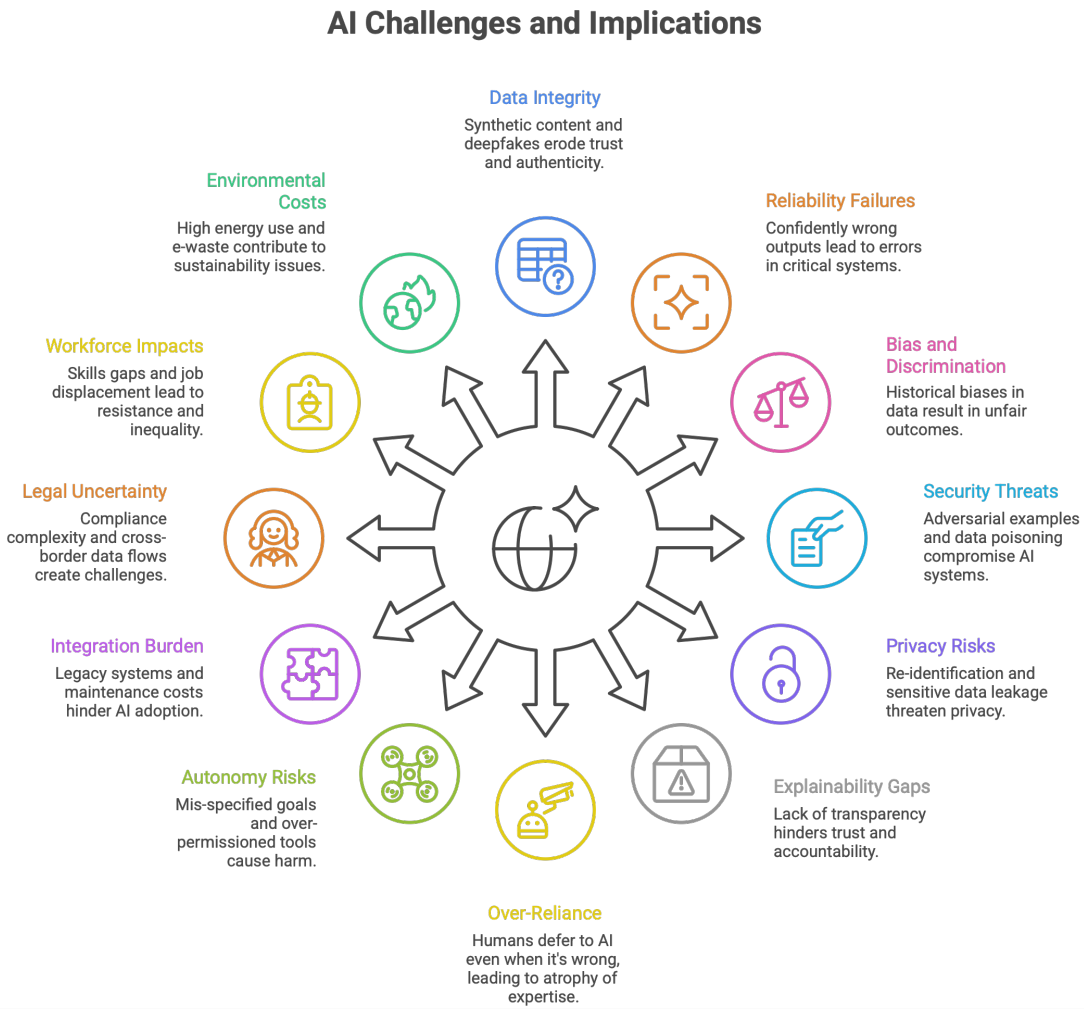
מה זה AI? הסברים במספר רמות: 3

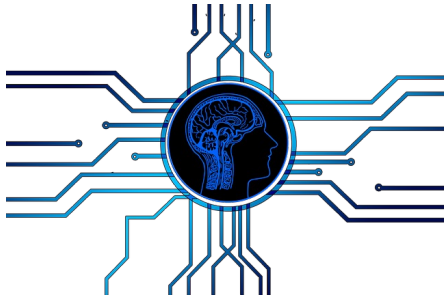
Artificial Intelligence: Types, Functions, and Technologies



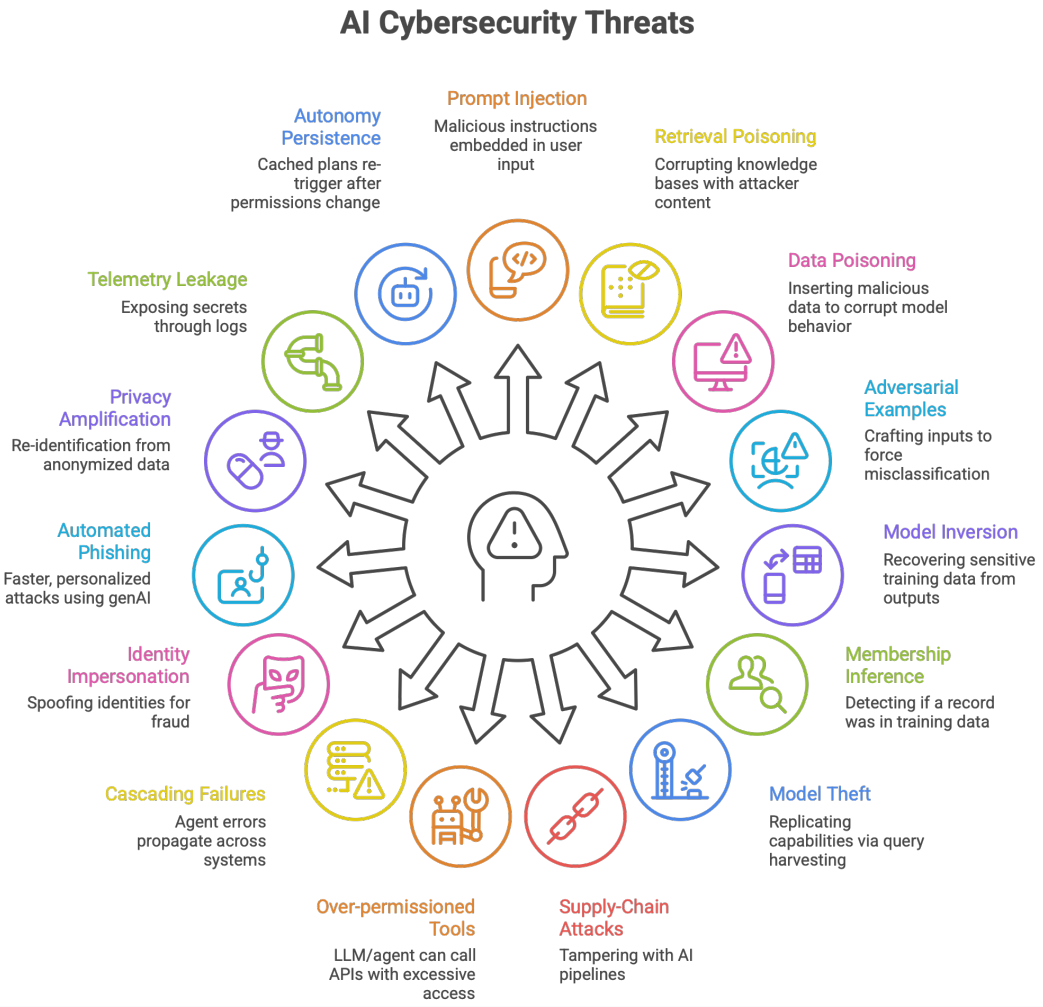


בעיות, סיבוכים וסיכונים





בעיות, סיבוכים וסיכונים



בעיות, סיבוכים וסיכונים

https://www.youtube.com/watch?v=6BYpdBx4Lyg

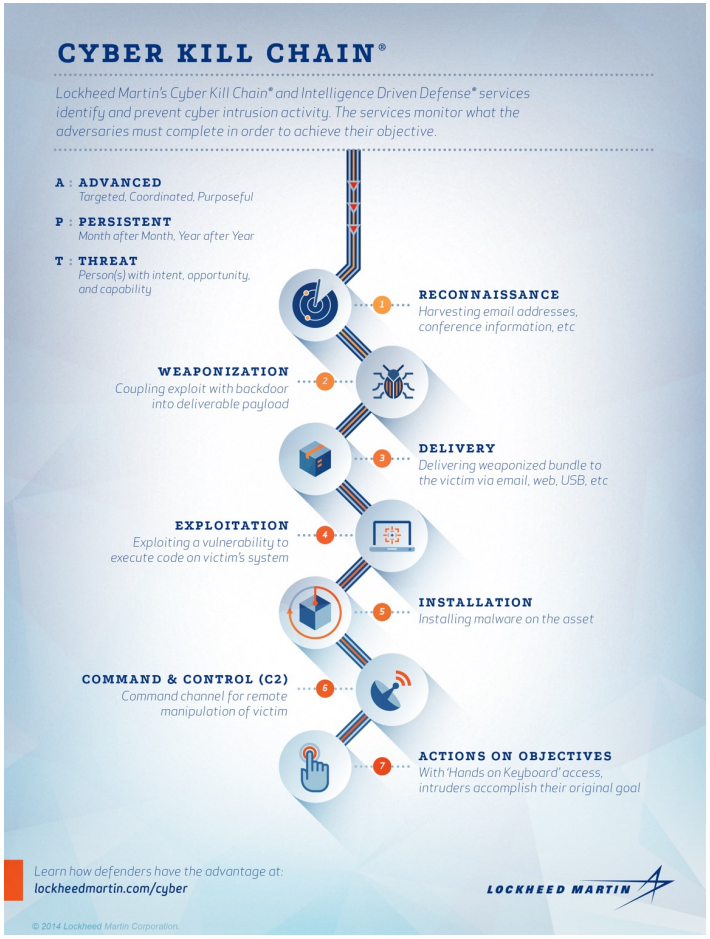
בעיות, סיבוכים וסיכונים

https://www.youtube.com/watch?v=RW944FGIdwg

בעיות, סיבוכים וסיכונים

https://www.youtube.com/watch?v=1Kv5njt9X5c

בעיות, סיבוכים וסיכונים



כיצד תוקפים משלבים בינה מלאכותית?

כיצד נשלב AI כדי להגן?

- ריגול – Espionage
- תחבולה – Sabotage
- מניעת גישה – Denial of Service
- מניפולציה או שינוי של מידע/תשתית – Data & Infrastructure Modification
- תעמולה – Propaganda (Dis/Mis-Information)

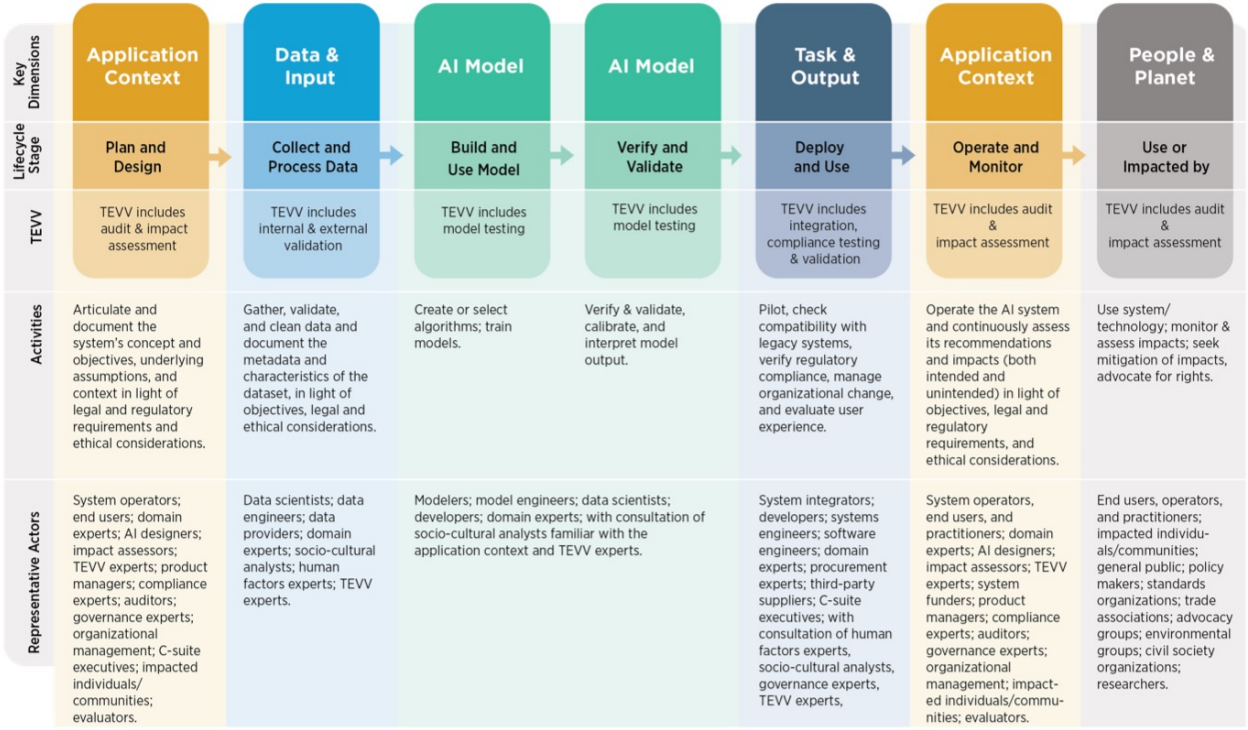
NIST AI Risk Management Framework 1.0

Fig. 1. Examples of potential harms related to AI systems. Trustworthy AI systems and their responsible use can mitigate negative risks and contribute to benefits for people, organizations, and ecosystems.

NIST AI Risk Management Framework 1.0

Fig. 2. Lifecycle and Key Dimensions of an AI System. Modified from OECD (2022) OECD Framework for the Classification of AI systems — OECD Digital Economy Papers. The two inner circles show AI systems' key dimensions and the outer circle shows AI lifecycle stages. Ideally, risk management efforts start with the Plan and Design function in the application context and are performed throughout the AI system lifecycle. See Figure 3 for representative AI actors.

NIST AI Risk Management Framework 1.0



Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Fig. 3. AI actors across AI lifecycle stages. See Appendix A for detailed descriptions of AI actor tasks, including details about testing, evaluation, verification, and validation tasks. Note that AI actors in the AI Model dimension (Figure 2) are separated as a best practice, with those building and using the models separated from those verifying and validating the models.

שיעור 11

המשך בינה מלאכותית, מדיניות ופרטיות