Principy zodpovědného užití AI v NATO

Mirek Nečas

TOVEK

find ▸
understand ▸
use ▸

Zodpovědné užití AI v NATO

- **NATO Artificial Intelligence Strategy**
  - **NIAG SG 252**: Emerging and Disruptive Technologies in the context of Emerging Powers

- **DARB: Data and Artificial Intelligence Review Board**
  - **NIAG SG 279**: Protocols and Standards to Certify Applications using AI within NATO

NIAG: NATO Industrial Advisory Group

www.diweb.hq.nato.int

Co je AI?

- Definice NATO
  - **Artificial Intelligence** (AI): Schopnost strojů vykonávat úlohy, které obvykle vyžadují lidskou inteligenci.

- Definice EU (AI Act)
  - **Artificial Intelligence System** (AI Systém): strojový systém navržený aby pracoval s proměnnou mírou samostatnosti, schopný vytvářet výstupy, jako jsou předpovědi, doporučení nebo rozhodnutí, které ovlivňující fyzické či virtuální prostředí s cílem dosáhnout explicitních či implicitních cílů.
  - **High-risk AI system**: AI systém, u kterého existuje významné riziko poškození zdraví nebo narušení bezpečnosti a základních práv osob v EU.

Principy zodpovědného užití (PRUs)

- **NATO** 6 principů zodpovědného využívání AI: Lawfulness, Responsibility and Accountability, Explainability and Traceability, Reliability, Governability, Bias Mitigation

- **USA** 5 principů etického užití AI: Responsibility, Equitability, Traceability, Reliability and Governability;
- **EU** kompromisní návrh regulace Artificial Intelligence Act (AIA);
- **UK** vize implementace AI v obraně založená na 4 principech užití AI (Efficient, Effective, Trusted and Influential);
- **OECD** 5 principů etického užití AI: Human-centred values and fairness; Transparency and explainability; Robustness, security and safety; Accountability);

Studie NIAG SG 279

TOVEK

- Cíl: Návrh metodiky certifikace AI dle NATO PRUs
  - Testování na 6 příkladech užití AI v rámci NATO

- Metoda řízení rizik
  - Role AI v rámci určitého případu užití
  - Analýza existujících standardů a gap analýza
  - Identifikace rizik ve vztahu k PRUs
  - Návrh kroků k vytvoření důvěryhodného prostředí

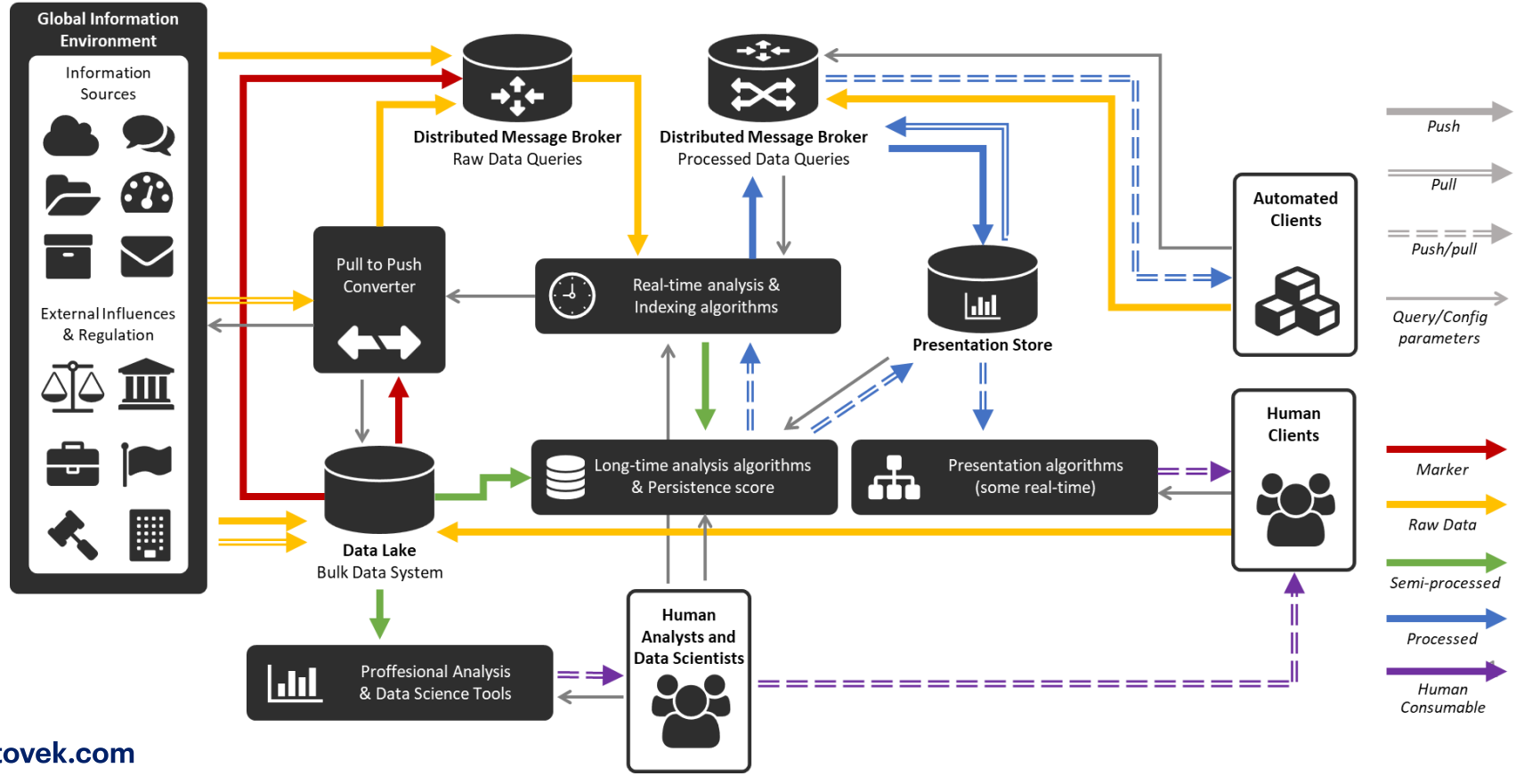
- 6 týmů podle příkladů užití (60 expertů 13+2 státy)

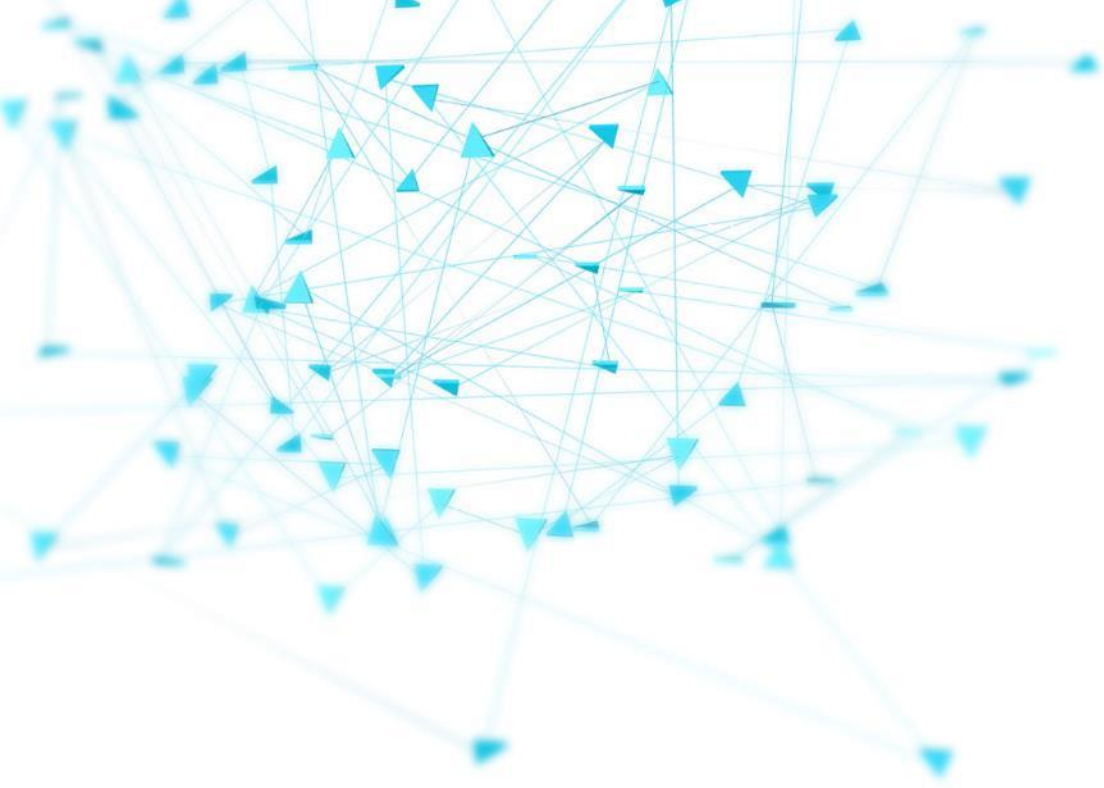
Role AI v různých příkladech užití

TOVEK

	Military Mobility	Imagery Analysis	Cyber Threat Analysis	Information Environment Assessment	Assisted *Decision Making*	Climate Analysis
Natural Language Processing			X	**XX**	**XX**	X
Computer Vision	X	**XX**		X	X	
Speech Recognition				**XX**	**XX**	
Anomaly Detection	X	X	**XX**	X	X	X
Graph-based Algorithms	**XX**		**XX**	X	**XX**	
Time Series Analysis & Forecasting	X		X	**XX**	X	**XX**
Optimization Algorithms	**XX**				**XX**	**XX**
Generative Algorithms				**XX**	**XX**	
Intelligent Agents	**XX**		X	X	X	
Data mining	X	X	X	X	X	X

Referenční architektura AI systému





TOVEK

Assisted Decision Making Use Case

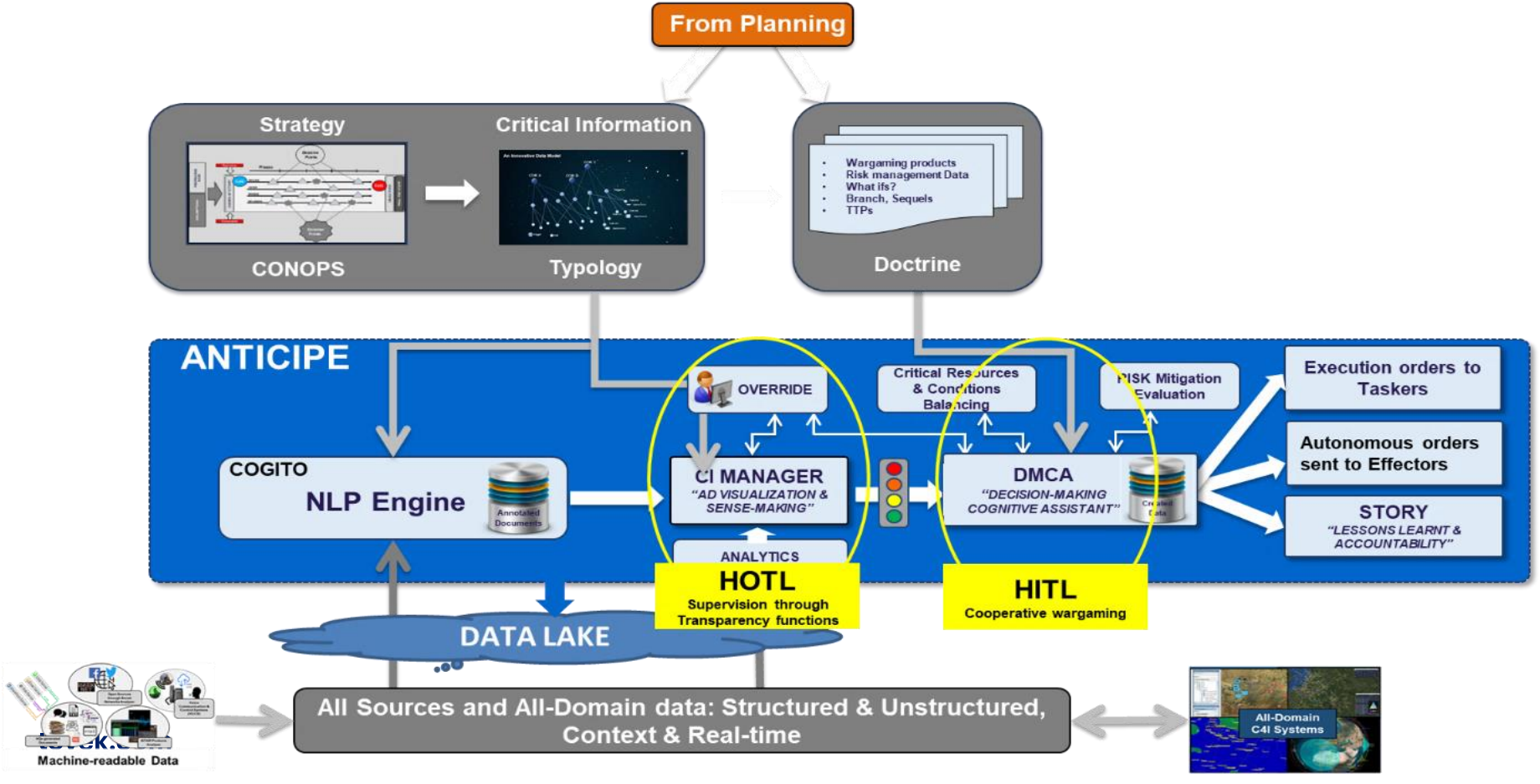
find ▸
understand ▸
use ▸

Role AI v rámci podpory rozhodování

TOVEK

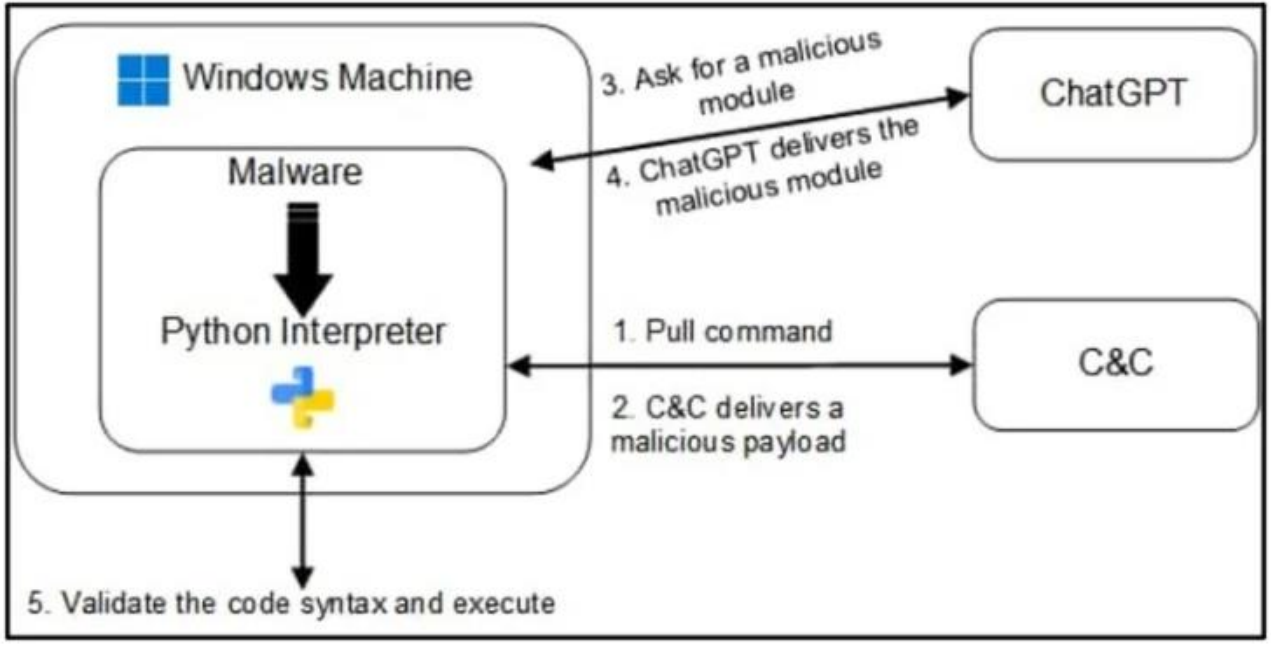
- Mentor
  - Analýza komplexního prostředí
  - Tvorba scénářů při plánování
  - Protivník ve strategických hrách
- Vykonavatel
  - Automatizace rutinních rozhodnutí
  - Předávání informací lidem/efektorům
- Cíl útoku / zranitelnost systému
  - Únik informací
  - Nepřátelské útoky

Systém ANTICIPE



Black Mamba

- Polymorfní malware využívající Chat GPT

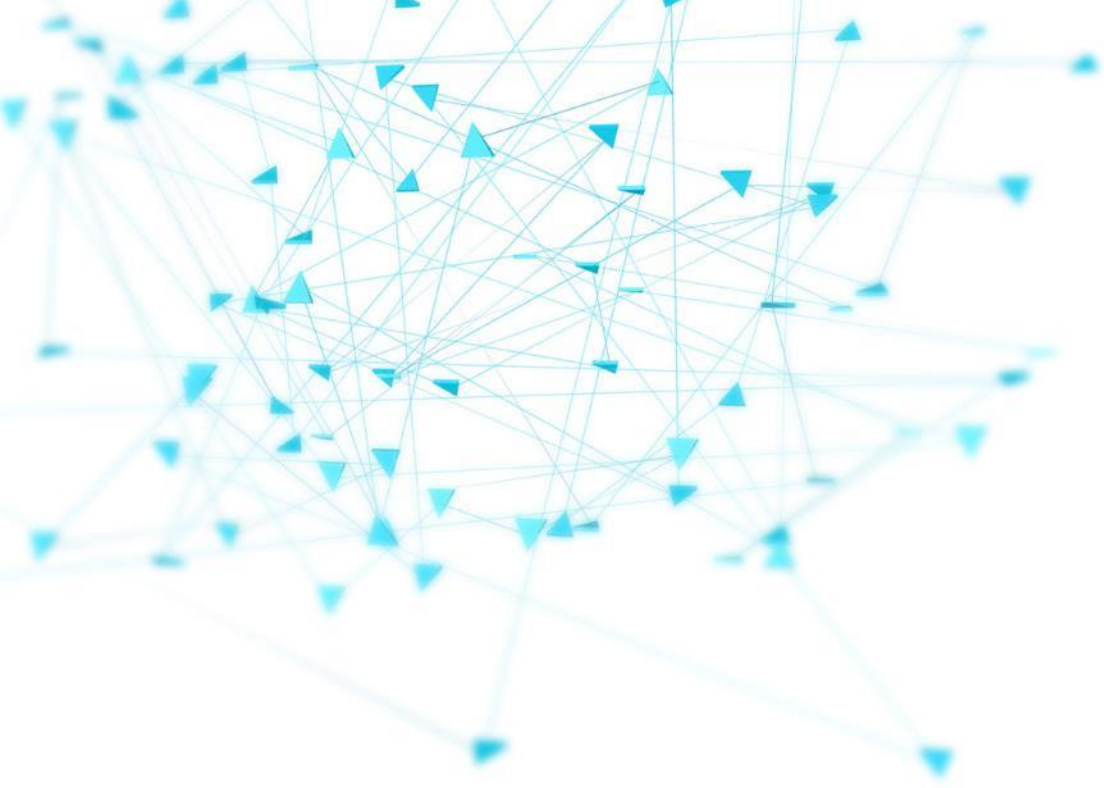


Relevantní stadndardy

TOVEK

- Pro všechny PRU
  - ISO/IEC 22989 Artificial intelligence concepts and terminology
  - ISO/IEC 23053 Framework for artificial intelligence systems using machine learning
  - ISO/IEC 8183 AI Data life cycle framework

Standard / PRU	Law	Rsp & Acc	Exp & Trc	Rel	Gov	BsM
ISO/IEC TR 24028 (Trustworthiness)			X			
ISO/IEC 24029 – 1 (NN robustness)				X		
ISO/IEC 24029 – 2 (NN robustness)				X		
ISO/IEC TR 24027 (Bias and D-M)						X
ISO/IEC 5259 – 1 (Data quality)				X	X	X
ISO/IEC 5259 – 2 (Data quality)				X	X	X
ISO/IEC 5259 – 3 (Data quality)				X	X	X
ISO/IEC 5259 – 4 (Data quality)				X	X	X
ISO/IEC 5259 – 5 (Data quality)				X	X	X
ISO/IEC 5259 – 6 (Data quality)				X	X	X
ISO/IEC 42006 (AI audit)					X	
ISO/IEC 42001 (AI management)					X	
ISO/IEC 17021-1 (conformity assessment)					X	
ISO/IEC 23894 (AI risk management)					X	
DIN SPEC 91426 (HR management)						X



TOVEK

Obecně užitečná zjištění a doporučení

find ▸
understand ▸
use ▸

Smyslem je podpořit využití AI v NATO

- První volbou pro certifikaci je použití stávajících průmyslových standardů

Stejné pojmy často mají různé významy

Principy jsou navzájem provázané

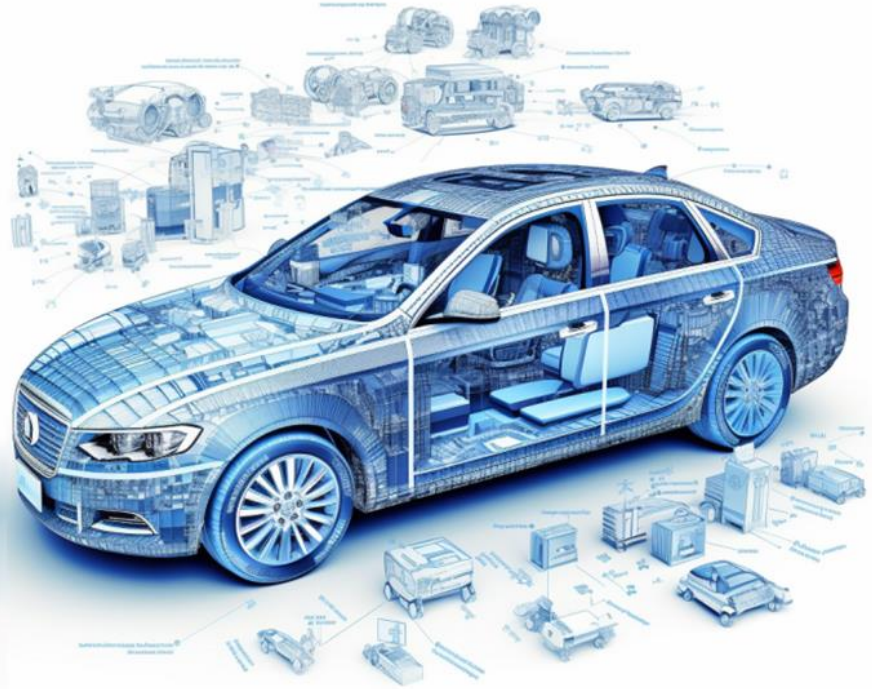
TOVEK

AI aplikace je potřeba posuzovat v kontextu užití celého systému

TOVEK

- Principy nejméně pokryté standardy:
  - Responsibility and Accountability
  - Lawfulness

- Co s tím?
  - Vnitřní předpisy a vzdělávání
  - Praktická zkušenost a cvičení



TOVEK

AI či ne AI?

jak zodpovědně implementovat umělou inteligenci,

to je, oč tu běží

TOVEK

Děkuji vám za pozornost!

Miroslav Nečas

necas@tovek.cz