MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

FUTURE IS AI

The Artificial Intelligence National Laboratory (MILAB) aims to strengthen Hungary's position in AI.

mi.nemzetilabor.hu

Artificial Intelligence National Laboratory

„Security and Privacy” subproject

Rudolf FERENC, PhD

November 11, 2021

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Cooperating MILAB partners

- **Budapest University of Technology and Economics – BME**
- **Special Service For National Security – NBSZ**
- **Institute for Computer Science and Control – SZTAKI**
- **University of Szeged – SZTE**
- **Centre for Social Sciences – TK**

MŰEGYETEM 1782

1996 NEMZETBIZTONSÁGI SZAKSZOLGÁLAT

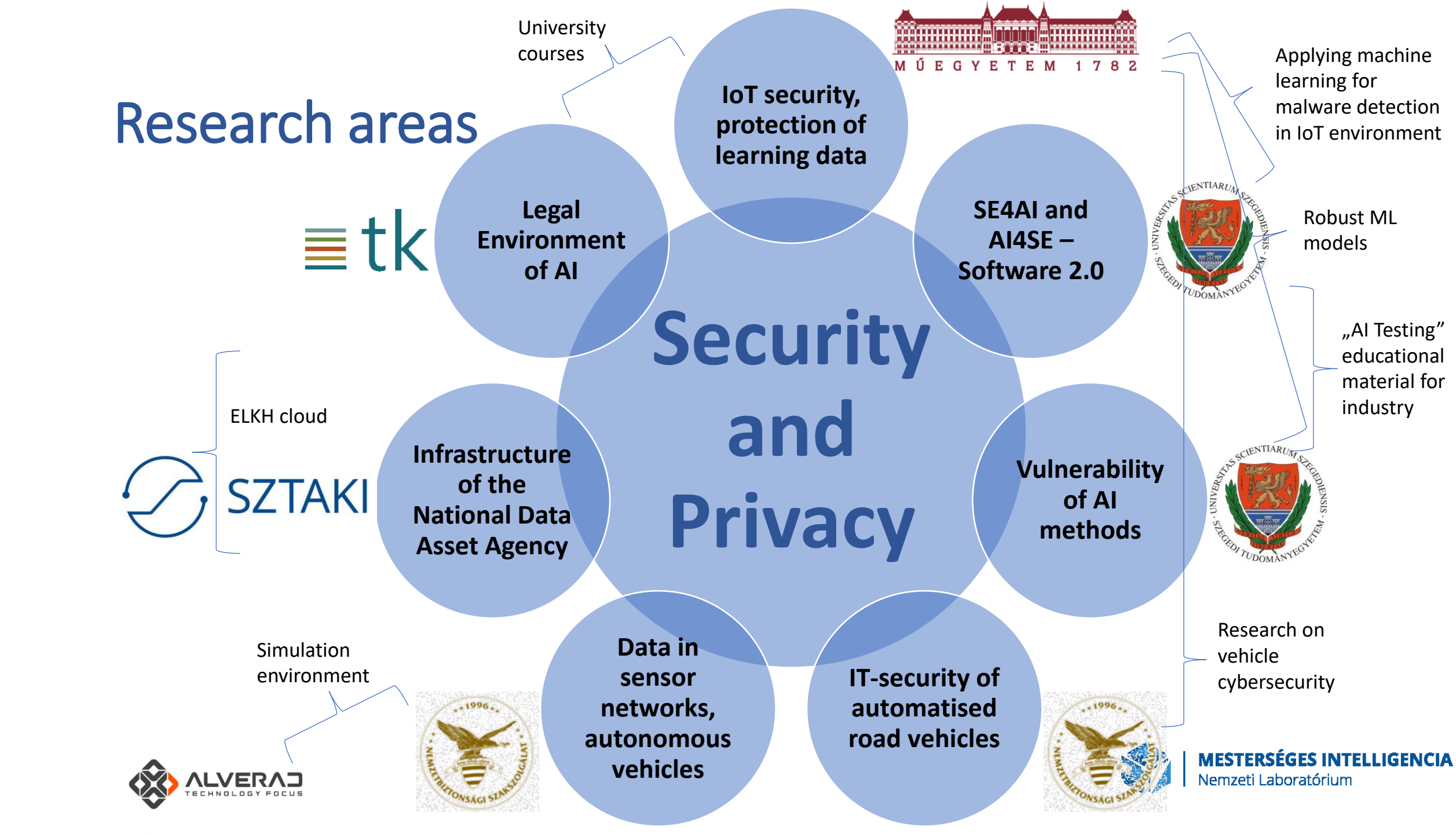
SZTAKI

UNIVERSITAS SCIENTIARUM SZEGEDIENSIS · SZEGEDI TUDOMÁNYEGYETEM ·

tk

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Research areas



Protection of learning data

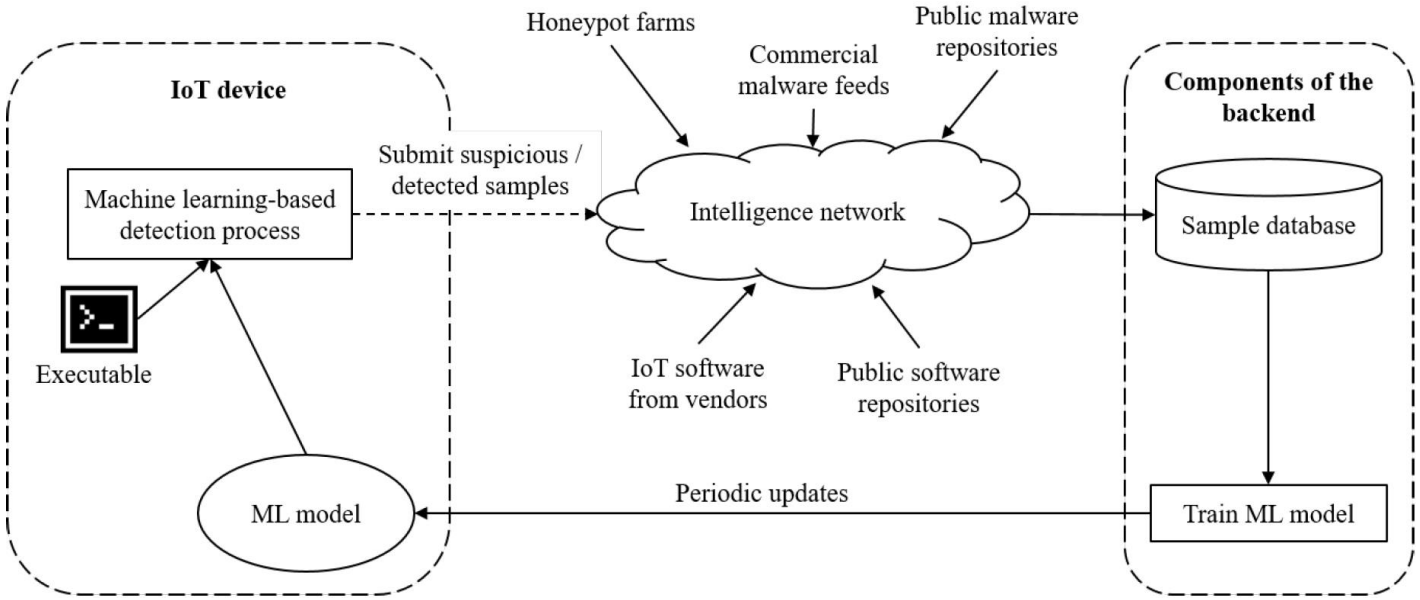
- Data protection in federated learning
  - Recovering gradients from aggregated gradients
  - Plans
    - ***Measuring contribution of participants*** – detecting free-riding and poisoning
    - ***Designing safe aggregation protocols*** – implementing weighted sum
- Anonymization of location data
  - Plans
    - ***Applying neural generative models***
- Robust ML models, detecting adversarial patterns
  - Plan
    - ***Detect robust features*** and use these during classification
    - [Possible cooperation with SZTE]

IoT security

- SIMBIoTA – malware detection method for IoT devices
  - Based on binary similarity instead of traditional signature database
  - Representative samples from clustered malware dataset
  - Database composed of fuzzy hash codes of the representative samples (size ~10 KB)
  - If a new file's fuzzy hash is similar to fuzzy hash in the database -> malware
  - 90% true positive detection, 0% false positives
- Plans
  - ***Develop SIMBIoTA/ML – Applying machine learning for malware detection***
  - ***Create adversarial examples for SIMBIoTA and SIMBIoTA/ML***
  - ***CFG-based malware detection with ML***

Applying machine learning for malware detection

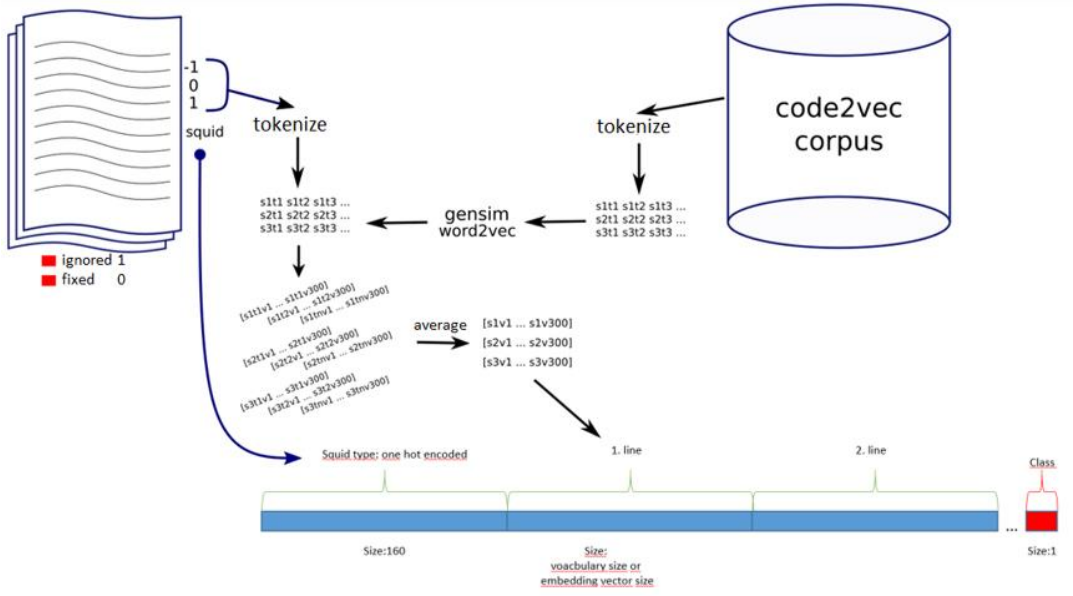
- *[BME-SZTE common project]*
- Instead of using traditional malware signature-based database
- Train a ML model on the back end on both malicious and benign samples
  - Dataset needs to be created
- Transfer the model to the IoT device
- Detect malware using the ML model
- Plans
  - ***Create dataset***
  - ***Perform thorough ML process to find the best ML model using the Deep-Water Framework***



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

SE4AI and AI4SE – Software 2.0

- SE4AI – **S**oftware **E**ngineering methods for **AI**
  - Development tools, testing tools, ML Ops workflow, …
- AI4SE – **AI** methods for **S**oftware **E**ngineering
  - Bug prediction, automatic error correction, …
- AI4SE project:
  Applying NLP methods for programming languages
  - Large corpus of Java programs
  - Java language model based on embedding
  - Filtering the results of static code analyser tools
  - Train ML model on true and false positives
  - Plan
    Improve ML model which already filters
    78% of false positives and keeps almost all true positives



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

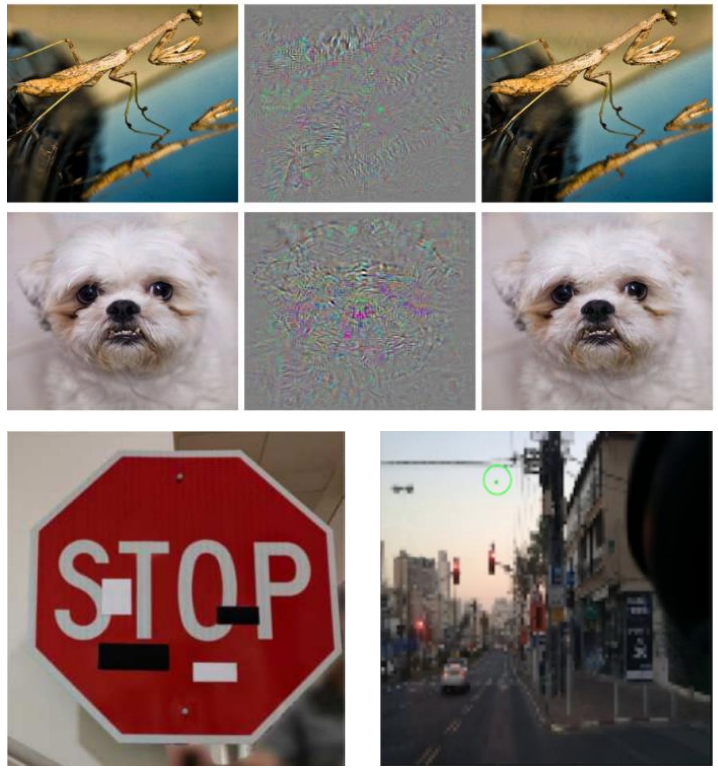
„AI Testing” educational material for industry

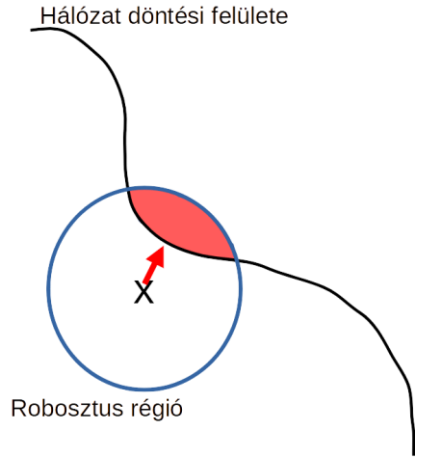
- *[SZTE-SZTE common SE4AI project]*
- <u>SE4AI project:</u>
- Plan
  - ***Create a thorough educational material (lecture and practice) for „AI Testing” suitable to hold courses for the industry and at the same time useful for university courses***

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Vulnerability of AI methods

- Adversarial perturbations
  - The rightmost column is Ostrich
  - Stickers: 45 mph speed limit
  - One pixel: Green light
- Formal verification
  - Improve known formal verification methods
    - Optimisation methods, interval methods
  - Vulnerabilities of formal verification
    - Complexity, number representation
    - Adversarial models against formal verification





MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Vulnerability of AI methods

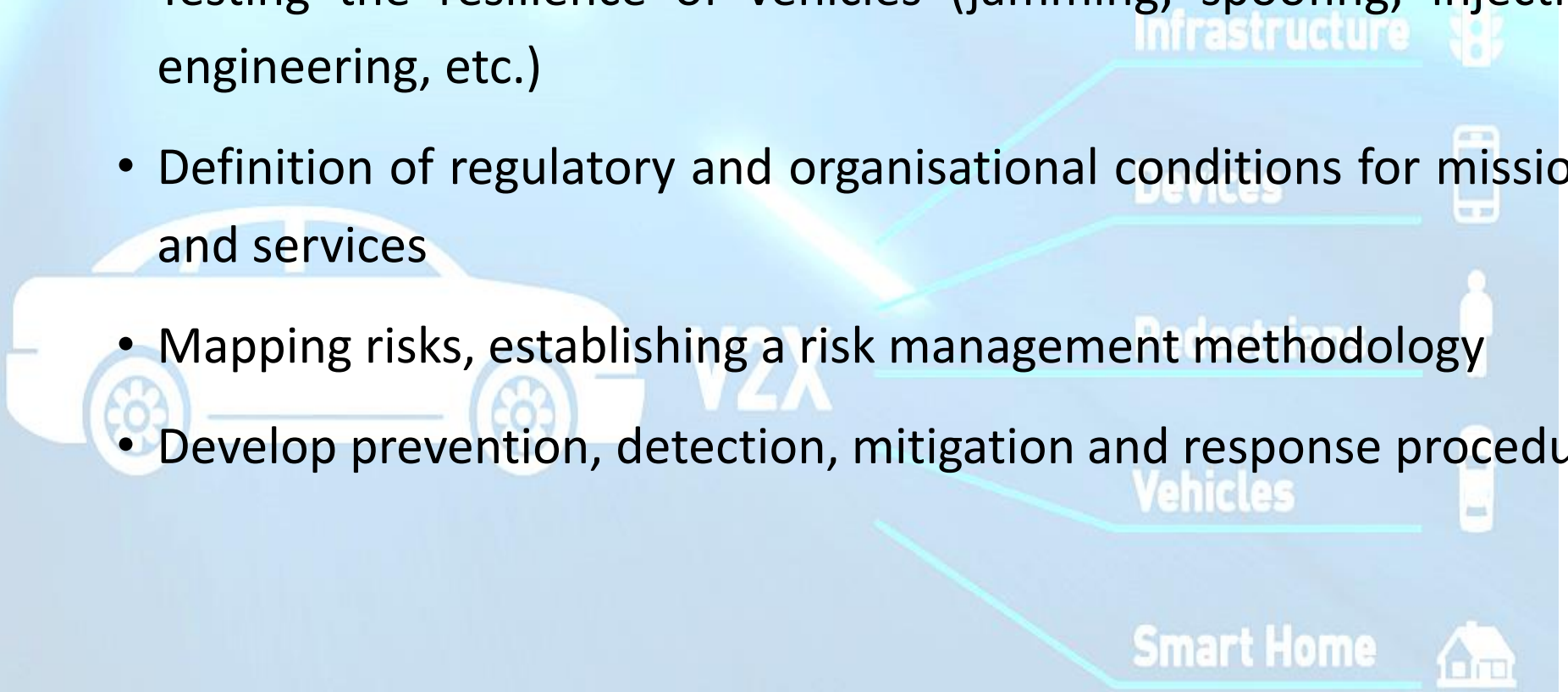
- Heuristic verification
  - Vulnerabilities (unreachable adversarial subspace)
  - Attacks against heuristic verification
  - Innovative heuristic searches (matching subspaces)
    - Also for the “nothing” (OOD) and “something” (ID) examples
- Characteristics of robust networks, knowledge representation
  - The evolution of the neural network during learning
  - Characteristic structural features
  - Increasing the effectiveness of robust teaching
  - [Possible cooperation with BME]

MŰEGYETEM 1782

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

IT-security analysis of automatised road vehicles

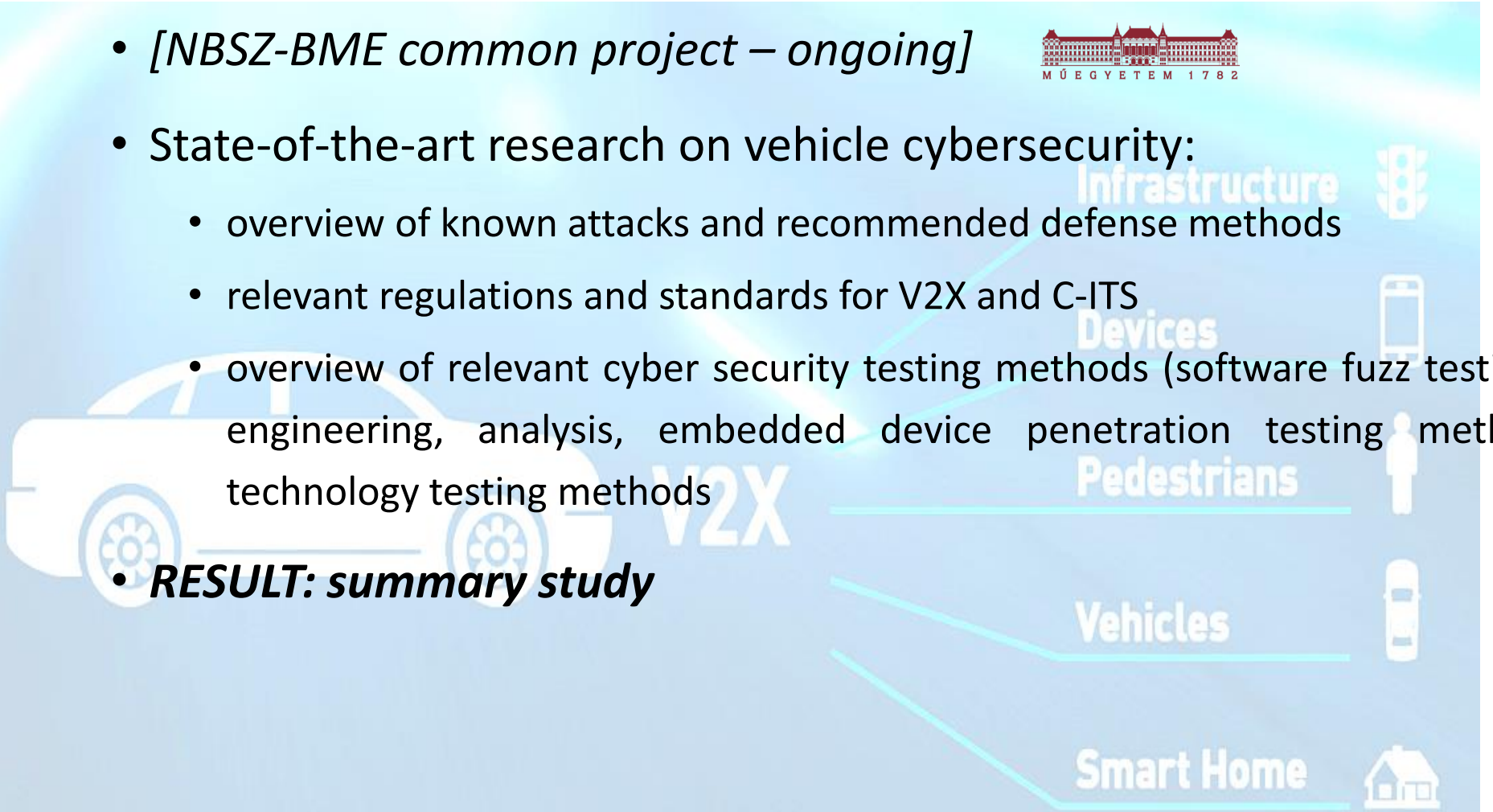
- Identification of risk-critical system components
- Testing the resilience of vehicles (jamming, spoofing, injection, DoS, reverse engineering, etc.)
- Definition of regulatory and organisational conditions for mission-critical systems and services
- Mapping risks, establishing a risk management methodology
- Develop prevention, detection, mitigation and response procedures



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

IT-security analysis of automatised road vehicles

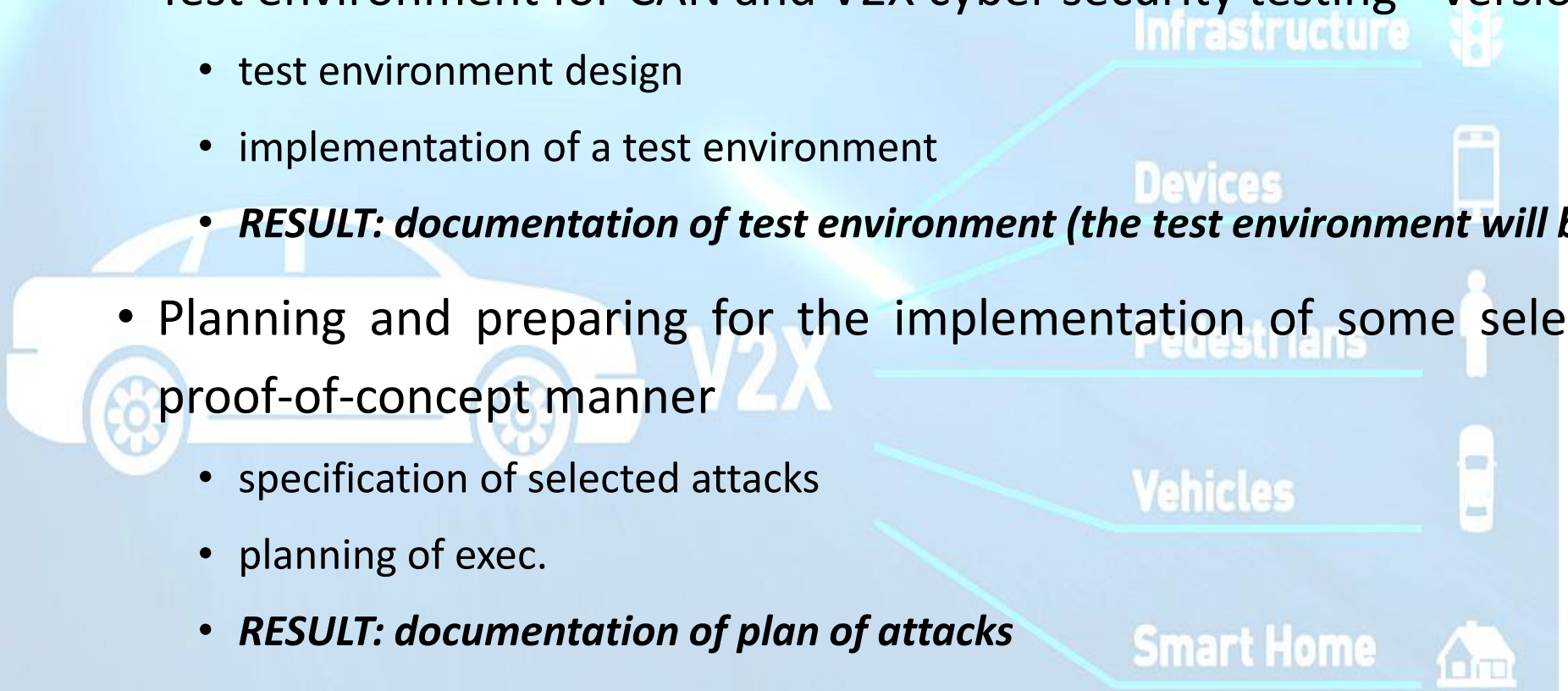
- *[NBSZ-BME common project – ongoing]*
- State-of-the-art research on vehicle cybersecurity:
  - overview of known attacks and recommended defense methods
  - relevant regulations and standards for V2X and C-ITS
  - overview of relevant cyber security testing methods (software fuzz testing, protocol reverse engineering, analysis, embedded device penetration testing methods, V2X security technology testing methods
- ***RESULT: summary study***



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

IT-security analysis of automatised road vehicles

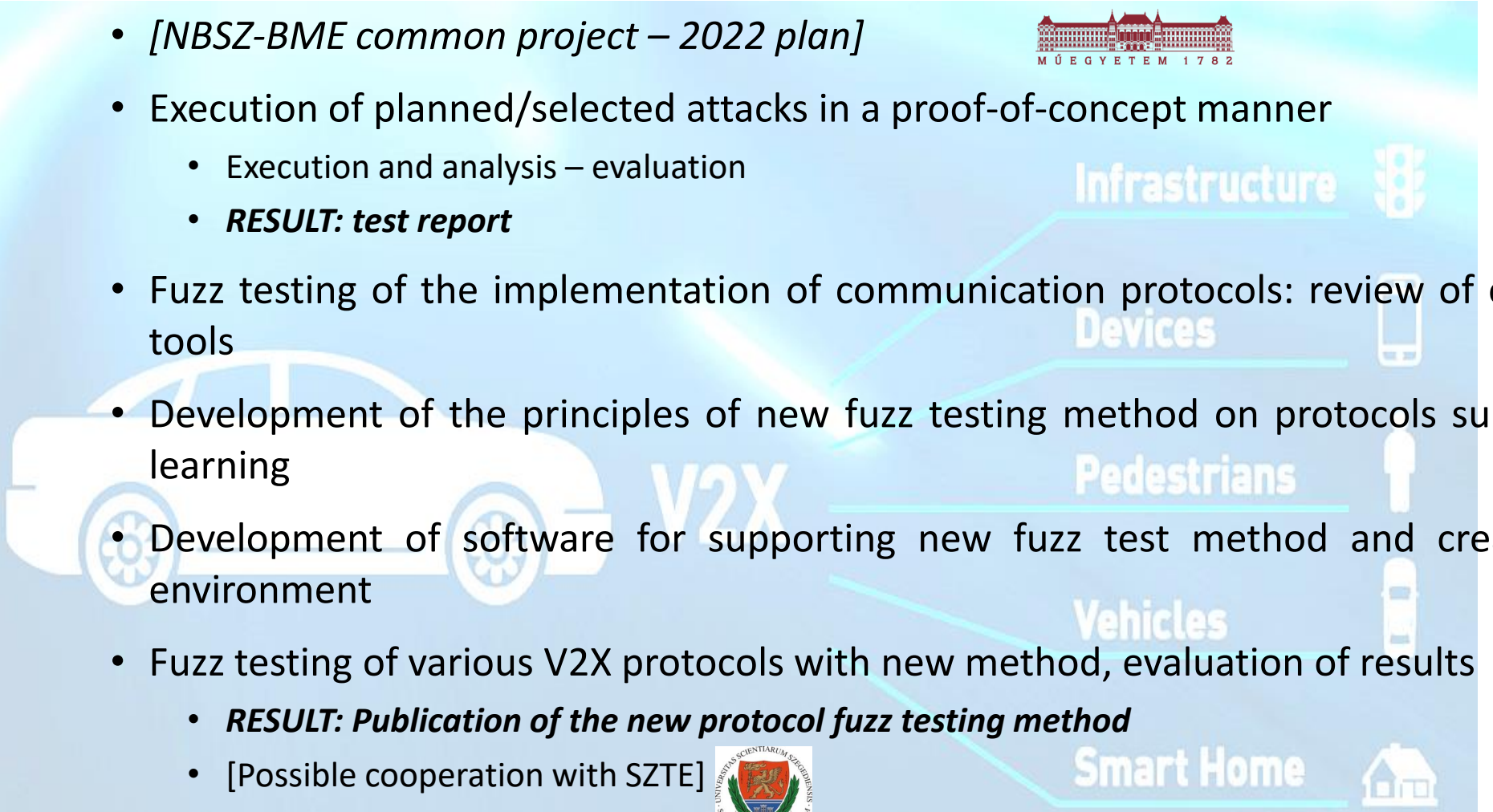
- *[NBSZ-BME common project – 2021 plan]*
- Test environment for CAN and V2X cyber security testing - Version 1
  - test environment design
  - implementation of a test environment
  - ***RESULT: documentation of test environment (the test environment will be set up at BME)***
- Planning and preparing for the implementation of some selected attacks in a proof-of-concept manner
  - specification of selected attacks
  - planning of exec.
  - ***RESULT: documentation of plan of attacks***



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

IT-security analysis of automatised road vehicles

- *[NBSZ-BME common project – 2022 plan]*
- Execution of planned/selected attacks in a proof-of-concept manner
  - Execution and analysis – evaluation
  - ***RESULT: test report***
- Fuzz testing of the implementation of communication protocols: review of exiting methods and tools
- Development of the principles of new fuzz testing method on protocols supported by machine learning
- Development of software for supporting new fuzz test method and creation of application environment
- Fuzz testing of various V2X protocols with new method, evaluation of results
  - ***RESULT: Publication of the new protocol fuzz testing method***
  - [Possible cooperation with SZTE]



MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Data in sensor networks, autonomous vehicles

- Subsequent extraction of forensic data from computers in self-driving or smart vehicles
- Define environment for modelling C-ITS (Cooperative Intelligent Transport Systems)
  - Components, data communication, protocols, data categories (creates, stored, sent, …)
- Plan:
  - ***Identification of possible ways of extracting data***
    - Data stored by vehicles using C-ITS technology
    - Tools for data extraction (non-destructive and destructive)
    - Intentionally malicious communications
  - ***Establish simulation environment***
    - Own or rented
- [Cooperation with Alverad Technology Focus Kft.]
- [Possible cooperation with another company]

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Infrastructure of the National Data Asset Agency

- Participation in the design and prototyping of the public data portal and the academic-public administration hybrid cloud
- [Cooperation with National Data Asset Agency]

Legal Environment of AI

- Identification of problems in the current and future legal environment related to AI
- Carrying out basic legal research with the aim of mapping out possible directions using legal comparison methods, in particular the legal framework developed by the EU
  - protection of basic human rights
  - data protection and cybersecurity
  - liability issues and intellectual property law
  - EU and international law, Artificial Intelligence Act
- Formulation of concrete, enforceable and textual legal recommendations for the legislator in line with international law practice
- National and international programme series (AI and Law)
- [Possible cooperation with BME – University courses for IT students]

MŰEGYETEM 1782

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

Thank you very much for your attention

„Security and Privacy” subproject

MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium