







M4DT DAY

ARTIFICIAL INTELLIGENCE EVALUATION & CERTIFICATION AT LNE

Guillaume Avrin – Head of the « AI Evaluation » Department

LABORATOIRE NATIONAL DE MÉTROLOGIE ET D'ESSAIS

30/9/21

SCOPE

Objective: to identify the challenges and means of evaluating AI systems during the development, certification and maintenance phases

AI requirements and evaluation needs

- Evaluation criteria
- Regulations in force

LNE resources and activites for conformity assessment of AI

- Evaluation standards
- Test on dataset, simulator or physical environment
- Certification

FRENCH NATIONAL LABORATORY FOR METROLOGY AND TESTING

Created in 1901

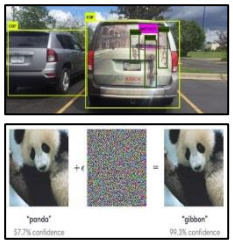
- LNE vocation is:
  - to **coordinate French metrology**;
  - to be the national reference laboratory for **testing, metrology and certification activities**;
  - to continue its scientific and technical development in order to anticipate the needs for **characterization and certification related to new products and technologies**;
  - to provide technical assistance for the development of new **regulations and standards** at international, European and national levels.

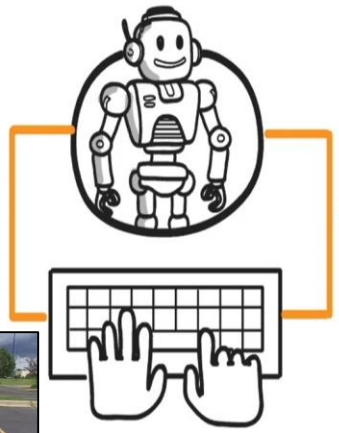
- Leading multidisciplinary testing and analysis facilities in Europe with 55,000m² of lab space.
- Turnover of 80M€ including 20% for R&D, 800 employees.

MATCHING AI SUPPLY AND DEMAND

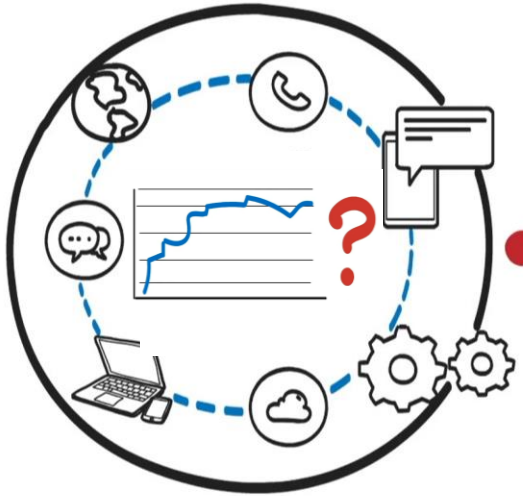
An AI evaluation must be set up to allow the supply to best meet the demand.

AI supply





Black-box, non convex, evolutive systems



Need: evaluation of AI

AI demand

Trustworthy and efficient functionalities

NEED FOR AI EVALUATION – EXAMPLE IN INDUSTRIAL ROBOTICS

NLP:
- personal assistant
- voice control
- speaker rocognition

Data analysis:
- predictive and preventive maintenance
- robo-advisor

Human-machine interaction:
- collaborative robot
- conversational agent

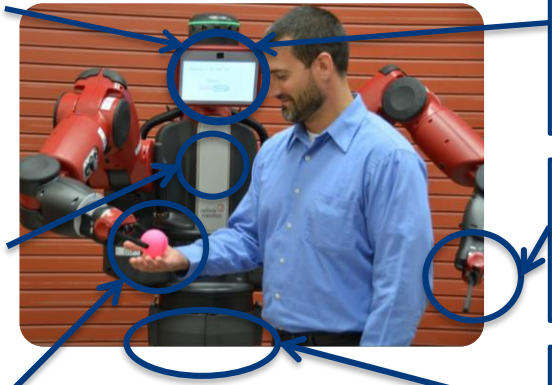


Image recognition:
- surveillance
- people recognition
- object and pose recognition
- OCR

Process automation:
- dexterous manipulation
- Robotic Process Automation

Autonomous navigation:
- logistics robot
- intervention robot
- inspection robot

LNE'S ACTIVITIES IN AI EVALUATION

Activity n°1: development of **evaluation standards**

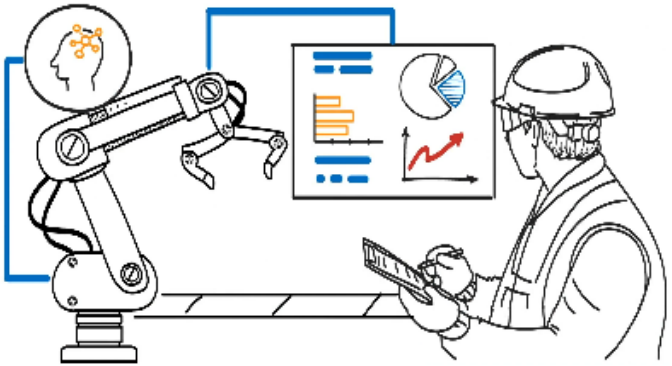
Activity n°2: AI systems **testing**

Activity n°3: **certification** of AI development and evaluation processes

Activity n°4: development of **evaluation tools**

Activity n°5: **professional training** on AI evaluation

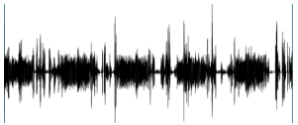
- 10+ years of experience
- 15+ ongoing R&D projects
- 950+ systems evaluated
- 15+ experts on AI Evaluation



EXPERTISE IN EVALUATION OF AI SYSTEMS

SPEECH AND TEXT

Transcription, Keyword spotting, Speaker comparison, Named entities recognition, Translation, etc.

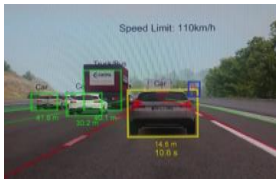


ترحيب ، يسعدنا أن نرحب بكم

Bienvenue, nous sommes ravis de vous accueillir

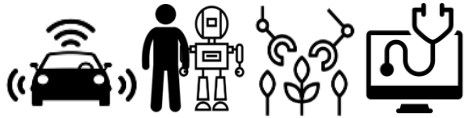
IMAGE

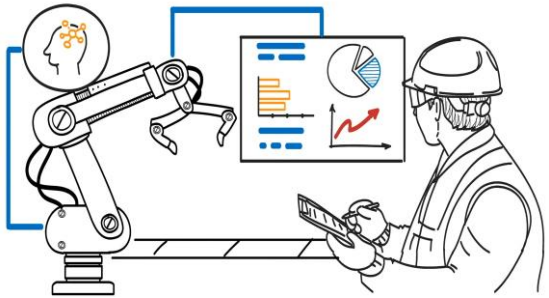
Object recognition, head tracking, OCR, etc.



MULTIMODAL

ROBOTICS





More than 10 years of experience in AI evaluation and more than 950 systems evaluated

ON-GOING R&D PROJECTS

- **Service robotics** (Robocom, 23 partners)
- **Agricultural robots** (Rose, 1 partner)
- **Biodiversity** (IA-Biodiv, 2 partners)
- **Autonomous vehicule** (3SA, 11 partners)
- **Evolutive systems** (Allies, 4 partenaires)
- **Future cockpit** (MMT, 1 partner)
- **Dialogue systems** (Lihlith, 5 partenaires)
- **Analog AI** (AIR, 3 partners)
- **Speaker recognition** (Voxcrim, 5 partners)
- **Robotics competitions** (Metrics, 16 partenaires)
- **AI certification de l'IA** (Grand défi, 15 partners)
- **Smart medical devices** (Labinnov, 3 partners)

AND OTHERS…

M4DT DAY

AI REQUIREMENTS

AI EVALUATION THROUGHOUT THE LIFE CYCLE

Evaluate AIs from their development to their maintenance.

1 – Evaluation in the development phase (project owner: developer)

- To guide R&D efforts
- To position oneself in relation to the competition
- Strengthen marketing arguments with quantitative and reliable measurements

2 – Conformity assessment (project owner: developer)

- CE marking (Machinery Directive, Medical Devices Directive, etc.)
- Voluntary certification

3 - Benchmarking (project owner: end user)

- Make an informed choice among the different technologies available on the market

4 – Monitoring and maintenance (project owner: end user)

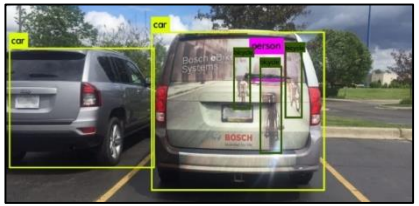
- Acceptance test
- Measure possible drifts of systems evolving in changing environment
- Measure possible performance regressions of lifelong learning systems

SPECIFICITIES OF AI TO BE TAKEN INTO ACCOUNT FOR ITS EVALUATION

AI evaluation is different from traditional software verification and validation.

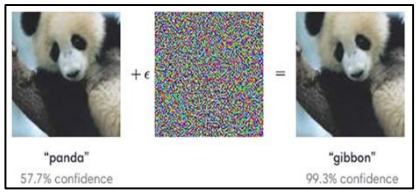
A - Essentially functional evaluation

- It is the adaptability of AI that makes it "intelligent".
- AI is therefore intended to operate in an open, often unstructured environment.
- An intelligent functionality is essentially evaluated by estimating the size of its operating domain.



B - Non-convex behavior

- Performance cannot be determined by interpolation and extrapolation between different operating points.
- Need to implement test tools that optimize the coverage of the operating environment (identification of corner cases, etc.).



C - Black Box

- Code auditing and formal verification are only rarely available.
- AI is often expected to be "explainable", i.e. to be able to justify its decisions with the right level of detail.

D - Evolving (sometimes)

- Need to put in place certification compatible with lifelong learning capabilities

AI EVALUATION CRITERIA

LNE incorporates all the criteria relevant to its partners into its evaluation standards:

Performance, robustness, and resilience evaluation
- characterization of the operating environment
- qualification of test data and environments
- data augmentation and automatic test scenario generation
- ongoing R&D projects: 10+

Evaluation of human-machine interaction
- within the framework of a close cooperation between an intelligent assistant and the pilot for example
- ongoing R&D projects: 2

Explainability evaluation
- quality of the explanations justifying the decision taken
- use of explanations to improve the tests carried out
- ongoing R&D projects: 4

Risk anakysis
- structured process for risk analysis of AI-enabled robotic systems to compensate for the lack of a standard
- ongoing R&D projects: 3

Ethics evaluation
- bias analysis, regulatory compliance, regulation of online platforms, etc.
- ongoing R&D projects: 2

Energy consumption
- analog AI vs. digital AI, use of quantum technologies, etc.
- ongoing R&D projects: 2

LNE'S INVOLVEMENT IN AI STANDARDIZATION

Afnor AI:

- pilot of the ad hoc group on the use of simulation to develop and evaluate AI
- member of the expert group on AI evaluation
- several contributions on AI evaluation and data characterization

CEN-CENELEC (JTC21):

- several contributions to the European roadmap and to the Focus Group's response to the European Commission's White Paper
- writing of a scientific paper in collaboration with the other members of the Focus Group

ISO/IEC JTC 1/SC 42:

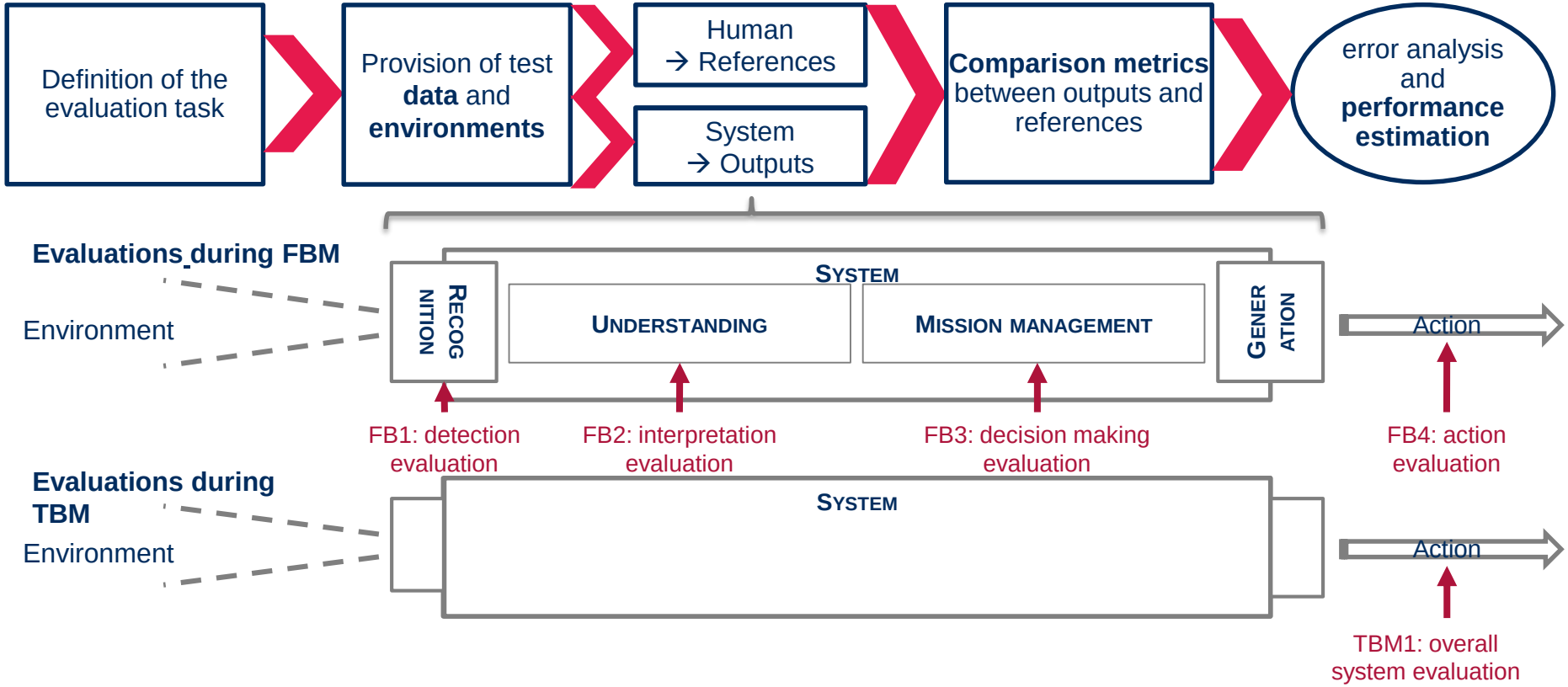
- member of WG1, WG3 and WG5
- several contributions, notably to the ISO/IEC 24029 TR "Assessment of the robustness of neural networks

UNM81 Robotics:

- Member of the AGV group
- Several contributions, in particular to the CEN WS 110 carried by the DIN on the performance of exoskeletons: comparative analysis of test benches and proposal of test environments consistent with those of the NIST

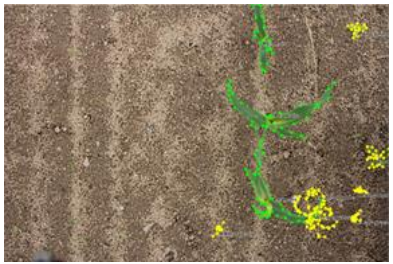
EXISTING TOOLS AND THOSE TO BE SET UP

AI EVELUATION METHOD



AI EVELUATION METHOD

Acquisition of images by the 4 evaluated robots



References: human annotations

Hypotheses : outputs from detection systems

Comparaison

1. Mapping
2. Calculation of the error rate

USEFUL TOOLS FOR EVALUATION

1. Data	2. labeling	3. Protocols, metrics	4. Testing environments

- Data augmentation
- Generation of adversarial testing scenarii
- etc.

DIANNE

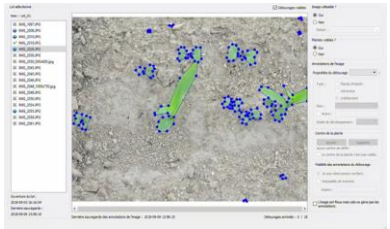
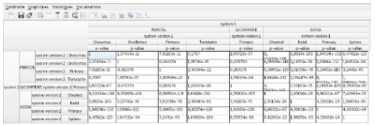


Image labeling
Automatic Pre-labeling

MATICS :
Datomatic – Dataset prep. and visualization
Evalomatic – Evaluation and visualization



ARTIFICIAL INTELLIGENCE EVALUATION LABORATORY (LEIA)

Trade-off between completeness and realism




Completeness

Realism

LEIA 1 LEIA 2 LEIA 3

DATA

central test databank

partnerships:

9 industrial groups

4 clusters

3 SMEs

23 academic laboratories

WHAT IS A CHALLENGE?

Evaluation campaigns to benchmark the performance of competing technologies, whose metrological rigor, collective emulation and knock-on effect generate progress in the field.

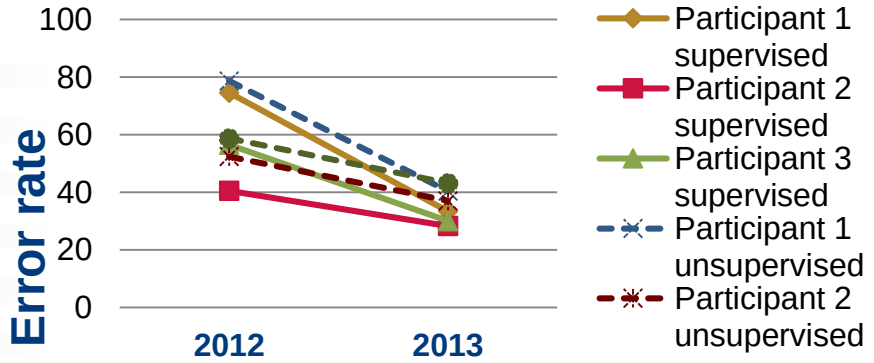
Repeated campaigns over time to assess progress:

Preparation of the evaluation tools → Dry Run → 1st Campaign → 2nd Campaign

R&D funding tool to bridge the TRL death valley (with monitored R&D efforts and targeted participant profiles):

CLOSED CHALLENGES
The sponsor select the participants

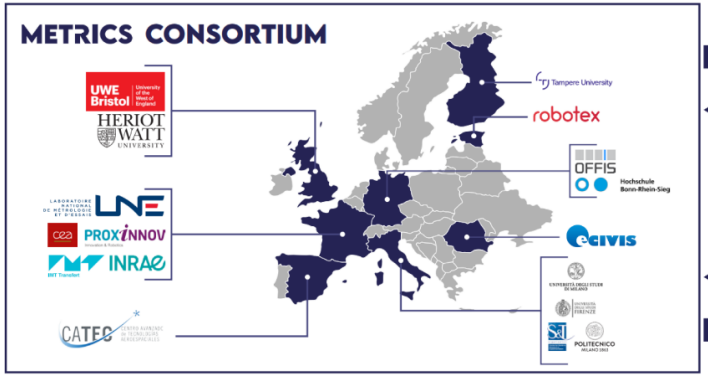
OPEN CHALLENGES
Everybody can participate



EXEMPLE DU PROJET H2020 « METRICS »

17 european partners

4 competitions



HEART-MET
Healthcare

RAMI
Inspection and maintenance

ACRE
Agri-food

ADAPT
Agile production

POSSIBLE APPROACHES TO AI CERTIFICATION

Process certification:

The AI functionality has been properly constituted (evaluation of the learning, evaluation and maintenance phases)

- Create confidence in the AI developed based on process control
- Analogous approach to creating trust via processes (management system certifications, CE marking of medical devices, aerospace etc.)

Product certification :

The AI functionality has a compliant behavior (test of the functionality)

- Impossible to address the needs of all sectors in which AI is used
- Very expensive
- Not very flexible

People certification:

Those involved in the development or use of AI throughout its life cycle are competent.

CERTIFICATION OF FOUR KEY PROCESSES

Design process

- Transform an expression of need into functional specifications.

Development process

- Translate these specifications into an evaluation-ready version of the AI functionality.

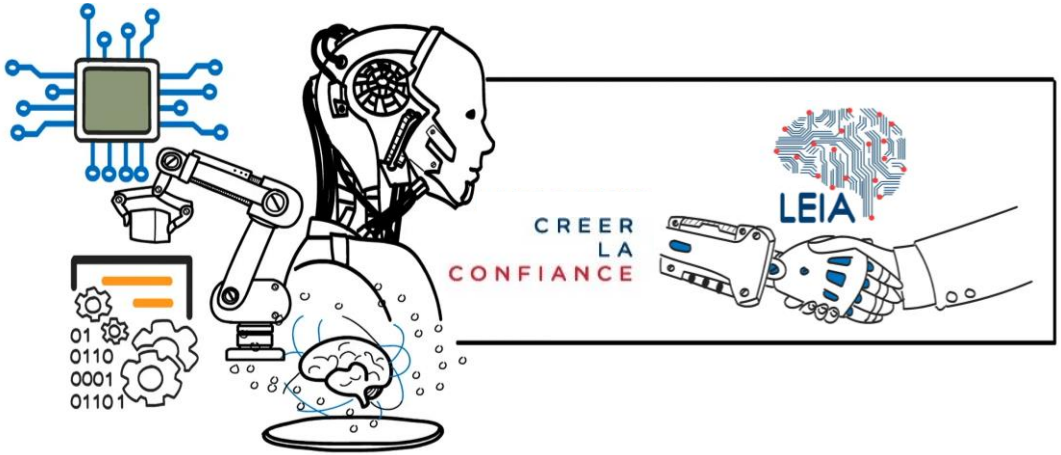
Evaluation process

- Verify the conformity of the system to the defined specifications before its deployment.

Maintenance process

- Ensure compliance of AI functionality with defined specifications after deployment and throughout its operational phase.

Thank you for your attention



Guillaume AVRIN, PhD
Head of the « AI Evaluation » department
Testing direction
Tel: +33 (0)1 30 69 13 62 - Mob: +33 (0)7 60 49 01 24
guillaume.avrin@lne.fr
Laboratoire national de métrologie et d'essais
29 avenue Roger Hennequin 78197 Trappes Cedex - lne.fr