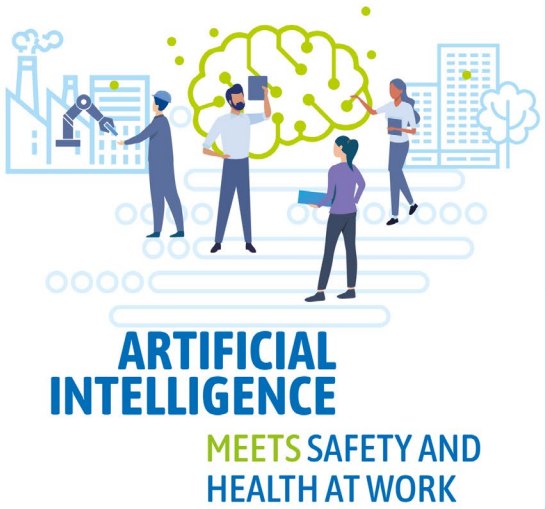
EUROSHNET



Evaluation and certification for safer artificial intelligence

Dr Agnes DELABORDE

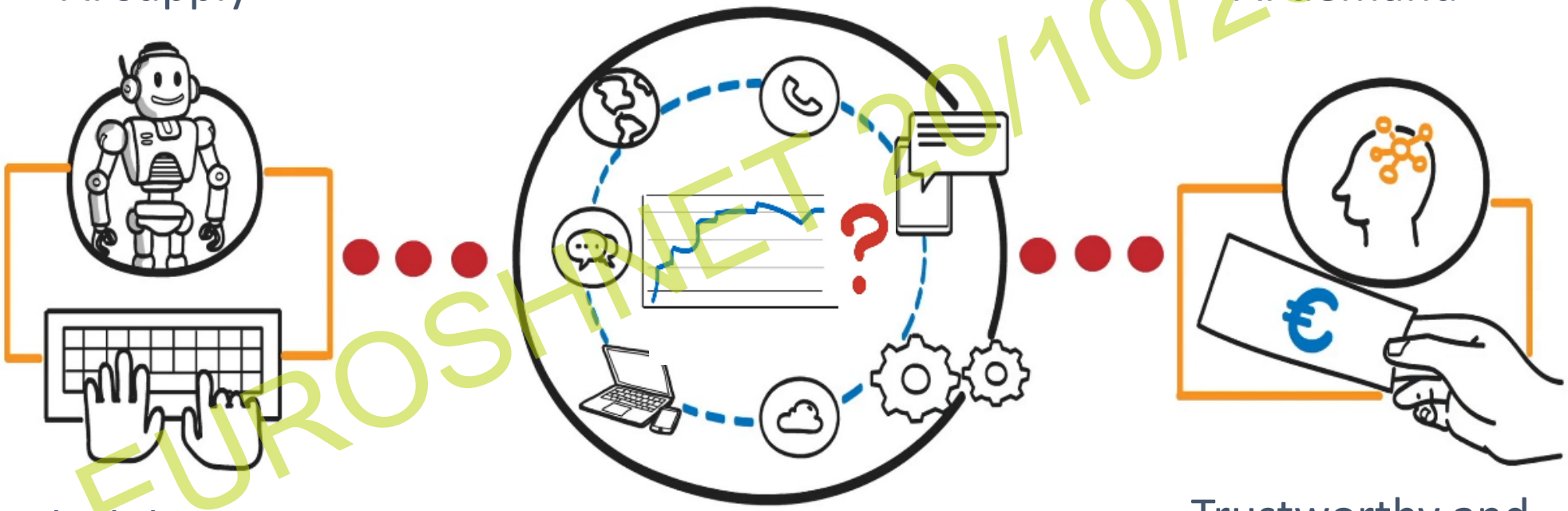
Research engineer in AI evaluation

LNE

Matching AI supply and demand

AI supply

AI demand



Black-box, non convex, evolutive systems

Need: AI evaluation & certification

Trustworthy and efficient functionalities

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

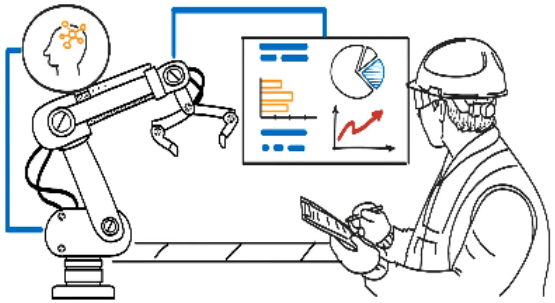
Activity n°4: development of **evaluation tools**

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

Application areas:

- **NLP:** speech-to-text, translation, speaker recognition, etc.
- **Image processing:** person recognition, object segmentation, OCR, etc.
- **Robotics:** Smart MD, industrial robots, inspection robots, autonomous cars, agricultural robots, etc.

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



How and why performing evaluation?

One-off evaluation

- **Description:** Evaluation of the performance of a system at a specific time in a specific test environment
- **Example:** To assess its compliance with regulations

One-off benchmarking evaluation

- **Description:** Comparative analysis of the performance of different systems on the same evaluation task in the same test environment at a specific time
- **Example:** To allow the user to make an informed choice between different existing technologies

Repeated evaluation campaign (« challenge »)

- **Description:** Comparative and repeated analysis of the performance of different systems on the same evaluation task
- **Example:** To evaluate the progress made by these different technologies and to encourage "coopetition"

Evaluation: overview of approaches

Definition of the evaluation task → Provision of test **data** and **environments** → Human → References / System → Outputs → **Comparison metrics** between outputs and references → error analysis and **performance estimation**

Evaluation in representative environments

Evaluation on representative data

References: human annotations

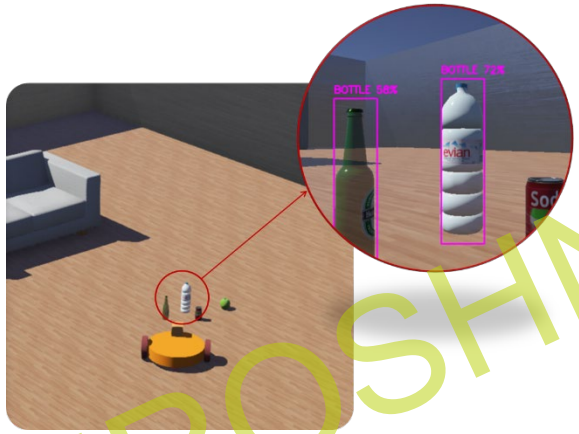
Outputs of the smart camera

Metric:

$$EGER = \frac{\sum_{k=1}^{N} C_k + FP_k + FN_k}{\sum_{k=1}^{N} R_k}$$

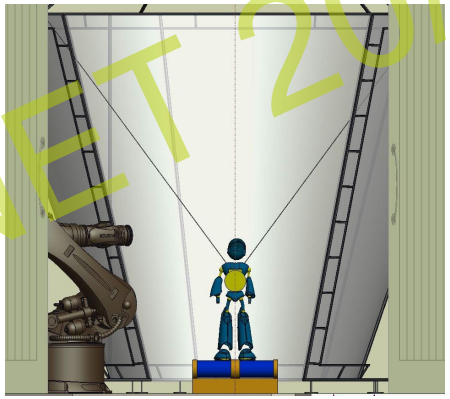
Test beds configuration (example: LNE's LEIA evaluation infrastructure)

LEIA1
Software In the Loop

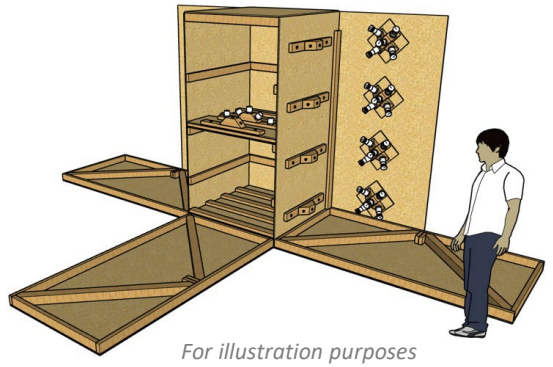


For illustration purposes

LEIA2
Robot In the Loop
Camera In the Loop



LEIA3
Testing in realistic environment



For illustration purposes
Source: ASTM E54.09

Exhaustivity

Realism

Does evaluation make AI safer?

- Some elements are required (and not fully available yet):
  - Identify forbidden and/or compulsory outputs
  - Trade-off between exhaustivity/realism (cost, existence of infrastructure)
  - Acceptable thresholds: minimum performance rates
- Contributes to safety:
  - Risk assessment drives the selection of test scenarios
  - Test results highlight areas of underperformance
  - Estimate the impact of mitigation strategies

Certification: overview of approaches

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)

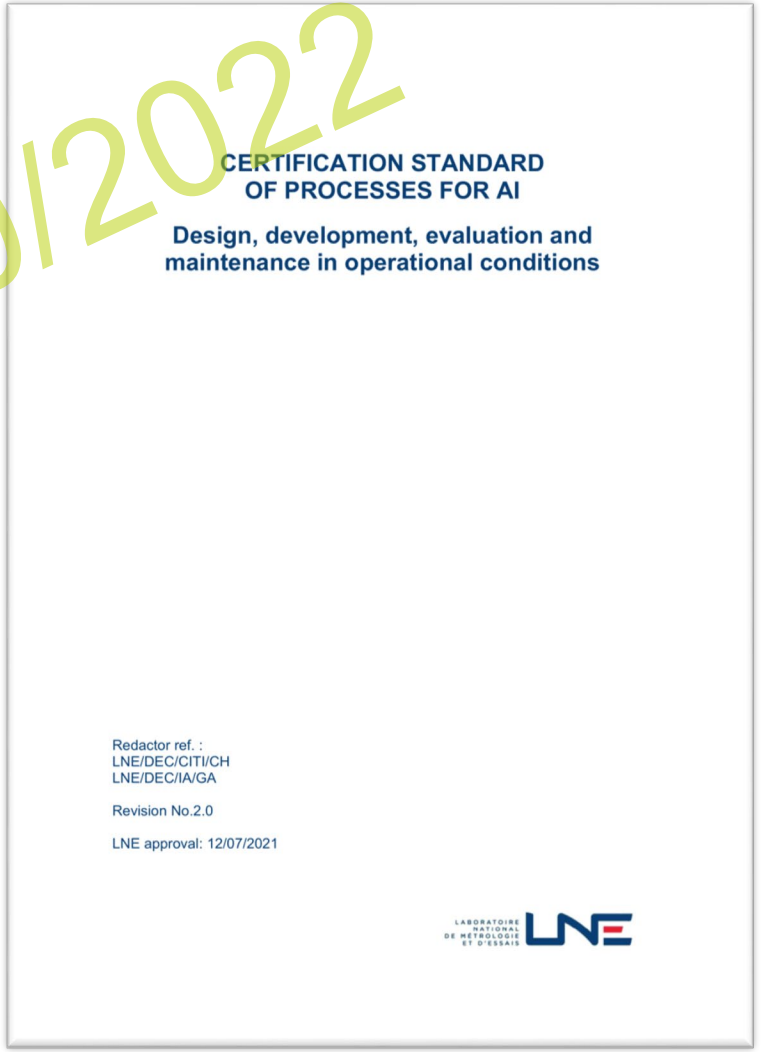
- ➔ Potential limitations to overcome (sectorial specificities, testing cost, test methods)

People certification:

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

Certification of processes for artificial intelligence

- https://www.lne.fr/en/service/certification/certification-processes-ai



Overview of the certification

- Not meant to certify the AI product itself, but guarantee that it has been **designed correctly**.
- Contributes to ensuring a trustworthy product, through **control of the processes and use of good practice**.
- Voluntary certification.
- For Machine Learning (and hybrid ML/expert).
- Processes analyzed:
  - Design, development, evaluation and maintenance in operational conditions

Contribution of evaluation and certification to safety

Evaluation

- Allows verification
- Provides valuable insight into the system's risks

REQUIRES

Exhaustive coverage of factors influencing safety
Methods (testing, data qualification, etc.)
Infrastructure (accessible, affordable, standardized)

Certification

- Allows validation
- Provides checkpoints that guarantee compliance

REQUIRES

Exhaustive coverage of factors influencing safety
Acceptable "thresholds"
Frame(s) of reference (derived from regulation)

Thank you for your attention

Dr. Agnes Delaborde
Research engineer in AI and robotics evaluation, LNE
agnes.delaborde@lne.fr

EUROSHNET

ARTIFICIAL INTELLIGENCE

MEETS SAFETY AND HEALTH AT WORK

QUESTIONS FROM PUBLIC

EUROSHNET

ARTIFICIAL INTELLIGENCE

MEETS SAFETY AND HEALTH AT WORK

LUNCH