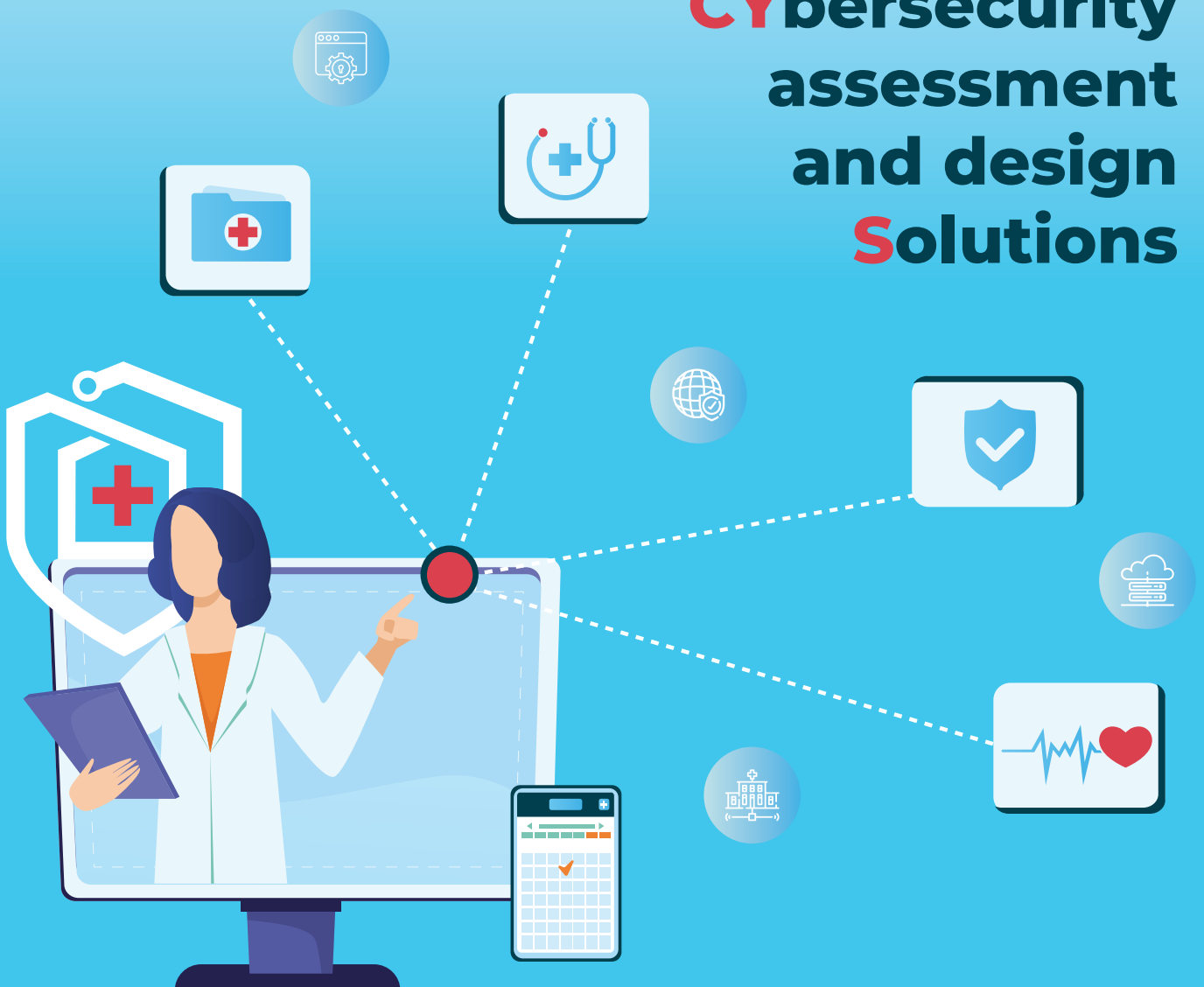




NEMECYS

NEW MEDICAL CYBERSECURITY AND DESIGN SOLUTIONS

NEW **ME**dic **CY**bersecurity assessment and design **S**olutions



NEMECYS Toolbox:
AI Privacy Risk Assessment (AIRA)

Developed by IBM

Recent studies show that a malicious third party with access to a trained Machine Learning (ML) model, even without access to the training data itself, can still reveal sensitive, personal information about the people whose data was used to train the model. For example, it may be possible to reveal whether or not a person's data was part of the model's training set, or even infer sensitive attributes about them, such as their salary. It is therefore crucial to be able to assess the privacy risk of Artificial Intelligence (AI) models that may contain personal information before they are deployed, allowing time for applying appropriate mitigation strategies. Such assessments are also important to enable comparing and choosing between different ML models.



The **AI Privacy Risk Assessment (AIRA) tool** is an end-to-end framework for privacy risk assessment of AI models. It aims to alleviate the issues of existing tools and frameworks that require a high degree of expertise or are tightly coupled with specific ML frameworks.

The first goal of this end-to-end risk assessment tool is to automate many of the decisions around which attacks and metrics to run, and all of the technical preparation required in order to run them. Secondly, since most non-technical users cannot understand the meaning of each individual attack or score, the tool summarizes these individual results into an overall privacy risk score.

The tool enables assessing models from different ML frameworks, using a variety of low-level privacy attacks and metrics, without requiring deep technical expertise.



Technical Inquiries for the Tool

If you are interested in learning more about the tool, its capabilities, or technical details, please contact **Abigail Goldsteen**, abigailt@il.ibm.com



**Scan to
access
the tool**

www.nemecys.eu

 [@NEMECYS_eu](https://twitter.com/NEMECYS_eu)

 [nemecys-horizon-eu-project](https://www.linkedin.com/company/nemecys-horizon-eu-project)



Co-funded by
the European Union

The NEMECYS project is co-funded by the European Union under grant agreement ID 101094323, by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee grant numbers 10065802, 10050933 and 10061304, and by the Swiss State Secretariat for Education, Research and Innovation (SERI). The information and views set out in this publication are those of the author(s) only and do not necessarily reflect those of the European Union, HADEA, UKRI or SERI. Neither the European Union nor the granting authorities can be held responsible for them.