**1/ Abstraction:**

* The literature on assurance,driven by contradicting motivations, assumptions, and intuitions.
* A new AI assurance defnition is adopted and presented, and assurance methods are contrasted and tabulated.
* Assurance: this term has been rarely applied to conventional software engineering; rather, it is used in the context of AI and learning algorithms. In this manuscript, based on prior definitions and recent AI challenges, we propose the following definition for AI assurance:
* **A process that is applied at all stages of the AI engineering lifecycle ensuring that any intelligent system is producing outcomes that are valid, verified, data-driven, trustworthy and explainable to a layman, ethical in the context of its deployment, unbiased in its learning, and fair to its users**.
* Our definition is by design generic and therefore applicable to all AI domains and subareas. Additionally, based on our review of a wide variety of existing definitions of assurance, it is evident that the two main AI components of interest are *the data* and *the algorithm*; accordingly, those are the two main pillars of our definition.

**2/ Relevant terminology and definitions:**

* We acknowledge the following relevant terms: (1) validation, (2) verification, (3) testing, and (4) assurance.
* Verification: “The process of evaluating a system or component to determine whether the products of a given development phase satisfy the conditions imposed at the **start of that phase”.**
* Validation: “The process of evaluating a system or component during or at the end of the development process to determine whether it satisfies specified requirements”
* testing is “the process consisting of all lifecycle activities, both static and dynamic, concerned with planning, preparation and evaluation of software products and related work products to determine that they satisfy specified requirements, to demonstrate that they are fit for purpose and to detect defects' '.
* Assurance: A process that is applied at all stages of the AI engineering lifecycle ensuring that any intelligent system is producing outcomes that are **valid, verified, data-driven, trustworthy and explainable to a layman**, **ethical** in the context of its deployment, unbiased in its learning, and fair to its users.

**3/ Description of included articles:**

**a/ AIA landscape:**

**b/ State of AIA:**

* For instance, in expert systems, the inference engine component creates rules and new logic based on **forward and backward propagation** [20]. Such processes require extensive assurance of the process as well as the outcome rules. Alternatively, for other AI areas such as neural networks, while propagation is used, **taxonomic evaluations and adversarial targeting** are more critical to their assurance
* For instance, in Lee et al. [121], layer-wise relevance propagation was introduced to obtain the efects of every neural layer and each neuron on the outcome of the algorithm.

**c/ The review and scoring of assurance methods:**

Each elemental metric is allocated one point, and each method is either given that point or not (0 or 1), as follows:

* **I. Specificity to AI:** some assurance methods are generically tailored to many systems; others are deployable only to intelligent systems; one point was assigned to methods that focused (i.e. specifc) on the inner workings of AI systems.
* **II. The existence of a formal method:** this metric indicates whether the manuscript under review presented a formal (quantitative and qualitative) description of their method (1 point) or not (0 points).
* **III. Declared successful results:** in experimental work of a method under review, some authors declared success and presented success rates, if that is present, we gave that method a point.
* **IV. Datasets provided:** whether the method has a big dataset associated with it for testing (1) or not (0). This is an important factor for reproducibility and research evaluation purposes.
* **V. AI system size:** methods were applied to a small AI system, other were applied to bigger systems for instance, we gave a point to methods that could be applied to big real-world systems rather than ones with theoretical deployments.
* **VI. Declared success:** whether the authors declared success of their method in reaching an assured AI system (1) or not (0).
* **VII. Mentioned limitations:** whether there are obvious method limitations (0) or not (1).
* **VIII. Generalized to other AI deployments:** some methods are broad and are able to be generalized for multiple AI systems (1), others are “narrow” (0) and more specifc to one application or one system.
* **IX. A real-world application:** if the method presented is applied to a real-world application, it is granted one point.
* **X. Contrasted with other methods:** if the method reviewed is compared, contrasted, or measured against other methods, or if it proves its superiority over other methods, then it is granted a point.
* Rising questions How is learning performed inside the blackbox? How is the algorithm creating its outcomes? Which dependent variables are the most infuential? Is the AI algorithm dependable, safe, secure, and ethical?

**4/ Conclusion:**

* The most successful methods presented in literature (scored as 8, 9, or 10), are the ones that were specifc to an AI subarea and goal; additionally, ones that had done extensive theoretical and hands-on experimentation.
* Accordingly, we propose the following five considerations as they were evident in existing successful works when defining or applying new AI assurance methods:
* (1) Data quality: similar to assuring the outcomes, assuring the dataset and its quality mitigates issues that would eventually prevail in the AI algorithm.
* (2) Specificity: as this review concluded, the assurance methods ought to be designed to one goal and subarea of AI.
* (3) Addressing invisible issues: AI engineers should carry out assurance in a procedural manner, not as an afterthought or a process that is performed only in cases of the presence of visible issues.
* (4) Automated assurance: using manual methods for assurance would in many cases defeat the purpose. It is difficult to evaluate the validity of the assurance method itself, hence, automating the assurance process can—if done with best practices in mind– minimize error rates due to human interference.
* (5) The user: involving the user in an incremental manner is critical in expert-relevant (non-engineering) domains such as healthcare, education, economics, and other areas.