TRUST IN AI 8

1. Trust-relevant situation

* Trust-relevant (or trust diagnostic) situations **involve vulnerability and require some stake (risk)**
* Without being vulnerable or at risk, there is no need for trust
* AI-produced credit scoring may represent a risk to the loan officer, or AI-supported diagnosis may inherent a risk for the patient to be misdiagnosed

1. Trustor‘s propensity to trust

* **refers to the general tendency for someone to trust others**
* the impact of trust propensity is most notable early in interpersonal interactions, when other information may not yet be available

1. Scales to measure a person‘s propensity to trust

* **General Trust Scale** -> a 6-item questionnaire that uses general statements to measure participants’ beliefs about honesty and trustworthiness of others, in general *(Yamagishi & Yamagishi, 1994)*
* **Trust in Technology Scale** -> a 6-item questionnaire that measures participants’ beliefs about general trustworthiness of technology *(Schneider, 2017)*

1. Trustee‘s trustworthiness

* Trustworthiness is the trustor’s perception of the trustee
* Perceptions are formed as a trustor interprets and ascribes motives to the trustees’ actions . Thus, perceptions of trustworthiness, although inherently within the trustor, are a function of the interaction of trustor and trustee as the trustor is processing information about the trustee. It is important to note these are the ascribed beliefs of the trustor and are not necessarily factual

1. Goals of research on interpretability and trust:

* Trust in a system must be appropriately calibrated to the actual system performance (Muir, 1994)

1. Theoretical foundations: Dimensions of Trust

* Affective/social trust: When a trustee is perceived as well-minded, warm-hearted and adhering to social norms
* Cognitive/performance-based trust: When a trustee is perceived as competent, understandable and predictable in terms of the performance required
* **Cognitive components of trust should be more relevant for robotics/AI than affective components**