Abstracts for Governance of Emerging Technologies and Science Conference May 2022

Submitted by: Helen A. F. Gould, Principal and Founder, Tech Strategizers, LLC (Intel, Retired)

Abstract #1:

Fostering Trust in AI: Focusing on Competency, Direct & Inferred Relational Trust, and Protective & Demonstrative Trust Mechanisms

Trust and Trustworthiness of AI technologies are universally recognized to be of critical importance but are still in the early stages of being defined and debated in multiple forums, from IEEE and ISO to the World Economic Forum and the Athens Roundtable on Artificial Intelligence and the Rule of Law.

This session explores what it means to have "trust" in AI and Autonomous Intelligent Systems (AI/AIS) and technologies and the necessity for these systems to be "worthy of our trust" and how trust of new technologies can, and needs to, be "earned."

The traditional definitions of technology trustworthiness are augmented to include concepts of Competency and Relational Trust. **Competency** poses the question of: "is the AI/AIS capable and able to effectively perform the job or task assigned?" with the necessary knowledge, abilities, skills, experience, and appropriate behaviors. **Relational Trust** refers to both "**Direct Trust**," as in trust between humans and machines (or between machines), and to "**Inferred Trust**," as in trust that is inferred with the organization that developed the technology, or the organization that is using the technology or the governmental organization that is regulating the technology

Additionally, ways to foster greater human trust in technology are proposed by utilizing "**Protective Trust**" and "**Demonstrative Trust**" mechanisms. **Protective Trust** can be regulated and enforced with soft law frameworks, laws, and regulations. **Demonstrative Trust** is earned when the product or service of technology has been "demonstrated" to be worthy of trust. Both protective trust and demonstrative trust mechanisms can be effective means to enhance trust of emerging technologies with end users, prospective customers, and regulators.

While much work remains to be done to foster greater and appropriate levels of trust (and even appropriate levels of distrust) in AI technologies, this session attempts to advance these important discussions by contributing additional pieces of the overall puzzle humanity needs to solve.