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

"Foreseeing the Unforeseen:" Considering Intended and Unintended Behaviors with Rule Abiding and Rule Breaking Results

There are understandable trepidations about how Artificial Intelligence (AI) technologies will be intentionally or unintentionally used and/or misused and/or abused. A structured risk assessment process is needed for technologists developing and using these technologies to consider potential consequences and to implement appropriate mitigations, precautions, and safeguards.

Inspired by the development of regulations for Unmanned Aircraft Systems (UAS) and ISO 21448, the Safety of the Intended Functionality (SOTIF) for automated vehicles, the concepts are applied more broadly to AI applications.

This session explores a straightforward approach to "foresee the unforeseen" with a four-quadrant scenario analysis framework for considering intentional and unintentional behaviors with potential rule abiding or rule breaking results. The four quadrants illustrate being: Careful, Clueless, Careless or Criminal. The objective function of the Four C's Scenario Framework is to maximize the Careful, to consider the Clueless and to minimize the Careless and the Criminal.

With AI, it is necessary but not sufficient to consider how a technology, product or service is intended to be used. It is equally, if not more important, to anticipate how AI could unintentionally be used with desirable or undesirable results and potentially be used nefariously or for illicit purposes.