**Measures Proposed**

**To Manage Risk of AI Development**

* **Types of Suggested operational limits and constraints**
  + Limit the problems posed to AIs to ones that do not demand moral choices.
  + Limit AI access to open, broad-based information resources – control inputs.
  + Limit AI access to the means of physical control of its external environment.
  + Limit AI capacity for self-programming.
  + Program AI to adhere to approved human moral values;
    - Provide exemplars that pose fundamental ethical/moral choice situations and appropriate responses -- while avoiding AI access to actual human responses.
    - Provide instruction/programming on value-rich decision making.
  + Require detailed reporting by AIs to human monitors of value-rich decision making.
  + Align AI objectives with human objectives so the self-interest of each is consistent with the other and so each works in partnership with the other.
* **Government regulation of development and deployment of new AI technologies:**
  + Require that all materials/products/communications produced by AIs be clearly identified as such for human consumers/clients/users.
  + License AI developers and require of standard safeguards and protocols throughout the AI development industry;
  + Monitor/audit all licensees for compliance;
  + Monitor/audit all AI technology in the field as it is placed in service;
  + Require government approval before implementation of any new AI technology;
  + Control the publication of works that inform high level AI development.
  + Adoption all requirements internationally by treaty.
* **Risk detection and monitoring-- develop new means to track AI development able to detect, track, and monitor:**
  + … when AIs are achieving self-consciousness or making use of Artificial General Intelligence (AGI),
  + … when AIs are using their intelligence to make themselves smarter without aid/management from their human monitors,
  + … when new, potentially dangerous technologies are under development or emerging,
  + … when dangerous trends, technologies, events for which AIs might be responsible appear in society or in the marketplace.
* **Voluntary industry self-management**
  + Adopt standards of care and self-governance to be adopted (and made publicly known) by responsible AI developers.
  + Count on human clients/customers using AI services to avoid AIs that seem to have developed problematic tendencies.
  + Industry-wide sharing of potentially dangerous AI research/implementation activities.