I'm writing to comment on the AI Risk Management Framework: Initial Draft (March 17, 2022).

Positives:

- Describes a broad set of risks and explains how those categories are related to related work on AI risk.
- Emphasizes the need to incorporate AI risks into enterprise risk management.
- Uses the likelihood and consequences dimensions of risk.

Negatives:

- The definition of "artificial intelligence" is limited to systems "that learn from data in an automated or semi-automated manner." Although learning is an important part of AI, many AI systems do not use learning.
- Section 5 discusses many important characteristics but does not clearly identify risks (Figure 2 presents only some examples). Risk identification is an important part of risk management, and system designers and users need a list of risks that can help inform their risk assessment.
- Section 6 presents a "AI RMF Core" that is very confusing. The definition in the first sentence (Page 14, lines 3-4) is unclear. This section does not reference existing standards and processes on risk management (cf. ISO 31000), which would be very helpful (and reinforce the connection to enterprise risk management).

Given that many fine resources on risk management already exist, the value-added content here should be the list of risks associated with AI systems, which can help system designers and users identify and assess risks within a risk management process. They need to know: What can go wrong? Whom (or what) will it harm? I recommend rewriting this to focus on the answers to those questions.

--

Jeffrey W. Herrmann, Ph.D.
Professor, Department of Mechanical Engineering and Institute for Systems Research
Web site: https://user.eng.umd.edu/~jwh2/jwh2.html
Director, Reliability Engineering Graduate Program
Director, Systems Engineering Graduate Program
University of Maryland, College Park