

# PUBLIC SUBMISSION

**Received:** May 17, 2025  
**Tracking No.** mas-q587-ftjz  
**Comments Due:** May 28,  
2025 **Submission Type:** API

**Docket:** NSF-2025-OGC-0001  
NITRD\_FRDOC\_0001

**Comment On:** NSF-2025-OGC-0001-0001  
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

**Document:** NSF-2025-OGC-0001-DRAFT-0081  
Comment on FR Doc # 2025-07332

---

## Submitter Information

**Name:** David Schweitzer

---

## General Comment

Efforts need to be made to adequately fund strong data sources to train these machine learning models. Typically, projects are not allowed to account for the often very expensive data procuring that is required, and data procuring must instead be sneaked into proposals at modest salaries or else there is a question of why a machine learning coding project costs hundreds of thousands of dollars. Unfortunately, data procuring is typically expensive and cannot be done for just a few thousand dollars as personnel need to be involved to plan the experimental design and/or scenarios from which the data will be derived, ensure there are adequate tools in place to capture said data, ensure there is compliance with partnering organizations if data is to be captured from environments managed by others, and ensure there is adequate support in managing this data so it is either intentionally made available to other organizations or is kept securely and only provided on a need-to-know basis, among other concerns. Often, data procuring is not only expensive, but also the most time consuming, and this must be accounted for in the funding.

There must also be more effort in funding multimodal basic research. For example, many data types are very different, such as image data vs. cyber data, but even many AI experts are not keen on utilizing models to adequately parse these. Instead, data types may be converted from one to another (such as taking pictures of documents) rather than exploring multimodal capabilities, and this is typically because multimodal understanding is both poor and not well-developed at the research level. Scientists and engineers will try, but they are typically limited in small, 1-year efforts that do not provide adequate time to research fundamental problems in multimodal algorithms.