

PUBLIC SUBMISSION

Received: May 05, 2025
Tracking No. mab-ymrm-1ega
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-0050
Comment on FR Doc # 2025-07332

Submitter Information

Name: James Worcester

General Comment

While the goals of the plan seem reasonable, it is vital to understand that none of these will be achievable unless the AI being used is first made in such a way that it can be reliably aligned. Second, we currently have no reliable methods for aligning an AGI. Even the limited capability LLMs we have now routinely do things the companies deploying them don't want (see OpenAI's recent problem with unwanted sycophancy).

Until we have found a way to reliably align an AGI, creating one is an irreversible step we must not take. By its nature, this is a situation where we will only ever get one shot on goal. If we miss, we will then be in an unrecoverable situation with an unaligned superintelligence that will not care what we thought we wanted. Should humanity be fortunate enough to survive such an outcome, we will never again be in control of the future we have ceded to a more intelligent entity. Please, decide responsibly how best to proceed, and remember that the fate of humanity depends on getting this right.