

We don't yet know which AI architectures are safe; learning more about this is one of the goals of FLI's grants program. AI researchers are generally very responsible people who want their work to better humanity. If there are certain AI designs that turn out to be unsafe, then AI researchers will want to know this so they can develop alternative AI systems.