Exploring the Civil-Military Divide

Data Safeguarding Plan

I. Project Description

This study will examine the opinions of software engineers, other employees at leading technology companies, computer science students and/or faculty, and members of the general public about the development of Artificial Intelligence (AI) and its potential applications for use by the military and United States (U.S.) government. It is being funded by the Department of Defense's (DoD's) Joint Artificial Intelligence Center (JAIC) and the Office of Net Assessment. It aims to explore the range of viewpoints about acceptable uses for Artificial Intelligence in defense applications as well as the factors that have shaped those beliefs. We hope to help each side understand the other's point of view and help bridge the gaps between the national security community and the engineering experts who deeply understand these new technologies.

II. Responsibility for Data Safeguarding

- The principal investigators, James Ryseff and Eric Landree, have overall responsibility for data safeguarding.
- All project staff will read and agree to the terms of this Data Safeguarding Plan.

III. Types of Data and Data Safeguarding Procedures – Focus Groups and In-Depth Interviews

Focus groups and in-depth interviews will collect opinions from computer science students and/or faculty, software engineers, and tech company employees about their trust in the U.S. federal government, DoD, and the acceptable uses of AI.

1. **Recruitment Information.** Names and contact information will be collected for recruitment only when necessary. When possible, anonymous focus groups will be conducted at events such as conferences. Additionally, students and professionals will be identified through professional societies, outreach to school department and student organizations and/or through vendors. Any files with contact information will be stored according to security guidelines presented in RAND's Data Protection Matrices; that is, they will be stored on an internal RAND internal server that is accessible only through the RAND internal network. Access to the server itself is password-protected, and the data will be stored in password-protected locations only accessible by approved project staff. The machine and directories are further protected by user authentication and passwords.
2. **Audio Recordings** – Focus groups and interviews may be audio recorded, and will be stored only until transcription is complete. At RAND, audio files will be stored according to security guidelines presented in RAND's Data Protection Matrices; that is, they will be stored on an internal RAND internal server that is accessible only through the RAND internal network. The data will be stored in locations only accessible by approved project staff. The machine and directories are further protected by user authentication and passwords.

3. **Transcripts.** Transcripts will be created by using a trusted external vendor. Audio files will be shared with the vendor via Kiteworks. Transcriptions will be de-identified and will not include any names or individually identifiable information.

4. **Handwritten or Typed Notes** from focus groups or interviews. These notes will not include any individually identifiable information.

IV. Types of Data and Data Safeguarding Procedures – Surveys

1. **Sample Files.** Sample files for contacting respondents for web data collection will be obtained through a variety of sources. We will purchase from sources such as Monster.com, Dunn & Bradstreet, and other sample vendors. Additionally, files will be created by RAND through internet searches. Sample files will be stored according to security guidelines presented in RAND's Data Protection Matrices. As the files will contain personally identifiable information (PII), they will be stored on a RAND Fixed internal server that is accessible only through the RAND internal network. Access to the server itself is password-protected and the data will be stored in password-protected locations only accessible by approved project staff. The machine and directories are further protected by user authentication and passwords. Sample files will contain variables such as:

    - Name
    - Employer
    - Title
    - Email Address
    - Phone Number

2. **Survey Sample.** Programmers will select the survey sample from the sample files and provide SRG with the survey sample file including only the variables needed for data collection. The programmers will assign a unique study identifier (RAND ID) to each sampled case. The survey sample will be loaded into RAND's Survey Research Group (SRG) study record management system (RMS). This sample will contain the variables needed for data collection, including the study identifier (RAND ID), respondent name, and respondent contact information. These files are used by the SRG to conduct the survey administration. All RMS data will be stored on SRG's secure segment.

3. **Web-based Survey Responses** from sampled software engineers and tech company employees. The survey will be administered using the ConfirmIt survey platform. Responses will be downloaded and stored in a secure location in alignment with the RAND's Data Protection Matrices.

4. **Analytic File** – Project programmers will create the analytic data file. The file will include web survey data and data from the original sample file. The SRG will also provide process and outcomes data, such as final outcome code and other data collection process measures to be determined. Analytic files will be de-identified.

V. Data Transmittal

- A RAND SRG programmer will program the web survey. The SRG will utilize ConfirmIt to collect web-based survey data.

- Hardcopy focus group and interview notes collected at remote locations will be hand-carried back to RAND and coded or typed into electronic records.
- Transfer of electronic files with identifiable data will occur via Kiteworks or internal drop boxes. Files will be password protected.

VI. Disclosure Risks

- Although personal identifiers are not included in the survey data collected, it is possible that the identities of some individuals could be inferred based on their title and employment history (e.g., if there is only one person with certain work experience completing the survey). This risk is minimized by the way the data are handled at RAND (see Data Safeguarding Procedures below). If this were to occur, there is some risk for individuals in the file if certain data elements, such as attitudes toward the U.S. government, were disclosed.
- The client will receive a copy of the deidentified analytic file and deidentified focus group/interview notes and transcripts.

VII. Auditing and Monitoring Plans

- The project programmer will periodically check permissions on directories holding the files.

VIII. Destruction of Data

- At the end of the project, all files that contain respondent contact information or that link survey identifier with respondent identity will be destroyed in accordance with RAND's procedures for the destruction of "privacy waste."