

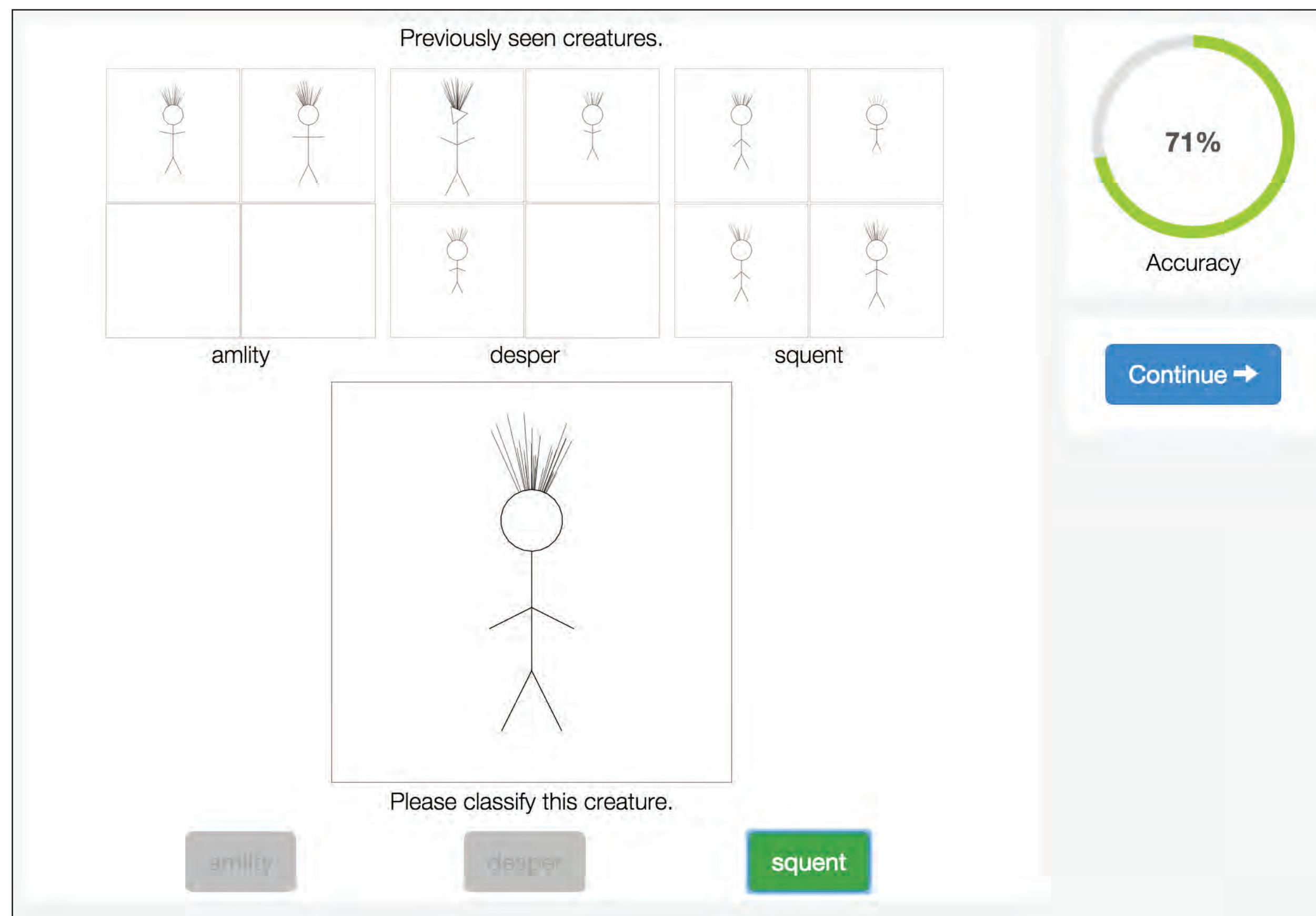
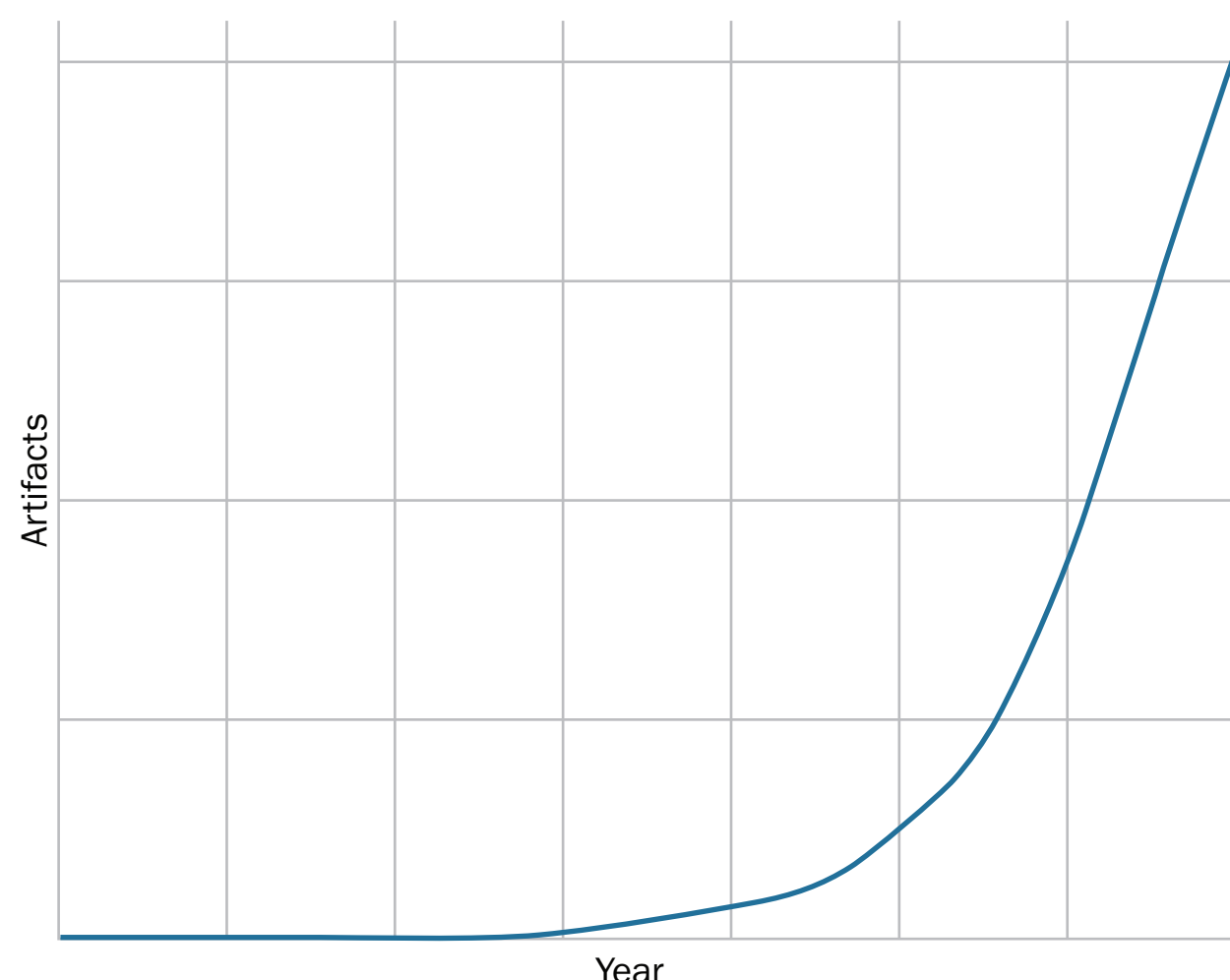
Human-Computer Decision Systems

Security decision systems aim to distinguish malicious activity from benign and often use a combination of human expert and automated analysis, including machine learning (ML). Systems using only human experts scale badly; pure ML systems are susceptible to structured attack by adversaries and, in most cases, have unsatisfactory performance on their own.

- Many operational security problems depend on a small number of skilled analysts to process a large and growing firehose of potentially malicious data.
- Traditional active learning tries to address this situation by suggesting allocation of limited analysis resources that optimize the convergence of a machine learning classifier.

Growth of CERT Artifact Catalog

Total Artifacts Over Time



A screenshot of the experimentation system built using Mechanical Turk and Psiturk.

- The human-computer collaboration model will improve upon traditional active learning by optimizing not simply for convergence of the ML component, but also for future performance of the overall system, including mutable human analysts.
- We test the performance of new models not only through simulation, but also through human-subject experiments.
- Because conducting these experiments using real security analysts performing their normal tasks would be prohibitively expensive, we instead developed a proxy problem of identifying fictional creatures and leveraged non-experts on Amazon's Mechanical Turk platform. The process of generating the fictional creatures adheres to the statistical distributions of real malware classes.

Dynamic Proactive Learning

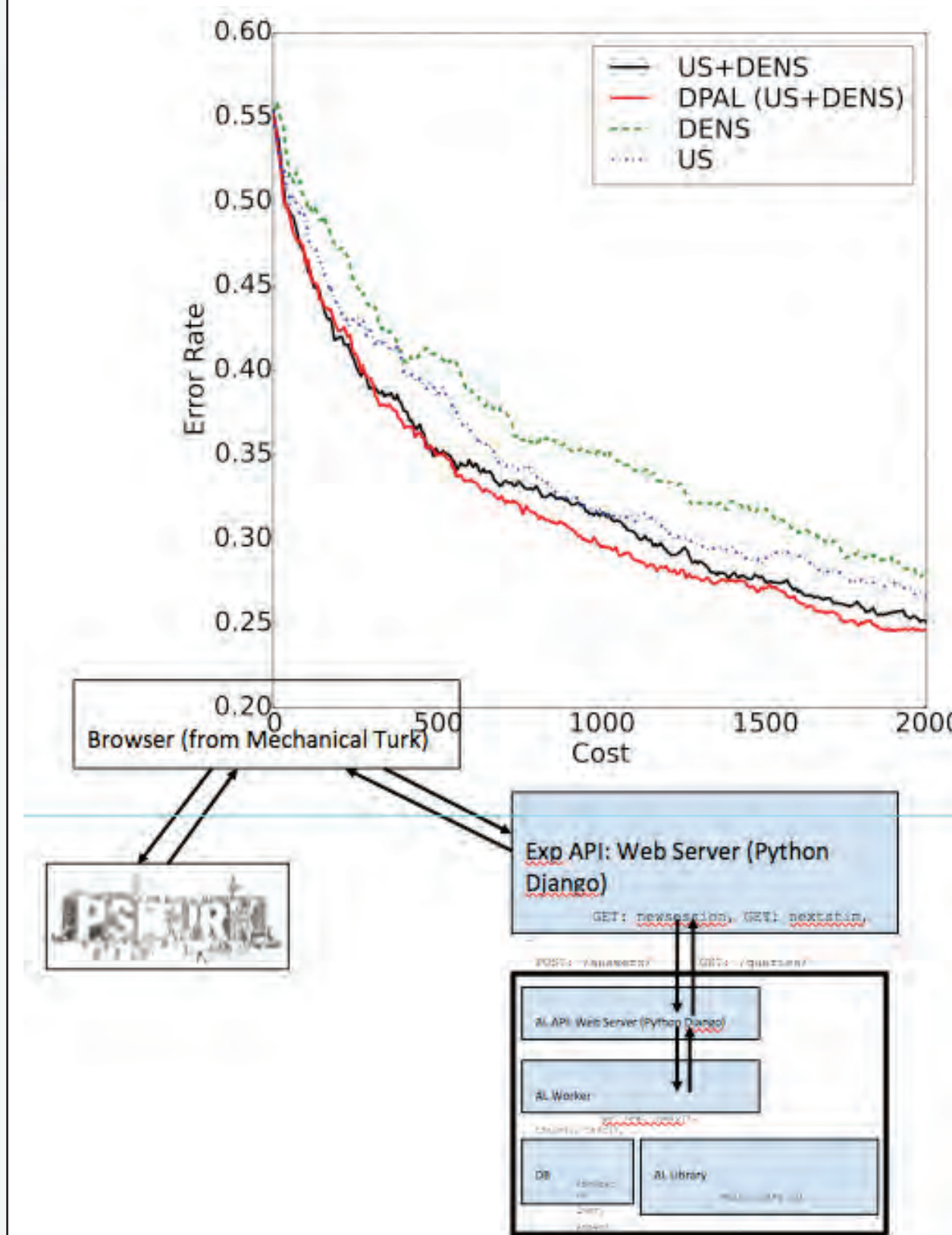
Weights for criteria W :

Multiple PAL criteria Ψ_j :

Utility of a sample $U_{ij} \in \mathbb{R}$: X

DPAL provides a framework to combine multiple factors in choosing points, including factors related to analyst performance. It shows promise in simulation and will be put to the test in a human-subject experiment.

Preliminary DPAL Results



Future work includes joint optimization of classifier and analyst objectives, extension of the experimentation software to support multi-session and team experimental trials, and a test of transferability of the model problem results to the target domain.

To keep pace with adaptive adversaries, our cybersecurity defenses must take advantage of both machine learning and human analyst strengths. Future solutions should optimize for success of the overall system.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution except as restricted below.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

* These restrictions do not apply to U.S. government entities.

DM-0002856