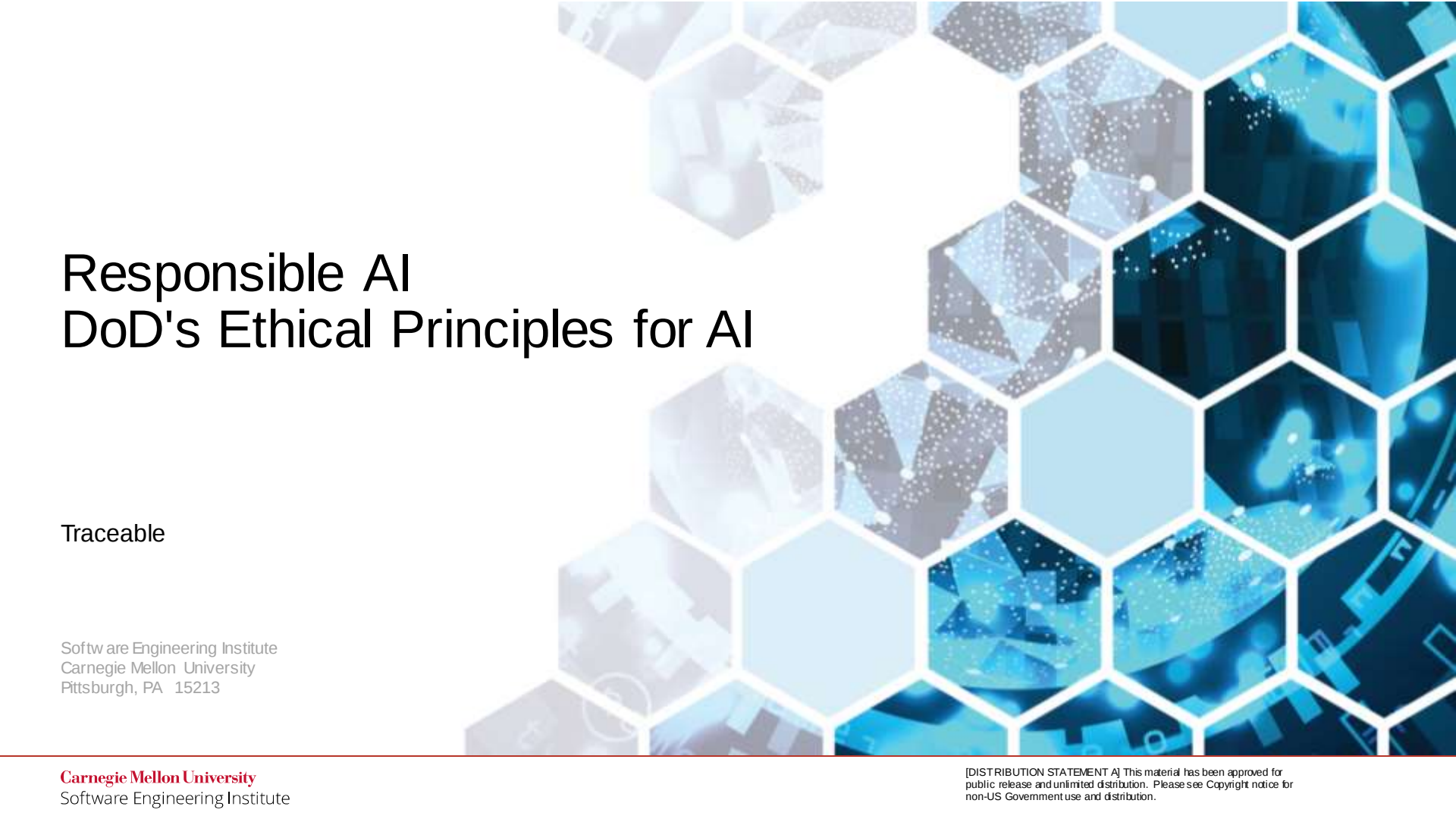




Responsible AI

DoD's Ethical Principles for AI

Traceable

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

Copyright 2022 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

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.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

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 use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM22-0011

Traceable

Traceable

AI capabilities will be developed and deployed such that relevant personnel possess an **appropriate understanding** of the technology, development processes, and operational methods applicable to AI capabilities, including with **transparent and auditable** methodologies, data sources, and design procedure and documentation.

AI transparency

Challenge "unknowable" and "opaque" statements.

Ask questions, get answers.

- How can we have accountability?
- What metrics are being used?
- Who is ultimately responsible?



Data transparency

Data should be understandable and explainable

- Provenance
- Creator's motivation, composition, and collection

Transparency improves with use of

- Datasheets for Datasets¹
- Model Cards for Model Reporting²

Datasheets for Datasets. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford. Communications of the ACM, December 2021, Vol. 64 No. 12, Pages 86-92 10.1145/3458723
<https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext#>

M. Mitchell et al., "Model Cards for Model Reporting," Proc. Conf. Fairness Account. Transpar., pp. 220–229, Jan. 2019, doi: 10.1145/3287560.3287596

Transparent and auditable methodologies and data sources

Datasheet: *Your Dataset Name Here*

Author: *Your Name Here*

Organization: *Your Organization Here*

Motivation

The questions in this section are primarily intended to encourage dataset creators to clearly articulate their reasons for creating the dataset and to promote transparency about funding interests.

1. For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

Your Answer Here

2. Who created this dataset (e.g. which team, research group) and on behalf of which entity (e.g. company, institution, organization)?

Your Answer Here

3. What support was needed to make this dataset? (e.g. who funded the creation of the dataset? If there is an associated grant, provide the name of the grantor and the grant name and number, or if it was supported by a company or government agency, give those details.)

Your Answer Here

4. Any other comments?

Your Answer Here

Datasheets for Datasets. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford. Communications of the ACM, December 2021, Vol. 64 No. 12, Pages 86-92
10.1145/3458723

<https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext#>

“markdown-datasheet-for-datasets” Josh Meyer.

GitHub: <https://github.com/JRMeyer/markdown-datasheet-for-datasets/blob/master/DATASHEET.md>

Transparency, trust, and confidence

Convey **appropriate** confidence and trust to personnel

- Evidence and transparency affects confidence
- Transparency $\neq$ trust
- Additional transparency can create additional complexity

Technology and methodologies are explained appropriately

- Access provided to details
- Rationale for decisions and recommendations are provided
- Context and urgency

Traceable – Review

Relevant personnel possess an **appropriate understanding** of the technology, development processes, and operational methods

- Transparent methodologies
- Data transparency
- Auditability
- Clear documentation



**Empower diverse teams in
inclusive environments.**

**Encourage deep conversations, speculation,
and imaginative thinking.**

Alex Van Deusen
Design Researcher
arvandeusen@sei.cmu.edu

Carol Smith
Sr. Research Scientist
cjsmith@sei.cmu.edu

Rachel Dzombak
Digital Transformation Lead
rdzombak@sei.cmu.edu